Parkinson's Disease Tremor Assessment Using Inertial Sensors

Versão final após defesa

Beatriz dos Reis Lopo Ferreira

Dissertação para obtenção do Grau de Mestre em
Bioengenharia
(2º ciclo de estudos)

Orientador: Prof. Doutor Nuno Manuel Garcia dos Santos
Co-orientador: Prof.ª Doutora Virginie dos Santos Felizardo

Abril de 2024.

Declaração de Integridade

Eu, Beatriz dos Reis Lopo Ferreira, que abaixo assino, estudante com o número de inscrição M11954 de Bioengenharia da Faculdade de Engenharia, declaro ter desenvolvido o presente trabalho e elaborado o presente texto em total consonância com o **Código de Integridades da Universidade da Beira Interior**.

Mais concretamente afirmo não ter incorrido em qualquer das variedades de Fraude Académica, e que aqui declaro conhecer, que em particular atendi à exigida referenciação de frases, extratos, imagens e outras formas de trabalho intelectual, e assumindo assim na íntegra as responsabilidades da autoria.

Universidade da Beira Interior, Covilhã 17 /04 /2024

Beatriz Ferreira

Dedication

Aos meus avós.

Acknowledgements

Em primeiro lugar quero agradecer ao meu orientador e à minha coorientadora, ao Professor Doutor Nuno Garcia, Faculdade de Ciências da Universidade de Lisboa e à Professora Doutora Virginie Felizardo, Universidade da Beira Interior. Agradeço o apoio, a disponibilidade e a ajuda para superar os desafios que surgiram ao longo da dissertação.

Agradeço também ao Vitaliy Tsymborskyy pela implementação da aplicação e colaboração no seu desenvolvimento.

Quero agradecer ao ALLab e aos seus membros pela simpatia e por me acolherem no laboratório durante os últimos dois anos.

Deixo também um agradecimento aos meus amigos, com quem partilhei muitos bons momentos e alguns mais complicados. Em especial quero agradecer aos meus antigos colegas de casa, José Mendes e Margarida Lopes, não só pelos dois anos em que partilhamos casa mas também por todos os outros antes e depois. Um obrigada pelas noites de domingo a fazer pão ou lasanha, pelas palavras de amigo nos momentos mais complicados, e pela partilha de muitas boas memórias e muitas mais que viram. À Carolina Salvador, à Júlia Ramos e à Lisana Daniel deixo também um obrigada pelo apoio e motivação em todos os momentos, pelo ombro amigo, pelas longas conversas e caminhadas. Agradeço-vos em particular pela vossa amizade.

Por último, o maior agradecimento é a minha família, aos meus pais, à minha irmã, aos meus avós e ao meu padrinho. Pelo apoio constante, pela paciência e por me ajudarem a chegar onde estou, sem eles não teria sido possível.

Resumo

A doença de Parkinson é uma doença crónica, progressiva e neurodegenerativa, que se prevê que vá ser diagnosticada a 12 milhões de pessoas até 2040. Um dos sintomas cardinais desta doença é o tremor. O tremor é caracterizado como um movimento involuntário e oscilatório de uma parte corporal e pode ser dividido em tremor de repouso, tremor postural e tremor cinético. O tremor associado a doença de Parkinson é caracterizado por um tremor regular e assimétrico de 3-6 Hz e é geralmente tremor de repouso e/ou tremor postural. Atualmente, a doença de Parkinson e o tremor são geralmente avaliados por um especialista treinado que avalia os sintomas de acordo com a *Unified Parkinson's Disease Rating Scale* (UPDRS). Contudo, devido a ser subjetivo e representar apenas uma pequena amostra de como os sintomas afetam o sujeito durante o dia, este método exibe uma alta variabilidade para o mesmo sujeito e uma baixa confiabilidade de teste-reteste. Consequentemente, outros métodos para avaliar o tremor, que não têm as mesmas limitações, estão a ser propostos e implementados. Estes métodos baseiam-se no uso de sensores inerciais, como o acelerómetro e o giroscópio, e na computação dos dados recolhidos usando esses sensores. Esta dissertação contém uma revisão sistemática da literatura e a proposta de uma aplicação móvel para a recolha de dados de acelerómetro e giroscópio durante a realização de cinco testes, três deles baseados em movimentos realizados para a avaliação com a UPDRS e dois deles que recriam atividades do dia-a-dia. Esta aplicação também inclui três questionários diários que contextualizam os sinais recolhidos. Para além disso, é proposto um framework computacional para a avaliação do tremor, incluindo as etapas de pré-processamento, extração de características e análise de dados. A etapa de análise de dados é dividida em duas tarefas, a distinção entre pessoas com a doença de Parkinson e do grupo de controlo, e a estimativa da pontuação da escala do tremor de repouso da UPDRS. O classificador *Bagging tree* foi implementado para ambas as tarefas, tendo apenas obtido bons resultados para a distinção entre os dois grupos, com uma taxa de acerto de 85.3%. Adicionalmente, foram propostos dois métodos um baseado no kurtosis e outro baseado base no número de janelas de 10 segundos do sinal em que a frequência fundamental está na banda de frequências associada ao tremor de repouso. Para estes métodos foram obtidas taxas de acerto de 83.3% e 87.88%, respetivamente.

Palavras-chave

Doença de Parkinson, Tremor, Sensores inerciais, Machine learning, kurosis, pontuação da UPDRS

Resumo Alargado

0.1 Introdução

Neste capítulo a dissertação e contextualizada e a sua motivação é apresentada. A motivação é a criação de um sistema completo para a recolha de dados de tremor com o acelerómetro e o giroscópio do telemóvel usando uma aplicação móvel e analise desses mesmos dados com o objetivo de suportar o diagnóstico da doença de Parkinson e/ou monitorizar a progressão do tremor. Para além disso, a organização da dissertação, sendo esta a mesma que a organização deste resumo alargado, é descrita. Esta dissertação está dividida em seis capítulos, sendo estes: A introdução; O estado da arte, onde e são apresentadas uma revisão bibliográfica e um revisão sistemática da literatura; O design da aplicação móvel, onde esta aplicação para a recolha de dados é proposta; O framework computacional, onde as etapas de pré-processamento, extração de características e análise de dados são detalhadas; Os resultados e discussão, onde os resultados dos métodos propostos são apresentados e discutidos; A conclusão e trabalho futuro, onde são tiradas as conclusões finais e são descritos os possíveis trabalhos futuros.

0.2 Estado da arte

Inicialmente neste capítulo é feita uma introdução a tópicos importantes para contextualizar o resto da dissertação. Estes tópicos incluem uma introdução à doença, ao tremor, ao método utilizado atualmente e aos novos métodos. Para além disso, uma revisão sistemática da literatura é apresentada, detalhando os métodos implementados para a recolha e processamento dos dados em diferentes artigos.

A doença de Parkinson é uma doença crónica, progressiva e neurodegenerativa. Mais de 10 milhões de pessoas em todo o mundo estão diagnosticadas com esta patologia e é estimado 12 milhões de pessoas terão o mesmo diagnóstico até 2040. Apesar de apresentar diversos sintomas motores e não motores, este estudo e focado no tremor, um dos seus sintomas cardinais. O tremor associado à doença de Parkinson é caracterizado como um tremor regular e assimétrico de 3 a 6 Hz, que é tipicamente tremor de repouso, apesar de alguns sujeitos também experienciarem tremor postural, e no início da doença, afeta principalmente as mãos. Atualmente, o método tipicamente usado para a avaliação da doença de Parkinson e do tremor associado a esta patologia é a *Unified Parkinson's Disease Rating Scale* (UPDRS). Apesar de ser um método validado, a avaliação clínica usando esta escala é subjetiva e referente apenas a uma amostra de como os sintomas afetam o paciente ao longo do dia. Devido a estas limitações, novos métodos para medir, analisar e quantificar tremor estão a ser propostos. Estes métodos utilizam sensores como acelerómetros e giroscópios para as recolhas de dados em ambiente controlado, como na clínica, ou para recolhas em casa ou durante o dia-a-dia normal do sujeito.

Após a aplicação dos critérios de exclusão, um total de 9 publicações foram incluídas na revisão sistemática da literatura. Esta revisão está dividida nos dados, na tecnologia, no design experimental, e no framework computacional, que contém as etapas de pré-processamento, extração de características e a análise de dados. Todas as publicações incluídas usaram acelerómetro, tendo algumas utilizado também outros sensores, para a recolha de dados. Estas recolhas foram realizadas em ambiente controlado ou durante as atividades normais do sujeito, fora da clínica, tipicamente com os sensores posicionados nas mãos ou pulsos. No pré-processamento o sinal é filtrado, segmentado e/ou a frequência de amostragem é reduzida. As características extraídas são bastante distintas entre estas publicações. Por último, para a análise dos dados foram aplicados principalmente métodos de *machine learning* ou análise estatística.

0.3 Design da Aplicação Móvel

Nesta capítulo, uma aplicação móvel para a recolha de dados de pessoas com a doença de Parkinson é proposta. Esta aplicação recolhe dados de sensores e as respostas a questionários que contextualizam estes dados. Os dados de sensores são recolhidos com o acelerómetro e o giroscópio do telemóvel a uma frequência de amostragem de 50 Hz, durante a realização de 5 testes. Estes testes são o teste do tremor de repouso, o teste do tremor postural, o teste do tremor cinético, o teste de escrever/desenhar, e o teste de levantar o copo. Os três primeiros testes são baseados nos testes realizados para a avaliação dos respetivos tipos de tremor na UPDRS e os dois últimos testes recriam atividades do dia-a-dia. A aplicação contém três questionários, sendo estes o questionário matinal, o questionário do almoço e o questionário noturno. Idealmente os testes são efetuados juntamente com os questionário três vezes ao dia.

0.4 Framework computacional

Neste capítulo, a base de dados utilizada é introduzida e o framework computacional, implementado em python, para a avaliação do tremor associado à doença de Parkinson é introduzido. O framework computacional está dividido no pré-processamento, a extração de características e a análise dos dados, que por sua vez contem dois métodos para a distinção entre pessoas com Parkinson e pessoas do grupo de controlo e dois métodos para estimar a pontuação da escala de tremor de repouso da UPDRS.

A base de dados utilizada para a implementação do framework computacional contem dados de acelerómetro de 17 pessoas com a doença de Parkinson e 17 pessoas saudáveis, como grupo de controlo. Os dados foram adquiridos com uma frequência de amostragem de 31.25 Hz, usando 5 sensores MC 10 BioStamp RC colocados no tronco, em ambas as coxas, e em ambos os antebraços, contudo para este estudo só foram usados dados dos antebraços. A recolha de dados foi iniciada durante a avaliação clínica da UPDRS e continuou durante cerca de dois dias, recolhendo dados durante a atividade normal do dia-a-dia do sujeito. Para além disso, esta base de dados contém a pontuação da escala de tremor de repouso da UPDRS determinada durante a avaliação clínica. Um intervalo de 4 horas, com início 5 minutos depois do fim da avaliação, foi selecionado para cada antebraço de cada paciente. Os pacientes estão sob o efeito de medicação durante este período.

Na etapa de pré-processamento, a magnitude do sinal do acelerómetro durante o período de 4 horas selecionado é calculada. Um filtro Butterworth passa-alto de 4ª ordem com uma frequência de corte de 0.5 Hz foi usado e o sinal dos eixos X, Y, e Z e da magnitude foram segmentados em janelas de 10 segundos.

As características extraídas do sinal foram características dos domínios do tempo e da frequência. As características extraídas do domínio do tempo foram o valor quadrático médio, o alcance, a média, a variância, a skewness e o kurtosis. As características do domínio do tempo foram extraídas da densidade espectral da potência, estimada através do método de Welch. Essas características foram a área debaixo da curva, o valor de pico, a frequência fundamental, a frequência central, a dispersão da frequência, e o valor absoluto da diferença entre a frequência central e a frequência fundamental.

Para a distinção entre pessoas com a doença de Parkinson e pessoas do grupo de controlo, foi implementado o classificador Bagging tree. Este classificador foi aplicado de modo a devolve a percentagem de janelas classificadas como Parkinson. Foi considerado que uma percentagem de 50% era classificada como inconclusiva, acima era como grupo de Parkinson e abaixo como grupo de controlo. O classificador foi aplicado ao sinal de cada antebraço e o paciente é considerado como parte do grupo de pessoas com Parkinson quando pelo menos um dos antebraços tiver a mesma classificação. Para além disso, os valores de média ou soma de cada característica, para cada eixo de cada paciente foi calculado. A correlação destes valores com a distinção entre os dois grupos de pacientes foi calculada. A característica com a correlação mais elevada foi o kurtosis no eixo Y. Posteriormente, os valores de média e desvio padrão para as somas desta característica foram calculados e usados para estabelecer intervalos padrão dos valores que esta característica assume para ambos os grupos.

Para a estimativa das pontuações da escala do tremor de repouso da UPDRS o mesmo classificador, Bagging tree, foi implementado. Antes disso, para a classificação o número de janelas com uma pontuação de 1, 2, e 3 na escala foi sinteticamente aumentado. Outro método proposto para esta estimativa é baseado na frequência fundamental. Em primeiro lugar, as janelas com frequência fundamental na banda de frequências associada ao tremor de repouso foram selecionadas. A pontuação da escala da UPDRS foi estimada através do número de janelas selecionadas para o sinal de 4 horas. Adicionalmente, o método também foi aplicado para intervalos de 2 horas e 1 hora. Para tal, intervalos do número de janelas correspondente a cada pontuação da escala para cada intervalo de tempo foram estabelecidos. Ao dividir o número de janelas pelo limite definido para a pontuação mais baixa para cada intervalo a predição da pontuação da escala é obtida.

0.5 Resultados e Discussão

Neste capítulo os resultados obtidos para os métodos propostos no capítulo anterior são apresentados e discutidos.

Para a distinção entre o grupo de pessoas com a doença de Parkinson e do grupo de controlo, o classificador obteve uma taxa de acerto de 85.3%. O antebraço esquerdo do paciente 22, o antebraço direito dos pacientes 24 e 45 foram classificados como inconclusivos. Contudo, através da classificação do outro antebraço de cada um destes pacientes, os pacientes 22 e 45 foram corretamente classificados como parte do grupo de controlo e o paciente 24 foi incorretamente classificado como parte do mesmo grupo. Os pacientes 13 e 44 foram classificados como parte do grupo de controlo em ambos os antebraços, apesar de pertencerem ao grupo de pessoas com Parkinson. Estes pacientes têm um tremor com uma pontuação de 0 na escala de tremor de repouso da UPDRS, para ambos os antebraços. Esta pontuação está associada à ausência de tremor, o que pode justificar esta classificação incorreta. Os pacientes 16 e 18 foram incorretamente classificados como parte do grupo de pessoas com a doença de Parkinson, devido à classificação incorreta de um dos seus antebraços.

O kurtosis no eixo Y, obteve uma correlação de -0.71 com a distinção entre os grupos de pessoas com a doença de Parkinson e pessoas do grupo de controlo. Os valores da média do kurtosis no eixo Y de todas as janelas de cada antebraço de cada paciente foi classificado como Parkinson ou controlo de acordo com o intervalos para os valores desta característica defendidos para cada grupo. Este método obteve uma taxa de acerto de 83.3%. Ao aplicar este método aos antebraços dos pacientes 22, 24 e 45, classificados como inconclusivos pelo classificador, o paciente 24 passa a ser classificado corretamente. Deste modo, a taxa de acerto para a conjugação dos dois métodos é 88.2%.

O classificador Bagging Tree implementado para a estimativa da pontuação da escala do tremor de repouso classificou a maior parte das janelas para cada paciente como tendo uma pontuação 0, não tendo bons resultados para esta tarefa. Estas classificações podem ser devido ao tremor não ser constante. O tremor associado à doença de Parkinson variar em intensidade ao longo do dia e não está constantemente presente. Em particular para as pontuações 1 e 2 da escala, o tremor é caracterizado como infrequente. Contudo, foi possível verificar que a percentagem de janelas classificadas com a pontuação 0 é inferior para janelas com as pontuações reais 1, 2 e 3, em comparação com as janelas com pontuações reais de 0.

Para o método que usa o número de janelas com a frequência fundamental na banda de frequências do tremor de repouso, a taxa de acerto foi determinada para os intervalos de tempo de 4, 2 e 1 hora e para os eixos X, Y e Z, para a magnitude e para a média dos eixos X, Y e Z. Concluímos que as previsões falham a maioria das vezes para a magnitude, que apresenta as taxas de acerto mais baixas para todos os intervalos de tempo. As melhores taxas de acerto para os intervalos de 4 e 2 horas foram obtidas com o eixo X, e para o intervalo de 1 hora foi obtida com o eixo Y. A taxa de acerto mais elevada dos eixos, magnitude e média para todos os intervalos de tempo foi 87.88% para o eixo X no intervalo de 4 horas.

0.6 Conclusão e Trabalho futuro

Neste capítulo são retiradas as conclusões finais da dissertação e tarefas para trabalho futuro são propostas.

A aplicação móvel proposta nesta dissertação para a recolha dados de tremor pode ser complementada com alguns dos métodos propostos para a avaliação de tremor, formando um sistema para a recolha e análise do tremor associado à doença de Parkinson. Para além disso, um dispositivo com sensores pode, no futuro, ser conectado à aplicação, permitindo a recolha continua de dados. Este dispositivo pode conter sensores para além de acelerómetro e giroscópio, o pode providenciar mais informação sobre o tremor e/ou informação sobre outros sintomas. Contudo, a aplicabilidade da aplicação para a recolha de dados em ambiente clínico e em casa teria que ser testada. A recolha de dados durante avaliações clínicas da UPDRS a uma frequência de amostragem de 50 ou 100 Hz pode melhorar os resultados obtidos nesta dissertação. Para a distinção entre o grupo de pessoas com a doença de Parkinson e o grupo de controlo, os resultados do classificador podem ser melhorados aplicando a seleção de características e os resultados do kurtosis podem ser melhorados testando diferentes argumentos desta característica. Para a estimativa da pontuação da escala de tremor da UPDRS, métodos de *machine learning* não supervisionados ou semi supervisionados podem obter melhores resultados para o classificador. Por último, também para a estimativa das pontuações da escala do tremor de repouso, o cálculo das correlações entre as pontuações e as características em janelas com a frequência fundamentar na banda de frequências do tremor de repouso pode revelar novas abordagens para a estimativa das pontuações.

Abstract

Parkinson's disease (PD) is a chronic, progressive, and neurodegenerative disorder, predicted to be diagnosed for 12 million people by 2040. One of the cardinal symptoms of this disease is tremor. Tremor is characterized as an involuntary and oscillatory movement of a body part and can be divided into rest tremor, postural tremor, and kinetic tremor. The tremor associated with PD is characterized by a 3-6 Hz, regular, asymmetrical tremor and is commonly a rest and/or postural tremor. Nowadays, PD and tremor are usually evaluated by a trained specialist who assesses the symptoms according to the Unified Parkinson's Disease Rating Scale (UPDRS). However, due to being subjective and representing only a small sample of how symptoms affect the subject during the day, this method exhibits a high within-subject variability and a low test-retest reliability. Consequently, other methods to evaluate tremor that don't have the same limitations are being proposed and implemented. These methods rely on the use of inertial sensors, like an accelerometer and a gyroscope, and the computation of data collected using these sensors.

In this dissertation, a systematic literature review is presented and a mobile app is proposed for the collection of accelerometer and gyroscope sensor data during the performance of five tests, three of them are based on movements performed for the UPDRS evaluation and two of them intend to recreate activities of daily living. This app also includes three daily questionnaires that contextualize the signals collected. Furthermore, a computation framework for the evaluation of tremor is proposed, including the preprocessing, feature extraction, and data analysis steps. The data analysis step is divided into two tasks, the distinction between people with Parkinson's disease (PwPD) and healthy controls (HC) and the estimation of UPDRS rest tremor scores. A Bagging tree classifier was implemented for both tasks, achieving a good result only for the distinction between the two groups, with a success rate of 85.3%. In addition, a method based on the kurtosis and a method based on the number of 10-second windows in the signal where the fundamental frequency is in the rest tremor frequency band. These methods obtained success rates of 83.3% and 87.88%, respectively.

Keywords

Parkinson's disease, Tremor, Inertial Sensor, Systematic literature review, Computation framework.

Contents

Dedication **v**

Acknowledgements **vii**

Resumo **ix**

Resumo Alargado **xi**

0.1 Introdução . . . xi
0.2 Estado da arte . . . xi
0.3 Design da Aplicação Móvel . . . xii
0.4 Framework computacional . . . xii
0.5 Resultados e Discussão . . . xiv
0.6 Conclusão e Trabalho futuro . . . xv

Abstract **xvii**

Contents **xix**

List of Figures **xxiii**

List of Tables **xxv**

Acronyms and Abbreviations **xxvii**

1 Introduction **1**

1.1 Context and motivation . . . 1
1.2 Problem statement and objectives . . . 1
1.3 Main contributions . . . 2
1.4 Dissertation organization . . . 2

2 State-of-the-Art **5**

2.1 Background . . . 5

2.2 Systematic literature review 11
2.2.1 Search strategy 11
2.2.2 Eligibility criteria 12
2.2.3 Extraction of study characteristics 12
2.2.4 Results 13
2.2.5 Discussion 19
2.2.6 Conclusion 27

3 Mobile App Design 29

3.1 Sensor Data 29
3.2 Questionnaires 30
3.3 Implementation 32
3.4 Context 34

4 Computation Framework 35

4.1 Dataset 36
4.2 Preprocessing 38
4.3 Feature Extraction 39
4.4 Distinction between PD and HC 40
4.4.1 Classification 40
4.4.2 Kurtosis 42
4.5 UPDRS rest tremor score estimation 42
4.5.1 Classification 42
4.5.2 Fundamental frequency within rest tremor frequency 43
4.6 Performance evaluation metrics 43

5 Results and Discussion 45

5.1 Dataset 45
5.2 Welch's periodograms 45
5.3 Distinction between PD and HC 48
5.3.1 Classification 48

5.3.2 Kurtosis . . . 50
5.4 UPDRS rest tremor estimation . . . 51
5.4.1 Classification . . . 51
5.4.2 Fundamental frequency within rest tremor frequency . . . 53

6 Conclusions and Future Work **57**
6.1 Conclusion . . . 57
6.2 Future Work . . . 58

Bibliography **61**

A Appendix **65**
A.1 Questionnaires . . . 65
A.1.1 Morning questionnaire . . . 65
A.1.2 Lunch questionnaire . . . 66
A.1.3 Night questionnaire . . . 68
A.2 Welch's Periodograms . . . 72

List of Figures

2.1 (A) Deterioration of the substancia nigra in PD, compared to normal. In advanced cases, the loss of dopaminergic neurons results in the depigmentation of the SN. (B) Lewy body in an SN neuron. 6

2.2 Assessment procedure for each type of tremor. 9

2.3 Confusion matrix. 10

2.4 PRISMA diagram of selection and exclusion of publications. 13

2.5 The NIMBLE patch.(A) Adhesive side that is placed on the skin; (B) Top of the patch. 21

2.6 Wearable IMU sensor glove. 22

3.1 Schematic representation of the data collected with the app. 29

3.2 Example text file with the data collected for the rest tremor test of the right arm using the mobile app. 30

3.3 (A) Page in the mobile app where the subject selects the questionnaire they want to answer; (B) Following page where the selected questionnaire is answered. 33

3.4 (A) Page in the mobile app where the subject selects the test they want to perform; (B) Following page where the subject selects the arm that they will use to perform the test; (C) Following page where the test can be initiated and the countdown until the end of the test is displayed. 34

4.1 Schematic representation of the proposed computation framework. 35

4.2 Top and bottom views of the BioStamp RC. 37

5.1 Welch's periodogram for the X, Y, and Z axes and magnitude (Mag) from the left forearm of patient 5. 46

5.2 Welch's periodogram for the X, Y, and Z axes and magnitude (Mag) from the right forearm of patient 12. 46

5.3 Welch's periodogram for the X, Y, and Z axes and magnitude (Mag) from the right forearm of patient 17. 47

5.4 Welch's periodogram for the X, Y, and Z axes and magnitude (Mag) from the left forearm of patient 36. 47

5.5 Welch's periodogram for the X, Y, and Z axes and magnitude (Mag) from the left forearm of patient 6. 48

5.6 Classification results with, where Recall(1) is the percentage with both forearms and Recall(1)_left and Recall(1)_right are the classification for the left and right forearms respectively. 49

A.1 Welch's periodogram for the X, Y, and Z axes and magnitude (Mag) from the left forearm of patient 5, with the scale of the PSD axis adjusted to 0.4 g^2/Hz. 72

A.2 Welch's periodogram for the X, Y, and Z axes and magnitude (Mag) from the right forearm of patient 12, with the scale of the PSD axis adjusted to 0.4 g^2/Hz. 72

A.3 Welch's periodogram for the X, Y, and Z axes and magnitude (Mag) from the right forearm of patient 17, with the scale of the PSD axis adjusted to 0.4 g^2/Hz. 73

A.4 Welch's periodogram for the X, Y, and Z axes and magnitude (Mag) from the left forearm of patient 36, with the scale of the PSD axis adjusted to 0.4 g^2/Hz. 73

A.5 Welch's periodogram for the X, Y, and Z axes and magnitude (Mag) from the left forearm of patient 6, with the scale of the PSD axis adjusted to 0.4 g^2/Hz. 74

List of Tables

2.1 Various symptoms that occur at each motor phase of PD. 6

2.2 UPDRS scores for resting tremor. 7

2.3 Information collected from the publications regarding the data, technology and experimental design. 15

2.4 Information collected from the publications regarding the data, technology, and experimental design (continuation). 16

2.5 Information collected from the publications regarding the computation framework. 17

2.6 Information collected from the publications regarding the computation framework (continuation). 18

2.7 Common baseline features extracted from the time and frequency domains in [11]. 26

4.1 Demographic and clinical data of individuals in the dataset. 38

4.2 Features extracted from the frequency and time domains. 39

4.3 Features extracted from the periodograms and theirs description. 40

4.4 Parametrization used for the tested classifiers. 41

4.5 UPDRS scores and the corresponding number of windows for the 4,2, and 1 hour intervals. 43

5.1 Results in percentage of the accuracy and recall for the five tested classifiers. 48

5.2 Mean and standard deviation of kurtosis, in the Y axis, values for the PD and control groups. 51

5.3 Predicted percentage of windows with each UPDRS rest tremor score for subjects' forearms with each UPDRS rest tremor score. 52

5.4 Predicted percentage of windows classified with each UPDRS rest tremor score for the PD and HC groups. 53

5.5 Analysis and prediction performed on the data on both forearms of patient 6 for the 4 hour time period. 53

5.6 Predicted and real UPDRS rest tremor scores for patient 6 in a 4 hour time period. 54

5.7 Success rate for all axes, the magnitude and the mean of the number of windows in the 4, 2, and 1 hour intervals. 54

Acronyms and Abbreviations

Acc	Accuracy
ADL	Activities of daily living
AUC	Area under the curve
CNN	Convolution neural network
DBS	Deep Brain Stimulation
DKS	Dyskinesia score
DT	Decision Tree
EMG	Electromyography
ECG	Electrocardiogram
F0	Fundamental frequency
F50	Central frequency
FDS	Fluctuating and dyskinesia score
FFT	Fast Fourier Transform
FN	False negative
FP	False positive
FT	Fourier Transform
HC	Healthy controls
IMU	Inertial measurement unit
kNN	k-nearest neighbors
La	Acceleration level
LOSO	Leave one subject out
MFCCs	Mel frequency cepstral coefficients
MIL	Multiple-instance learning
MLP	Multi-Layer Perceptron
PD	Parkinson's Disease
PD-VME	Parkinson's Disease Virtual Motor Exam
PKG	Parkinson's KinetiGraph
PPG	Photoplethysmography
PRISMA	Preferred Reporting Items for Systematic Review and Meta-Analysis
PSD	Power spectral density
PTT	Percent Time Tremor
pv	Peak value
PwPD	People with Parkinson's Disease
REM	Rapid eye movement
RF	Random forest
RMS	Root mean square
RT	Rest tremor
SD	Standard deviation
SF50	Frequency dispersian

SLR	Systematic literature review
SMOTE	Synthetic Minority Over-sampling Technique
SN	Substancia nigra
SVM	Support vector machine
tip	Tremor intensity parameter
TN	True negative
TP	True positive
UPDRS	Unified Parkinson's Disease Rating Scale

Chapter 1

Introduction

1.1 Context and motivation

Parkinson's Disease (PD) is the fastest growing neurological disease and the second most common neurological disorder [1, 2]. This disease can cause several motor and non-motor symptoms, with tremor being one of the cardinal symptoms [2]. Tremor is the most common movement disorder and an individual with PD can have rest tremor, postural tremor, and/or kinetic tremor. In PD this symptom typically manifests at rest, and during the onset of the disease, tends to affect the hands [1, 3]. The method commonly used to evaluate PD and tremor associated with PD is the Unified Parkinson's Disease Rating Scale (UPDRS) [1]. Despite being a validated method the UPDRS has limitations, which lead to the proposal of new methods, such as the use of inertial sensors, like accelerometer and gyroscope, and machine learning algorithm to measure and quantify tremor [1, 3].

The main motivation for this dissertation is to provide a complete system for the collection of tremor data using an accelerometer and gyroscope with a mobile app, that could be used in the clinic or at home, and the analysis of that data to help support diagnosis and/or monitor the progression of tremor and the disease.

1.2 Problem statement and objectives

PD manifests differently for different people, which makes the monitoring of the disease progression and symptoms extremely important in the selection and changes necessary in the treatment plan [2, 4]. However, the UPDRS, currently used for the monitoring of PD and its symptoms exhibits a high within-subject variability and low test-retest reliability [1, 5]. Furthermore, the clinical assessments using this scale are subjective to the specialist performing the assessment and do not provide an accurate representation of how tremor affects the patient during the day [1, 5]. The implementation of new methods based on wearable inertial sensors and data analysis are promising for the accurate monitoring of PD and tremor since they are not subjected to the same limitations as the UPDRS.

The mobile app proposed collects accelerometer and gyroscope sensor data using sensors in the phone and answers to three daily questionnaires to contextualize the sensor data.

For the computation framework, steps of preprocessing, feature extraction, and data analysis are implemented using the dataset collected in [6]. The data analysis will include a short evaluation of Welch's periodograms extracted from the signals of each subject of the dataset. However, this step will be mainly focused on the distinction of between people with PD (PwPD) and healthy controls (HC) and the estimation of UPDRS rest tremor scores. A Bagging tree classifier will be implemented for both these tasks and a different analytical method will be proposed for each task.

The main objective of this dissertation is to propose different approaches for the evaluation of tremor associated with PD and obtain promising results for both tasks mentioned previously.

1.3 Main contributions

The main contributions of this study are the following:

- A systematic literature review summarizing the datasets, technology, experimental design, and computation framework used in different publications for the assessment of PD tremor;
- A proposed mobile app for the collection of sensor data;
- A method using a Baging tree classifier and values of kurtosis extracted from the signal's features to distinguish between PwPD and HC;
- A method based on the number of windows from the signal where the frequency of their highest peak in Welch´s periodogram is in the typical rest tremor frequency band for the estimation of UPDRS rest tremor scores.

1.4 Dissertation organization

This dissertation contains the following chapters:

- Chapter 1: This chapter contains a brief introduction of the topics of this study, contextualizing the dissertation, the motivation, and objectives, as well as the main contributions of this study;
- Chapter 2: This chapter contains a state-of-the-art, where context on the disease and the methods used in its evaluation are provided, and a Systematic Literature Review summarizing and detailing the methods used for the assessment of tremor in different publications;
- Chapter 3: In this chapter a mobile app for the collection of sensor data and answers for three daily questionnaires is proposed;

- Chapter 4: In this chapter, the dataset is introduced and methods for the distinction between PwPD and HC and the estimation of UPDRS rest tremor scores are proposed;
- Chapter 5: This chapter contains the results of the proposed methods and their discussion;
- Chapter 6: This chapter discusses the conclusions of this dissertation and establishes prospects for future work.

Chapter 2

State-of-the-Art

2.1 Background

Parkinson's disease (PD) is a chronic, progressive, and neurodegenerative disorder [1]. It is the fastest growing neurological disease [7], and it is estimated that 12 million people will be diagnosed by 2040 [5]. Currently, more than 10 million people worldwide are diagnosed with PD, making it the second most common neurodegenerative disorder [2]. Despite the existence of risk factors, like head injury or exposure to toxic chemicals, PD is likely to have a multifactorial etiology, resulting from both environmental and genetic factors [8].

This disease is defined by the degeneration of dopaminergic neurons in the brain's substancia nigra (SN) and is associated with intraneural protein aggregates, the Lewy bodies, and Lewy neurites. Lewy bodies are cytoplasmic inclusions containing α-synuclein aggregates. Furthermore, recent studies found that the loss of dopaminergic terminals in the striatum is crucial to the onset of PD motor symptoms. Nonetheless, PD also involves non-dopaminergic neurons [8]. In addition, the Braak model suggests that the disease starts in the medulla and olfactory bulb, and, in this phase, is mainly associated with rapid eye movement (REM) sleep behavior disorder and decreased smell. The disease then progresses to the SN pars compacta and other basal forebrain and midbrain structures and is associated with the typical motor symptoms. In a more advanced stage, it progresses to the cerebral cortices and leads to cognitive impairment and hallucinations [9]. The effect of PD in the substancia nigra as well as a Lewy body are shown in figure 2.1.

This disease may cause the patients to feel several motor and non-motor symptoms, such as postural instability, dyskinesia (involuntary muscle movement, sometimes caused by long-term use of medication), muscular rigidity, dementia, impaired speech, hyposmia, rapid eye movement sleep behavior disorder, personality change, and depression [2, 10, 11, 1]. Furthermore, the symptoms most associated with each motor stage of the disease are described in table 2.1. Nonetheless, the cardinal symptoms, generally used to make a diagnosis, are bradykinesia, rigidity, and tremor [2].

From these cardinal symptoms, the tremor is considered the most common movement disorder and is characterized as an involuntary, oscillatory movement of a body part, classified according to its clinical features and cause. An individual with PD can display one or more types of tremor, those being rest tremor (RT), postural tremor and kinetic tremor [3].

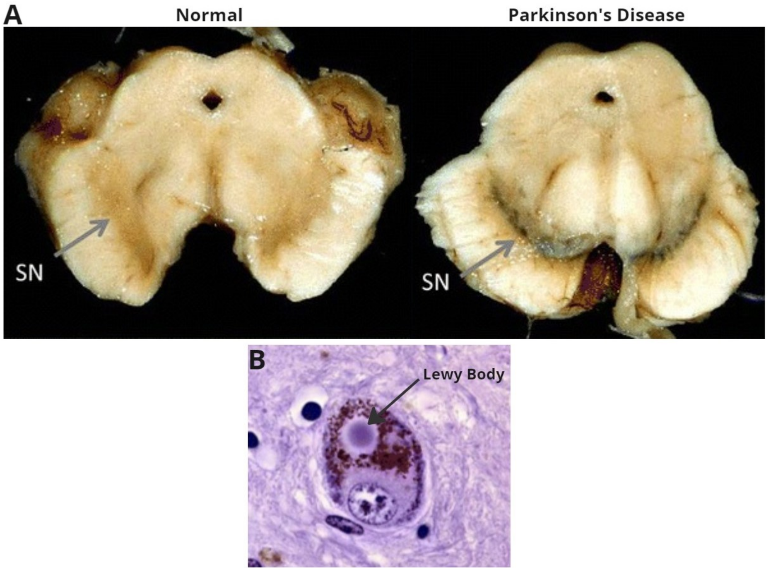


Figure 2.1: (A) Deterioration of the substancia nigra in PD, compared to normal. In advanced cases, the loss of dopaminergic neurons results in the depigmentation of the SN. (B) Lewy body in a SN neuron. Adapted from [8].

Table 2.1: Various symptoms that occur at each motor phase of PD. Retrieved from [10].

Stages	**Symptoms**
Stage 1	One sided resting tremor, misaligned posture, changes in walking pattern, and altered facial expressions.
Stage 2	Symmetric tremor and movement disorders, altered gait and posture.
Stage 3	Imbalance, poor gait, and sudden fall due to imbalance are often present.
Stage 4	Aggravated movement disorders, would require some assistance for standing and walking.
Stage 5	Usage of the wheelchair becomes necessary as legs become stiff, making standing or walking impossible.

The tremor associated with PD in particular is characterized by a 3-6 Hz, regular, asymmetrical tremor. This symptom usually manifests itself at rest and, at the onset of the disease, commonly affects the hands. Regardless, a moderately severe postural tremor is experienced by half the PD patients [1, 3, 12].

A patient experiencing postural tremor typically also experiences rest tremor. Despite the distinction between these types of tremor being difficult, it is an important step in the correct diagnosis of different subtypes of PD and differentiation between PD and essential tremor [1].

Due to the heterogeneity of this disease, accurate monitoring and assessment of symptoms, such as tremor, is extremely important in selecting the correct treatment and in making changes to the treatment plan [2, 4]. Changes in the patients' treatment plan are essential as the disease progresses and the severity of the symptoms increases, helping maintain the symptoms controlled. Currently, the available treatments for PD motor symptoms include pharmacological, surgical, Deep Brain Stimulation (DBS), and other therapeutic interventions [7].

Nowadays the golden standard to quantify the severity of PD and characterize tremor is the Unified Parkinson's Disease Rating Scale (UPDRS). This scale is a validated method that requires an experienced professional to perform the assessment. Due to the scores being given according to the visual perception of the examiner, this evaluation is subjective to him, making this method have a low degree of objectivity, impartiality, and sensitivity [1]. In addition, the UPDRS is, sometimes, supplemented with patients' diaries, where the patients take note of their symptoms throughout the day. Despite providing a better understanding of the severity of symptoms, these diaries are not completely reliable due to often not being maintained over a long period of time and depending on the patient's perspective [7].

The UPDRS is divided into four parts, those being:

1. Non-Motor Aspects of experiences of Daily-Living;
2. Motor Aspects of Experiences of Daily Living;
3. Motor Examination;
4. Motor Complications.

Tremor is evaluated in the second and third parts. In the second part, the patient is asked how their shaking or tremor was during the past week, and the scores from 0 to 4 are given according to the answer. In the third part, the patient performs specific movements, presented in figure 2.2, and a specialist visually assesses the amplitude of the tremor, giving the corresponding score [13]. Furthermore, the given scores and their guidelines are presented in table 2.2.

Table 2.2: UPDRS scores for resting tremor. Retrieved from [10].

RT scores	**Guidelines**
0	Absent/normal
1	Slight (infrequently present for a short duration)
2	Mild (infrequently present for a considerably long duration)
3	Moderate (present most of the time with moderate amplitude)
4	Severe (present most of the time with marked amplitude)

Current methods exhibit a high within-subject variability and low test-retest reliability, and thus cannot accurately quantify the tremor presented by an individual [1, 5]. Furthermore, physical exams performed in a clinic provide only a small sample of PD signs, which may not accurately represent a patient's function at home [5].

As a result of the disadvantages stated, new methods used to measure, analyze, and quantify tremor, that don't rely exclusively on the interpretation of the experienced professional, are being proposed. Some of these methods make use of transducer-based methodologies, like accelerometry, electromyography, gyroscope data, electromagnetic tracking, actigraphy, or digitizing tablets [3]. Methodologies such as accelerometry or gyroscope data require the use of inertial sensors, that transduce inertial force into a measurable electrical signal. These sensors measure voluntary or involuntary movements. Their use helps discriminate different types of tremor and support a clinical diagnosis [1]. Research has shown the feasibility of using these sensors to quantify motor signs of PD [5].

Accelerometers are considered the minimum necessary sensor to characterize human activity. This sensor consumes a relatively low amount of power when compared to other inertial sensors, like gyroscopes. Nonetheless, it is not certain whether accelerometer data alone is sufficient or whether other sensors are necessary to characterize the symptoms and improve detection. Data collected from gyroscopes attached to the forearm or finger of a patient are also effective in the measurement of tremor [2]. Additionally, the sampling rates of these wearable sensors can be adjusted to increase temporal resolution, improving model accuracy at the cost of an increase in the power and memory necessary [7].

The mechanism underlying the measure of acceleration is based on the principles of Hooke's Law, $F = kx$, and Newton's second law of motion, $F = ma$. In addition, this mechanism can be described as a mass-spring system. When the system is subjected to a compression or stretching force due to movement, the spring, whilst returning to its state, will generate a restoring force proportional to the compression or stretching. The mass, m, and the stiffness, k, of the spring can be controlled and thus, the acceleration of the mass element can be determined from the displacement, as is shown in equation 2.1 [14].

$$F = kx = ma, \text{ so } \quad a = \frac{kx}{m} \quad (2.1)$$

Despite being accurate, cost-effective, and widely available, most of these devices are not used for clinical or home monitoring [3, 4]. This is due to their time-consuming measurements and complexity. Furthermore, different setups and criteria are not standardized and there is a lack of evaluations in large clinical trials of the diagnosis accuracy using this technology [3]. However, accuracies from the existing trials exceed 85% for the detection of tremor and bradykinesia, indicating the viability of these technologies for the monitoring of PD symptoms [7].

The implementation of this technology in continuous in-the-wild monitoring is promising, however, it is heavily limited by practical considerations, such as the battery life and memory capacity of the devices. These limitations can lead to a need to remove or replace the devices, which may lead to intermittent usage or inconsistent device positioning [7]. Another problem is the difficulty in distinguishing between symptoms like tremor and normal daily activities, due to noise in a signal collected in uncontrolled living conditions [11]. Nonetheless, since resting tremor is an occasional symptom, continuous monitoring is important to determine the severity of PD [10].

The patient is typically asked to perform different movements or have their arms in different positions during both clinical evaluations by a specialist and clinical or controlled evaluations using wearable sensors. Typically, the patient is asked to position their arm or hand in a resting state, for the rest tremor evaluation. For the postural tremor evaluation, the patient is asked to hold their arm against gravity. Lastly, to evaluate kinetic tremor, the patient is asked to perform the finger-to-nose maneuver [2, 4]. These movements are represented in figure 2.2 and are the same movements performed during UPDRS evaluation of these symptoms, where the tremor is accessed by the specialist. Furthermore, when sensors are used to classify tremor in a controlled environment it is also common to have patients perform Activities of daily living (ADL) to recreate signals generated in their daily lives.

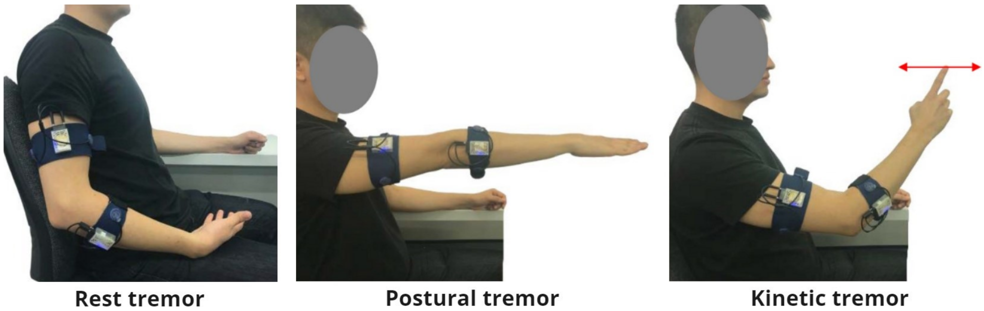


Figure 2.2: Assessment procedure for each type of tremor. Adapted from [2].

In general, the methods used to quantify PD symptoms can be divided into two categories. In the first approach, external sensor data is correlated with PD symptoms through the creation of a model based on the physical and physiological mechanisms of symptoms. This is then used in correlation with the UPDRS scores. On the other hand, the second approach consists of the extraction of features from the signal and the implementation of a regression or classification model to minimize prediction errors in the UPDRS scores. Due to the lack of an established model for the correlation of UPDRS scores with PD symptoms, machine learning algorithms are typically used for the analysis [2].

Most of the feature sets proposed for PD detection consist of time domain features, like the mean or range, and/or frequency domain features, like the dominant frequency, the energy content in a particular band or signal entropy [11]. Additionally, there is an important trade-off between the costs and benefits associated with data complexity in disease monitoring. For example, features based on complex processing techniques, like Fast Fourier Transform (FFT) or sample entropy, capture subtly meaningful characteristics of a signal, potentially increasing the accuracy of symptom detection. However, these techniques require more computation time, device memory and use more device power, increasing the costs [7].

Machine learning algorithms create mathematical models that produce an output using information learned from the input data. These algorithms can be supervised, unsupervised, or semi-supervised. The supervised machine learning requires labeled data, with each training input being associated with an output value. These methods are divided into regression, when output data is a continuous range of values, and classification, where the output data is a finite discrete set [15]. The unsupervised machine learning does not require the data to be labeled and instead tries to find the output by grouping the input data by their representative characteristics. Lastly, semi-supervised machine learning uses a mixture of labeled and unlabeled data for the training [15].

The performance of machine learning algorithms is typically evaluated with a confusion matrix, like the one shown in figure 2.3 for a problem with two classes, where true negatives (TN) is the number of input windows or samples of the signal with a negative label correctly classified as belonging to the negative class, false positives (FP) is the number of inputs with a negative incorrectly classified, false negatives (FN) is the number of inputs with a positive label incorrectly classified and true positives (TP) is the number of inputs with a positive label correctly classified [16].

	Predicted Negative	Predicted Positive
Actual Negative	TN	FP
Actual Positive	FN	TP

Figure 2.3: Confusion matrix. Retrieved from [16].

Some metrics commonly in the evaluation of the performance are accuracy, recall, and precision, calculated using values in the confusion matrix as shown in equations 2.2, 2.3, and 2.4, respectively. The accuracy shows how often the algorithm makes the correct prediction overall, the recall shows the amount of samples from the positive class that is correctly predicted to belong to that class, and the precision shows the amount of samples correctly predicted from all the samples predicted to be positive [16].

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (2.2)$$

$$Recall = \frac{TP}{TP + FN} \quad (2.3)$$

$$Precision = \frac{TP}{TP + FP} \quad (2.4)$$

Currently, the processing of these signals is mainly done using machine learning. Machine learning can be defined as the application of mathematical algorithms to find patterns or structure data and make predictions, based on previous training with data. Machine learning expands on statistical analysis methods, dealing better with high dimensional and nonlinear data, which is important in the processing of wearable sensor data [15].

2.2 Systematic literature review

This chapter will explore the methodologies currently used to quantify the tremor of individuals with PD. For this purpose, the state-of-the-art will be in the form of a systematic literature review (SLR). This type of review establishes a method to select publications and extract and summarize their content. Furthermore, for the SLR, research questions can be defined and a research protocol that can be replicated must be established.

The constant increase in the number of new technologies and methods, as well as the evolution of previously used methods in this area, lead to the constant necessity of new reviews. A SLR has the advantage of methodically extracting and presenting the methods used in the field, providing a comparison between them. There are recent publications of SLR in this area. Despite this, it was still opted to do an SLR for this thesis, due to the benefits of this review. This will also allow the selection and extraction of important information for the practical part of this thesis.

Therefore, in this chapter, an SLR comparing the different datasets, technologies, experimental designs, and computational frameworks, currently used is presented.

2.2.1 Search strategy

The databases selected for the research were Scopus, IEEE Xplore, ScienceDirect, ACM Digital Library, and PubMed. These databases were chosen based on their disciplinary fields of focus. ScienceDirect, ACM Digital Library, and IEEE Xplore are focused on computer science. Pubmed is focused on the biomedical field and Scopus is an interdisciplinary database.

The keywords used were parkinson AND tremor AND (wearable AND sensor). This research was made in the title, abstract, and keywords of the publications, except for in PubMed where the research was only on the title and abstract due to the impossibility of including the search in the keywords.

Furthermore, the research was limited between the years 2018 and 2022, to include the last five years.

2.2.2 Eligibility criteria

Initially, the duplicate articles are removed. Then the remaining publications obtained through the database research were screened for their eligibility according to previously defined exclusion criteria. The web application Rayan QCRI [17] was used for this screening. Furthermore, the exclusion of publications can be divided into two phases.

The exclusion criteria were defined based on the research questions that this review aims to answer. Those research questions are:

- Most common approach/application of wearable technology on tremor assessment?
- Main limitations of the current state-of-the-art?

In the first phase of exclusion, the publications were screened by their title and abstract. These publications were excluded if they were a review, editorial, book chapter, or erratum. Furthermore, publications that were not regarding tremor, sensors, and PD were also excluded. Publications that evaluated more than one disease, used additional data collection methods besides wearable sensors, or evaluated additional symptoms of PD besides tremor were included, as long as they also evaluated tremor associated with PD using sensors. Furthermore, in case of doubt that these publications fit these criteria they were included and screened in the second phase.

In the second phase of exclusion, publications with less than 10 patients were excluded since the patient sample size did not allow for the reliability of the study results. Publications that did the study for less than 5 days were also excluded since this review is intended to focus on continuous monitoring of tremor. If a publication had the wrong study design, meaning that the tremor was not from PD patients and was instead simulated or the publication did not present isolated results for the tremor assessment, or for the processing of the tremor signal, it was excluded. Furthermore, publications with no access or that were not written in English were also excluded. Lastly, publications that did not regard sensors, tremor, or PD and were not excluded in the first phase will be excluded.

2.2.3 Extraction of study characteristics

The information extracted from the publications was divided into tables 2.3, 2.4, 2.5, and 2.6 and belonged to one of four categories. These categories were chosen in order to extract the most relevant information regarding the topics this review intends to encompass. The first tables, 2.3 and 2.4, contain the categories of data, technology, and experimental design. In turn, the remaining tables, 2.5 and 2.6, contain the extracted information referring to the computation framework.

The data category aims to assess the different characteristics of the studied datasets. The technology category defines the sensor used for the data collection as being accelerometers, gyroscopes, magnetometers, or others. The experimental design establishes the settings in which the data was collected for the different publications. Lastly, the computation framework, determines the algorithms and methods used in the preprocessing, feature extraction, and classification of the signal data.

2.2.4 Results

The publication selection process followed the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) flowing diagram [18]. This flowing diagram is presented in figure 2.4.

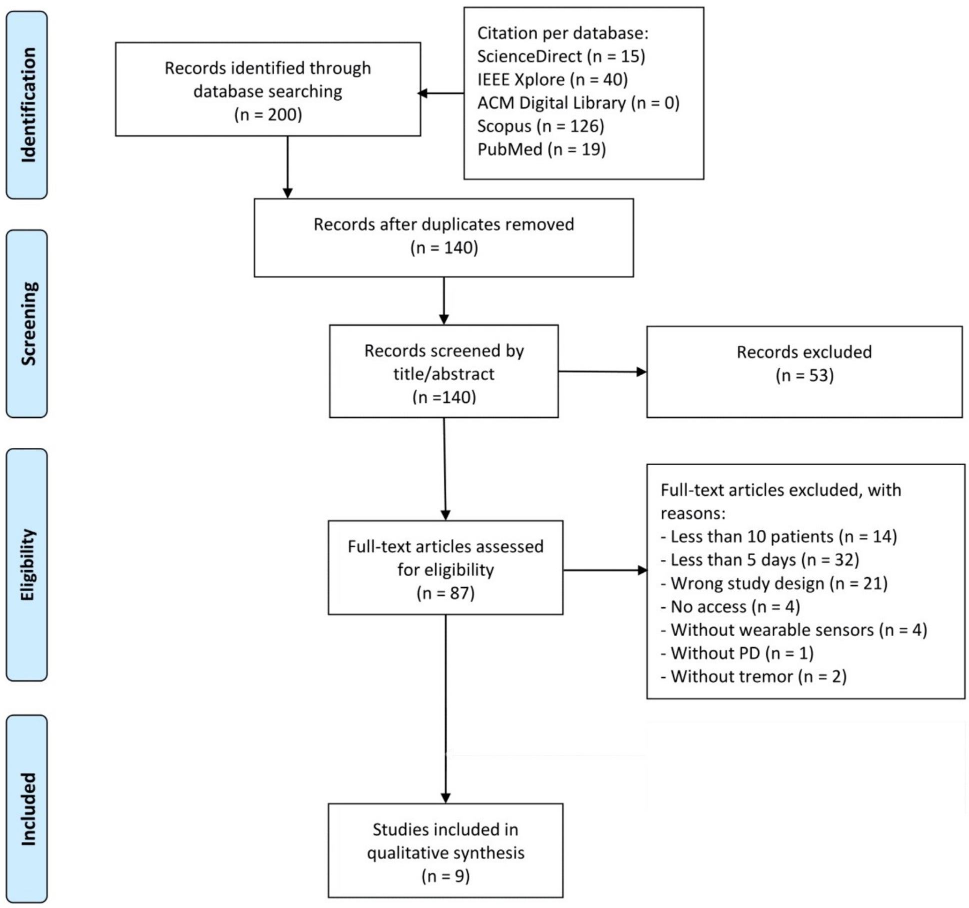


Figure 2.4: PRISMA diagram of selection and exclusion of publications.

The database research resulted in a total of 200 publications. From those 15 were from ScienceDirect, 40 from IEEE Xplore, 0 (zero) from ACM Digital Library, 126 from Scopus, and 19 from PubMed. Firstly the duplicates were removed, leaving 140 unique publications.

Then, in the first exclusion phase, the remaining publications were screened by their title, keywords, and abstract. In this phase, 53 publications were excluded. The 87 publications left were assessed based on the criteria previously defined for the second exclusion phase.

A total of 77 publications were excluded, 14 were due to including less than 10 patients, 32 were due to collecting data from less than 5 days, and 21 for having a wrong study design. Furthermore, 7 publications were excluded due to not including wearable sensors, PD, or tremor. Lastly, 4 publications did not have open access and, thus were also excluded.

After this selection, a total of 9 publications were included in the SLR. The information extracted from these publications can be seen in the tables 2.3, 2.4, 2.5, and 2.6.

Table 2.3: Information collected from the publications regarding the data, technology and experimental design.

Year	Study	Data					Technology				Experimental design
		Number of patients	Age	Sex (M/F)	Type of treatment	Treatment state	A	G	M	Other	
2022	[10]	30	62.1 ±5.78	18/12	Medication	Both	X	X			Resting tremor was measured 24/7, for 10 days, using wearable sensor glove placed on the tremor-affected hand. Experiment was home-based and patients were asked to rest the hand in determined positions when not performing tasks. OFF state recordings were taken as samples
2022	[5]	198 (Set 1) and 370 (Set 2)	62.9 ±8.9 and 62.3 ±8.8	128/70 and 231/139	Medication	Both	X	X		PPG and skin conductance sensors	Set 1: patients did MDS-UPDRS evaluations, in clinic, through clinical and video, rating and wore the sensors, on the wrist, collecting data up to 23h/day at home; Set 2: besides doing the set 1 data collections, patients did specific evaluation tasks in the clinic, during the OFF state, within 1h of MDS-UPDRS, and once a week at home, 2 times a day in ON and OFF states or 2 time 1h apart (not taking medication); Tremor evaluation tasks: rest, postural tremor, and arm twist.
2021	[19]	18	64.9 ±7.6	8/10	Medication	ON	X				Weekly evaluation in care center during between 2 and 8 weeks. Perform 6 scripted UPDRS activities with at least 30s resting interval in between, with sensor on the wrist of the most affected side.
2020	[20]	100	66.36±0.92	60/40	medication	ON	X				Data was collected between 6 am and 10 pm over 7 days, with the sensor on the wrist of the most affected side.
2020	[11]	10	62-85	ND	ND	ND	X				Acc positioned on both wrists. Laboratory recordings (ADL and motor tasks from Part III of UPDRS) and in-the-wild recordings of 6 patients throughout the day for 4 weeks.
2020	[12]	11	64.7	ND	Medication and 2 DBS	Both	X				Data collected in naturalistic conditions with a smartphone as participants played a game for 10s.

Table 2.4: Information collected from the publications regarding the data, technology, and experimental design (continuation).

Year	Study	Data					Technology				Experimental design
		Number of patients	Age	Sex (M/F)	Type of treatment	Treatment state	A	G	M	Other	
2019	[21]	10 (PD) and 7 (HC)	66.4 ±6.9 and 53.3 ±6.7	5/5 and 2/5	ND	Both	X	X	X		Subjects wore the sensors, collecting data, on each wrist for approximately 7 days.
2019	[22]	25	64.2 ±7.8	15/10	Medication	Both	X			EMG	Part 1: patches placed on the right and left forearms, right and left shins, chest and back of tremor-dominant hand. In the clinic patients did a full MDS-UPDRS assessment (day 1), multiple MDS-UPDRS Part III and PDSS-2 (day 2); Part 2: patches applied to the chest, shin, forearm, and back of the hand of the most affected side, use diary app, conduct a series of motor tasks (day 1, in clinic), subset of UPDRS-III motor tasks completed before and after medication (day 2, in clinic), motor tasks following instructions before and after medication (day 3, at home).
2018	[23]	57	66.4 ±9.0	35/22	Medication	Both	X	X	X		One sensor unit per limb, on top of the hand and the ankle. Complete scale motor examination. Some patients repeated the examination on different days.

M - Male; F - Female; A - Accelerometer; G - Gyroscope; M - Magnetometer; ND - Not disclosed; PPG - Photoplethysmography; MDS-UPDRS - Movement Disorder Society-Sponsored Revision of the Unified Parkinson’s Disease Rating Scale; UPDRS - Unified Parkinson’s Disease Rating Scale; UPDRS - Unified Parkinson’s Disease Rating Scale; Acc - Accelerometer; ADL - Activities of daily living; DBS - Deep Brain stimulation; PD - Parkinson’s Disease; HC - Healthy Control; EMG - Electromyography; PDSS-2 - Parkinson´s Disease sleep scale 2.

Table 2.5: Information collected from the publications regarding the computation framework.

Year	Study	Computation framework		
		Preprocessing	Feature extraction	Data analysis
2022	[10]	Band-pass filter	Estimate tremor frequency, position, and occurrence rate	Assign UPDRS score and statistical analysis: Bland Altman's analysis
2022	[5]	Filtering out gravitational component	Total tremor acceleration and lateral tremor acceleration: computing median absolute tremor acceleration of the norm of the acceleration vector and along the lateral axis of the acc signal, respectively; Lateral tremor amplitude	Statistical analysis: Spearman R, Test-retest ICC, Cohen's D
2021	[19]	Context classifier and Tremor detector	Acceleration level of the RMS amplitude in the tremor band (3.5-7.5 Hz)	Estimation of the indicator of the amplitude of tremor
2020	[20]	ND	Percent time tremor	Statistical analysis: Pearson's correlation test, ICC, independent t-test, X^2 test
2020	[11]	Down-sampling to 50 Hz, gravity removal (high-pass 3^{rd} order Butterworth filter at 0.3 Hz), division into 3s windows (150 samples/window) with 2s overlap, FT module, tremor spectrum extraction, labeling	Energy in the 3-9 Hz band, MFCCs, MFCCS-T/NT, CNN and CNN-T/NT, common baseline features, Welch's one-sided PSD	RF and MLP
2020	[12]	Labeling according to medication time	Welch's periodogram (PSD): AUC, peak value, F0, F50, SF50, |F50-F0|, tremor intensity parameter	Statistical analysis
2019	[21]	Subtract gravity component and calculate spectrogram of the acceleration, representing PSD over time, using a Blackman-Tukey nonparametric estimator of the PSD, normalize estimation of PSD	Detect periods of gait through the PSD, and find a candidate tremor frequency, where the power associated with the candidate tremor peak, γ_t(f), the power of the baseline immediately before or after the peak, γ_b(f), and the center frequency of the candidate tremor peak, f_t, are determined	If gait is not detected and power of tremor peak over the baseline is above a threshold, tremor is detected

Table 2.6: Information collected from the publications regarding the computation framework (continuation).

Year	Study	Computation framework		
		Preprocessing	Feature extraction	Data analysis
2019	[22]	Filtering and segmentation	Range, maximum value, minimum value, spectral entropy, RMS, signal entropy, cross-correlation at zero lag, dominant frequency, magnitude of dominant frequency, ratio of dominant frequency to power spectrum energy, range of auto covariance, line length, range ratio, kurtosis	Statistical analysis: Post hoc analysis, Pearson's correlation coefficients
2018	[23]	Tremor detection and Limb displacement estimation	Amplitude, amplitude average and % of tremors higher or equal to 90% of the maximum, % of tremors lower than 90% of the maximum and the % of tremors lower than 70% of the maximum	Fuzzy inference system

UPDRS- Unified Parkinson's Disease Rating Scale; ICC - intraclass correlation coefficient; RMS - Root mean square; FT - Fourier transform; MFCCs - Mel frequency cepstral coefficients; MFCCs-T/NT - Mel frequency cepstral coefficients after tremor spectrum extraction; CNN - Convolution neural network; CNN-T/NT - CNN trained on spectra after tremor spectrum extraction; PSD - Power Spectral Density; RF - Random Forest; MLP - Multi-Layer Perceptron; AUC - Area under the curve; F0 - Fundamental frequency; F50 - Central frequency (frequency which divides PSD into two equal parts); SF50 - frequency dispersion (describes the width of frequency band around F50 containing 68% of the total power of the signal).

2.2.5 Discussion

2.2.5.1 Data

Tables 2.3 and 2.4 show the data regarding the patients from which sensor data was collected in each publication. This category is divided by the number of patients, age, sex, type of treatment, and treatment state.

Every publication extracted data exclusively from PD patients, except for [21], where data was also collected from HC. The authors of [5] divided the PD patients into two sets. The first set did only in-clinic evaluations and the second set did in-clinic and at-home evaluations. Furthermore, some patients in set 1 are also included in set 2, and thus there is an overlap between these sets. The evaluations performed for each set are described in table 2.3 and subsection 2.2.5.3.

The mean age of PwPD is in the 60s for all the selected publications. Moreover, the only mean age in the 50s is regarding HC, in [21]. In [11], the ages are presented in a range and the lowest and highest ages are 62 and 85, respectively. Every dataset has a higher number of males than females, except for [19], where the opposite occurs, and [21], where, for the PD group, the number of males and females is the same and for the HC group the number of females is higher.

The type of treatment is medication in seven publications, [22, 10, 12, 5, 19, 20, 23]. Out of those studies, the medication was specified as Levodopa in [22, 20, 23]. In addition, [19] specified the medication as Levodopa and/or dopamin agonists. The remaining three studies with this type of treatment did not specify the medication used. Nonetheless, the authors of [5] indicate that 185 patients of set 1 and 297 of set 2 were taking any symptom medication, and 117 of set 1 and 202 of set 2 were taking fast acting medication. In two publications, [11, 21], the type of treatment was not disclosed and in [12] 2 of the 11 patients were on DBS, however, it is not specified if these patients also took medication. Additionally, [11] where the type of treatment was not disclosed, did not have information about the treatment state during the assessment.

No publication did the evaluation exclusively in the OFF state when the patients are not under the effect of the medication. Contrarily, two studies, [19, 20], evaluated the tremor exclusively on the ON state. On the six studies, [22, 10, 12, 5, 21, 23] that evaluated both ON and OFF treatment states, the approach was typically to do the first assessment before the patient usually took their first dose of medication or to have the patients skip or delay the intake of that dose. Afterward, patients would take their medication and be evaluated after it took effect, and they were in the ON state. It is also important to note that in [10] the recordings during the OFF state were performed due to some PD subjects not exhibiting tremor when taking the medication. In those cases, the OFF state recording was taken as a sample and the analysis was done for that data.

Given the overall advanced age of the patients, a plausible conclusion is that PD appears and/or is diagnosed at an advanced age. Moreover, it is known that symptoms' severity increases as the disease progresses, confirming that the elderly population is more affected by this disease. Furthermore, in most cases, the first non-motor symptoms can occur several years before diagnosis [9]. These conclusions are also in accordance with the information discussed in the background.

This data is in accordance with the knowledge that PD is uncommon in individuals younger than 50 years, the prevalence increases with age, and this disease is more common among men [9].

These parameters will influence the raw data and, consequently, the final result. A higher age will probably be associated with more severe symptoms, and on the contrary, an ON treatment state will have the opposite effect.

2.2.5.2 Technology

The sensors typically used to collect tremor data are accelerometers, gyroscopes, and sometimes, magnetometers. The accelerometer gives information on the linear acceleration along an axis. The gyroscope provides the angular velocity, meaning it measures how quickly an object turns around an axis. The magnetometer measures the magnetic field, which can interfere with the measurements of the other sensors and, thus may need to be removed [1]. The accelerometer and the gyroscope are accepted sensors for tremor quantification, however, the magnetometer is used as a support sensor. Alone the magnetometer can't quantify tremor, but when combined with other sensors it can provide information on the three-dimensional position of the measuring unit. This is especially useful during ADL, like writing or drawing [3]. Moreover, accelerometer data alone is sufficient to detect tremor, and simple data collection approaches are enough to monitor PD symptoms without compromising accuracy [7].

An Inertial measurement unit (IMU) is used in some publications, [10, 5]. These devices use a combination of the sensors previously mentioned to measure a body's specific force, angular rate, and occasionally, orientation.

From the studies included in this SLR, four, [11, 12, 19, 20], used only accelerometers. In [10], both an accelerometer and gyroscope were used, and in [21] and [23] the authors used these sensors and a magnetometer. Lastly, [22] and [5] both used an accelerometer and another sensor to collect data. Additionally, [5] also collected data from gyroscope sensors.

For [5], the authors used the Verily Study Watch, a wrist-worn wearable device, to collect data from an IMU, containing an accelerometer and a gyroscope, a photoplethysmography (PPG), and skin conductance sensors. The sampling rate of the IMU was 100 Hz, except for during the UPDRS or Parkinson's Disease Virtual Motor Exam (PD-VME) exams, where it was 200 Hz. The PD-VME was an exam performed by the subjects of set 2.

On the other hand, the authors of [22] used an accelerometer and electromyography (EMG), that records electrical activity in the muscles. These sensors were embedded into the NIMBLE patch. This patch is shown in figure 2.5 and is a flexible patch that attaches to the skin with an adhesive sticker.

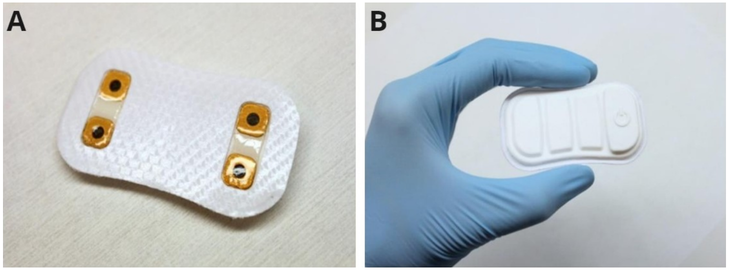


Figure 2.5: The NIMBLE patch.(A) Adhesive side that is placed on the skin; (B) Top of the patch. Adapted from [22].

Another option is using a smartphone or a smartwatch, since these devices contain accelerometers and gyroscopes, like in [12] where a smartphone was used and [19] where the authors opted to use a smartwatch. Moreover, in [12], the sensor data is collected through a mobile app. This app collects data from an accelerometer, linear accelerometer, gyroscope, and rotation vector sensors, even though the authors only use accelerometer data. Additionally, both these publications collected data at a 50 Hz sampling rate. In [23], the sensor unit was used with a sampling rate of 50 Hz, and in [11] the sampling rate was 100 Hz.

Furthermore, in [10] the authors implemented a wearable sensor glove, shown in figure 2.6, by placing an IMU on hand gloves. This was done to avoid friction and other errors that could affect the measurement. In [20], a Parkinson's KinetiGraph (PKG) was used to collect accelerometer data [24]. This wearable wrist-worn sensor system provides ratings of PD motor symptoms, like bradykinesia, dyskinesia, tremor, and fluctuations, as well as measuring sleep-related parameters and providing medication dosing reminders. Additionally, the authors of [21] used an Opal wearable sensor with a sampling rate of 128 Hz.

2.2.5.3 Experimental design

Due to the evaluation for the tremor associated with PD being mainly focused on the arms, the sensors were typically placed on the hands and wrists. Additionally, sometimes sensors were also placed on the forearm, like in [22], or ankle, like in [23].

In publications like [22], [11], [5], [19], and [23] the patient was asked to perform the specific tasks targeted at the assessment of each type of tremor. Those tasks were generally

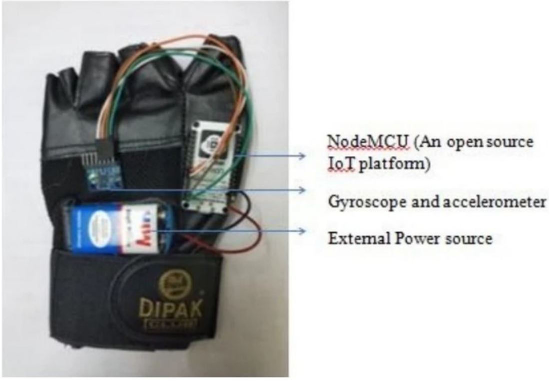


Figure 2.6: Wearable IMU sensor glove. Adapted from [10].

the same for every study and were based on the UPDRS tasks mentioned in section 2.1.

Besides the previously mentioned assessments, when being evaluated in a controlled environment, patients were also asked to perform ADL in [11]. Furthermore, evaluations in the clinic were usually either watched by a trained professional, that would perform the UPDRS classical evaluation or filmed so the same evaluation could later be performed.

When recordings were in the wild, meaning that data was collected outside of a controlled environment like the clinic, data was collected during free-living normal daily activities, [10, 11, 5, 20, 21], or the subjects were asked to perform specific tasks and record the data, like in [22]. Additionally, in [22] subjects were asked to register their medication, symptoms, and sleep in the patient's diaries.

For publications where recordings were in the wild, these recordings were mostly done continuously over several days. Contrarily, in [19] the evaluations are done weekly in a care center, and in [23] some patients did a follow-up evaluation, by repeating the same evaluation after between one and six months.

For the main study of [5], after their first clinic visit, sensor data was passively collected at home. The study duration was three years, with the last year being optional. Moreover, clinic visits were done yearly, after the first visit. In these visits, live clinical ratings and video ratings were performed. The final UPDRS Part III scores given were the median of in-person clinical rating and video rating. In addition, data was collected for a substudy, the PD-VME. In this substudy, participants performed the proposed task in the clinic, 1 hour after the UPDRS evaluation and before taking dopaminergic medication. At home, the participants performed the same tasks once a week, two times on the same day, either in the ON and OFF states of medication or one hour apart when not taking medication.

Lastly, in [12], data was collected through a mobile app. This mobile app features a ball game for quantifying patients' hand tremor, a medication log journal, and a daily survey to report the overall severity of symptoms. The game sessions lasted ten seconds and the patient placed the phone horizontally on the palm of the hand and tried to keep the ball inside a circle in the center of the screen, while accelerometer data was recorded.

2.2.5.4 Preprocessing

Typically in the preprocessing the signal can be, for example, filtered, segmented or down-sampled. Except for [12] and [23], every article that was reviewed filtered the signal. Despite being less frequent, only [10], [12], [21], and [23] did not segment the signal. Nonetheless, it is important to note that in [12], the data is already divided into 10 second intervals since that was the duration of the game sessions.

In regard to the filtering, most of the publications don't mention the type of filter used. In some of them, it is only mentioned that this step was done to remove the gravity component. However, in [11] and [19] the filter used was a 3^{rd} order Butterworth filter.

The two studies that mention the segmentation of the signals chose similar intervals of 2.56 seconds, in [19], and 3 seconds, in [11]. Even though, for [11], the windows had a 2 second overlap.

In [12], for each game session, the closest medication intake was identified and the data was labeled as "before" or "after" the intake. Furthermore, in [11] the laboratory recordings data were labeled according to the presence or absence of tremor. These labels were manually obtained by segmenting periods of time with or without tremor, using video recordings as a reference. Labels for the in-the-wild recordings, collected in this publication, were provided by the participants.

In [11], besides filtering and segmenting, the signal was also down-sampled to 50 Hz. Additionally, for some feature sets extracted, the Fourier transform (FT) module and tremor spectrum extraction were implemented, to enhance the tremor signal.

The authors of [23], first detected tremor using gyroscope signals, since this sensor is less affected by gravity, compared to the accelerometer, and by soft/hard iron distortions, compared to the magnetometer. The spectrogram of gyroscope signals in every axis was computed using short-time FT. Afterwards, a signal segment is considered for tremor quantification if there is no voluntary movement and tremor is present. A voluntary movement is considered as present when the maximum power frequency components below 3 Hz is higher than the maximum power of frequency components above 4 Hz by at least 15 dB. On the other hand, tremor is considered present when the opposite happens, with a difference between the maximums higher than 10 dB.

In the same study, limb displacement was calculated. The authors consider the limb displacement caused by the tremor as the tremor's amplitude. To calculate the limb displacement, first, the unit orientation concerning the earth frame is computed. The gyroscope is used to detect orientation changes and the accelerometer and magnetometer are used to compute correction parameters. The static acceleration is computed by projecting the gravity vector onto the sensor frame. The dynamic acceleration caused by tremor on each axis was calculated by subtracting the static acceleration from the accelerometer readings. Afterward, a high-pass filter with a cutoff frequency of 4 Hz is applied to the dynamic acceleration, and the velocity vector norm is calculated through a simple integration of the dynamic acceleration. Since the tremor is oscillatory, the authors consider it occurs in the time between two contiguous points of low velocity. Considering starting at rest with a velocity of zero, the arm starts accelerating until a maximum velocity and then returns to the velocity of zero by decelerating. For this study, it is considered that each tremor happens between the values of zero velocity on either side of the maximum velocity. Lastly, a single integration of the velocity for that window, representing limb displacement, was computed and tremor that were not in accordance with the tremor detection method previously mentioned were discarded.

In [19], the data was labeled based on the analysis of the magnitude in the tremor frequency band (3.5-7.5 Hz) and thresholds were empirically established to identify the presence of tremor. The resting periods for the context classifier were labeled manually, by annotating the intervals in the video recordings.

Afterward, the time segments when the participant is at rest and doesn't make movements with their hands are identified. This process is called context classification and is implemented to prevent false positives, due to other types of activities, in the detection of rest tremors. To validate the performance of the context classification, the Leave one subject out (LOSO) cross-validation is used. The data is segmented into 2.56 second windows with 50% overlap. These windows were labeled as resting or nonresting when more than 50% of the samples had that same label. The signal was then filtered using a 3rd order Butterworth band-pass filter with cut-off frequencies of 0.5 and 10 Hz. The authors implemented two approaches for this classification. The authors extracted 290 features, that include time and frequency characteristics. This data representation was then evaluated using an AdaBoost classifier with 100 estimators. Furthermore, as a second data representation, the FFT was obtained from the Euclidean Norm of the signals. In turn, this data representation was evaluated using AdaBoost and Gradient Boost classifiers. Then, for the second approach, a Convolution Neural Network (CNN) was evaluated for the classification with raw signals, since the CNN can automatically extract discriminating characteristics. For this evaluation, the data was filtered in the 0.5 to 10 Hz interval and normalized in a range from 0 to 1. The remaining data was divided into 80% for training and 20% for validation.

A tremor detector was also implemented to detect the presence of resting tremors. For this step, the signal was segmented into 2.56 second windows, where a window was labeled as tremor if more than 50% of its samples had that label. Furthermore, the signals were filtered between 0.5-10 Hz and 0.5-15 Hz. Afterward, 203 features, used in other studies, were extracted. These features and the raw accelerometer signal were used to evaluate the tremor detector. The features extracted from the signal filtered between 0.5-10 Hz were the symmetric part of FFT obtained from the Euclidean NORM, Mel frequency cepstral coefficients (MFCCs) adapted to inertial signals, root mean square (RMS), signal range, signal entropy, the ratio of the dominant frequency band to total energy, spectral flatness, and raw triaxial accelerometer signal. For the signal filtered between 0.5-15 Hz, the extracted features were the power in bands 4-6 Hz and 0.5-15 Hz, autocorrelation features, spectral entropy, 1st and 2nd dominant frequencies and magnitudes, cross-correlation between pairs of the axis. In addition, similarly to what was done in the context classifier, a CNN was evaluated for tremor detection using the raw signals as an input.

Moreover, multitask CNN capable of analyzing the context and occurrence of tremor simultaneously is proposed. This model can reduce the number of trainable parameters that would be necessary if, instead, two individual CNN models were used for the context classification and the tremor detection. This model was tested on the raw signals and FFT data representation. It is important to note that these steps are only used to select the windows where features will be extracted and, consequently were considered as part of the preprocessing and not described in debt in table 2.5.

2.2.5.5 Feature extraction

Commonly features are extracted from the time domain and the frequency domain. Time domain features can be extracted, for example, from the RMS and frequency domain features can be extracted through power spectral density (PSD) or FFT. Features extracted from the frequency domain can be the peak frequency, the mean frequency, the peak power, and the mean power. In turn, some examples of features extracted from the time domain are the mean amplitude, the average regularity, and the standard deviation. It is also common to extract the entropy, the correlation, and derivative features, like in [7]. The extracted entropy is typically the Spectral entropy.

In [5], the authors extracted features and evaluated rest tremor, upper-extremity bradykinesia, and arm swing during gait. However, the features enumerated in table 2.5, are only the features extracted to measure rest tremor severity.

In [19], analysis windows, where the context detector and the presence tremor identified that the patient was at rest and detected the presence of tremor, respectively, were identified. Then, the acceleration level (La) of the RMS amplitude in the tremor band between 3.5 and 7.5 Hz was calculated for each analysis window, in order to provide an amplitude indicator. The RMS values were calculated from the frequency spectrum, using Parseval's theorem. Afterward, the acceleration level was calculated according to the equation $La = 20 \log \frac{a}{a_0}$, where a is the RMS acceleration, in m/s^2, and a_0 is the reference acceleration of 1 μm/s^2.

The features used in [20] were provided by the PKG. Besides the numerical scores for tremor, the Percent Time Tremor (PTT), this wearable sensor system also provided the sleep time, Percent Time Immobile, the dyskinesia score (DKS), the bradykinesia score (BKS), and the fluctuating and dyskinesia score (FDS).

In [11], the authors implemented a CNN that was trained using the raw signal, similar to what the authors of [19] did with this neural network in the context classification and tremor detector. Furthermore, the common baseline features are also extracted from this signal. Some of those features were extracted from both the time and frequency domains, and others only from the frequency domain. For the frequency domain, the features were computed from the module of the FFT. The features extracted from these domains are enumerated in table 2.7. Moreover, these features were extracted from the X, Y, and Z axes, and the magnitude vector of acceleration and jerk. Additionally, the correlation coefficient between every pair of axes for the acceleration, jerk, FFT of acceleration, and FFT of jerk were computed.

Table 2.7: Common baseline features extracted from the time and frequency domains in [11].

Domain	**Features**
Time and Frequency domains	Mean, standard deviation, median, max, min, signal magnitude area, energy, inter-quartile range, empirical cumulative distribution function, entropy, and auto-regression coefficients.
Frequency domain	Dominant frequency, average frequency, spectral power, skewness, kurtosis, and energy of six equally spaced frequency bands.

Furthermore, in [11], the energy in the 3-9 Hz frequency band and the MFCCs were computed from the FT module. Lastly, the MFCCs were computed from the tremor spectrum and a CNN was trained on spectra. The CNN, like the one trained on raw data, learned features from the training on spectra, after the extraction of the tremor spectrum.

The authors of [12] analyzed the signals exclusively in the frequency domain. The Welch method was used to generate a non-parametric estimation of the PSD of every patient's game sessions. In this study, a tremor intensity parameter (tip), calculated as pv divided by SF50, is proposed and extracted from the signals.

2.2.5.6 Data analysis

Regarding the methods used for the analysis of the extracted features, five publications, [22, 10, 12, 5, 20], applied statistical analysis using the methods shown in tables 2.5 and 2.6. Another two publications, [11, 19], used classification algorithms to classify this data. Lastly, in [21] the authors implemented their own tremor detection method, and in [23] a Fuzzy inference system is implemented.

In [10], the UPDRS scores were assigned based on the vibration frequency, position, and occurrence rate. These resting tremor scores, besides being found by using the sensor glove data from ten days, were also estimated with only one day. In the last case, the occurrence rate was not used. Afterward, a statistical analysis was performed to compare the results from 1 and 10 days with the clinically given scores.

The authors of [5] concluded that the lateral tremor acceleration measurement showed the strongest correlation to the UPDRS ratings and the strongest ability to discriminate between the ON and OFF medication states. Thus, the authors opted to only discuss the statistical analysis results for this measurement.

In [19], after obtaining the acceleration level for all the analysis windows, the 75th percentile was obtained. This percentile was used to simulate the clinical evaluation using the UPDRS scale, since in this evaluation the amplitude of tremor is considered the highest amplitude identified during the evaluation.

Two different classifiers were used in [11], those were fully supervised learning and weakly supervised learning classifiers. The data collected in the lab was used to train two classification algorithms for fully supervised learning. The algorithms used were random forest (RF) and Multi-Layer Perceptron (MLP). For weakly supervised learning a multiple-instance learning (MIL) algorithm, proposed by another author, was trained on data collected in the wild. To initialize the MIL algorithm, the 3 second windows in a five minute interval are sorted according to their estimated energy of the power spectrum. Afterwards, the class of every three second window is predicted. These predictions can be thought of as pseudo-probabilities. Then, the windows in the interval are sorted according to the probabilities and a subset of the windows are selected for retraining, according to the weak label attributed.

2.2.6 Conclusion

From this SLR it can be concluded that the methods used tend to vary for most parts of the selected studies. The variances between studies are particularly evident in the computation framework.

The data, technology, and experimental design are similar for the 9 studies selected for the SLR, as it can be seen in tables 2.3 and 2.4. Regarding the data, the number of patients varies according to the dimension of the study, and the number of patients from each sex varies accordingly, with the ages being similar. The technology used is similar, with all studies using accelerometers and five studies implementing the use of gyroscopes, magnetometers, or other sensors. For the experimental design, the publications can generally be divided into three settings for the data collection. Those being in the clinic while recreating the movements used for the evaluation of the different types of tremor in the UPDRS, while recreating ADL, or during normal daily activities. The information presented in these tables provides an overview of the methods typically used in the data collection, for the experimental design and technology, and the characteristics of each dataset. This information was essential for the idealization of the mobile app proposed in chapter 3 and the implementation of future data collections.

The computation framework shown in tables 2.5 and 2.6 is divided into preprocessing, feature extraction, and classification. In the preprocessing, it is common to implement downsampling, filtering, and/or segmentation steps. No common features were extracted between the studies selected, however, this might be due low number of studies that were selected for this review. The inclusion of more studies through the adjustment of the exclusion criteria could help establish common features between the studies. For the analysis of the data, the methods used were mostly statistical analysis or classification algorithms. The information extracted from these studies was important in the development of the computation framework.

Chapter 3

Mobile App Design

In this chapter, a mobile app for the collection of data from PD patients is proposed. A schematic representation of the data collected is presented in figure 3.1. The data, collected through the mobile app to evaluate the tremor associated with PD, can be divided into two main types of data. Those are data from sensors and data from questionnaires, described in sections 3.1 and 3.2, respectively. Lastly, section 3.3 contains figures and a description of the mobile app implementation.

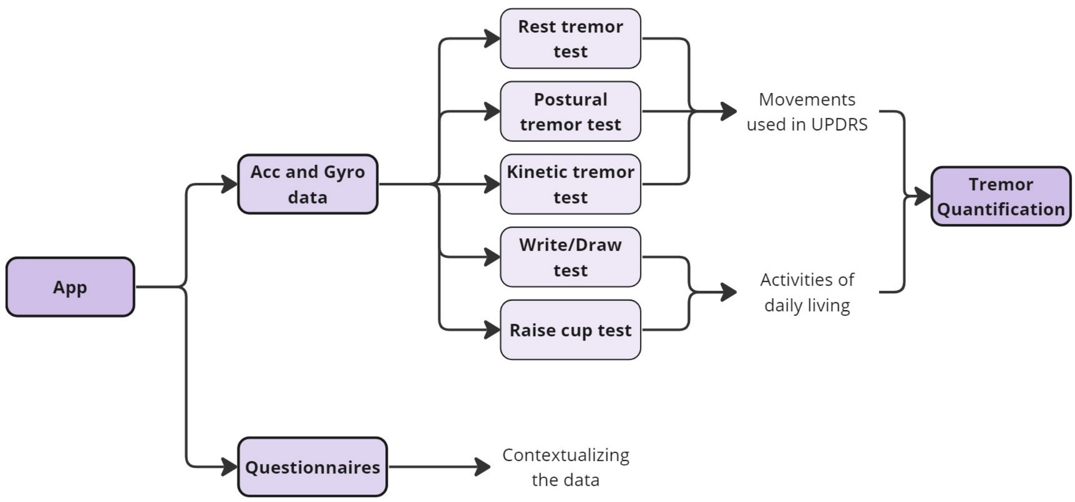


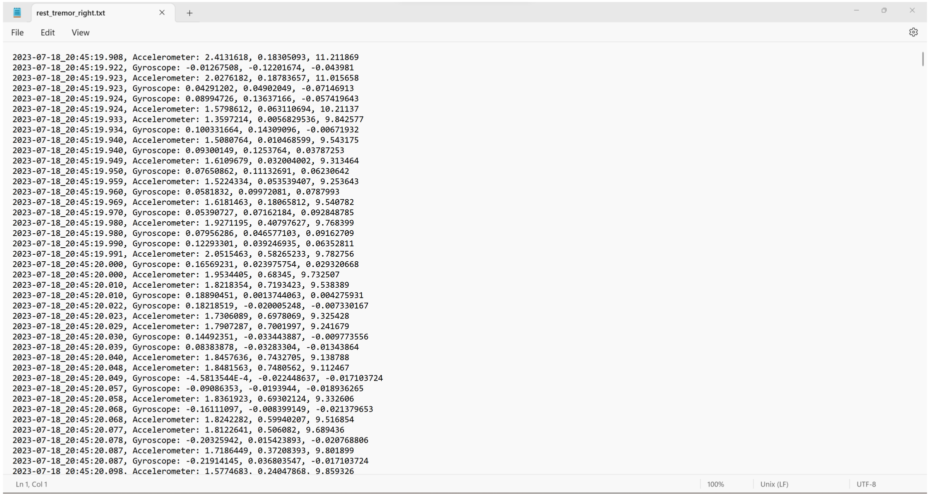
Figure 3.1: Schematic representation of the data collected with the app.

3.1 Sensor Data

Sensor data is collected from the accelerometer and gyroscope sensors in the smartphone, at a sampling rate of 50 Hz. This sampling rate was chosen due to being used in multiple studies and considered appropriate for the analysis of tremor in the upper extremities [11, 12, 19]. Moreover, the recorded data is saved in text files, with the date and time of recording, as shown in figure 3.2. Each line in the file corresponds to one of the sensors and contains the name of that sensor and the values for the X, Y, and Z axes, in that order.

Data is collected while the subject performs five tests, each one with a different movement of the arm. These tests are a rest tremor test, a postural tremor test, a kinetic tremor test, a writhing/drawing test, and a raising a cup test.

The first three tests are based on the movements the patient performs during the UPDRS


rest_tremor_right.txt

File Edit View

```
2023-07-18_20:45:19.908, Accelerometer: 2.4131618, 0.18305093, 11.211869
2023-07-18_20:45:19.922, Gyroscope: -0.01267508, -0.12201674, -0.043981
2023-07-18_20:45:19.923, Accelerometer: 2.0276182, 0.18783657, 11.015658
2023-07-18_20:45:19.923, Gyroscope: 0.04291202, 0.04902049, -0.07146913
2023-07-18_20:45:19.924, Gyroscope: 0.08994726, 0.13637166, -0.057419643
2023-07-18_20:45:19.924, Accelerometer: 1.5798612, 0.063110694, 10.21137
2023-07-18_20:45:19.933, Accelerometer: 1.3597214, 0.0056829536, 9.842577
2023-07-18_20:45:19.934, Gyroscope: 0.100331664, 0.14309096, -0.00671932
2023-07-18_20:45:19.940, Accelerometer: 1.5080764, 0.010468599, 9.543175
2023-07-18_20:45:19.940, Gyroscope: 0.09300149, 0.1253764, 0.03787253
2023-07-18_20:45:19.949, Accelerometer: 1.6109679, 0.032004002, 9.313464
2023-07-18_20:45:19.950, Gyroscope: 0.07650862, 0.11132691, 0.06230642
2023-07-18_20:45:19.959, Accelerometer: 1.5224334, 0.053539407, 9.253643
2023-07-18_20:45:19.960, Gyroscope: 0.0581832, 0.09972081, 0.0787993
2023-07-18_20:45:19.969, Accelerometer: 1.6181463, 0.18065812, 9.540782
2023-07-18_20:45:19.970, Gyroscope: 0.05390727, 0.07162184, 0.092848785
2023-07-18_20:45:19.980, Accelerometer: 1.9271195, 0.40797627, 9.768399
2023-07-18_20:45:19.980, Gyroscope: 0.07956286, 0.046577103, 0.09162709
2023-07-18_20:45:19.990, Gyroscope: 0.12293301, 0.039246935, 0.06352811
2023-07-18_20:45:19.991, Accelerometer: 2.0515463, 0.58265233, 9.782756
2023-07-18_20:45:20.000, Gyroscope: 0.16569231, 0.023975754, 0.029320668
2023-07-18_20:45:20.000, Accelerometer: 1.9534405, 0.68345, 9.732507
2023-07-18_20:45:20.010, Accelerometer: 1.8218354, 0.7193423, 9.538389
2023-07-18_20:45:20.010, Gyroscope: 0.18890451, 0.0013744063, 0.004275931
2023-07-18_20:45:20.022, Gyroscope: 0.18218519, -0.020005248, -0.007330167
2023-07-18_20:45:20.023, Accelerometer: 1.7306089, 0.6978069, 9.325428
2023-07-18_20:45:20.029, Accelerometer: 1.7907287, 0.7001997, 9.241679
2023-07-18_20:45:20.030, Gyroscope: 0.14492351, -0.033443887, -0.009773556
2023-07-18_20:45:20.039, Gyroscope: 0.08383878, -0.03283304, -0.01343864
2023-07-18_20:45:20.040, Accelerometer: 1.8457636, 0.7432705, 9.138788
2023-07-18_20:45:20.048, Accelerometer: 1.8481563, 0.7480562, 9.112467
2023-07-18_20:45:20.049, Gyroscope: -4.5813544E-4, -0.022448637, -0.017103724
2023-07-18_20:45:20.057, Gyroscope: -0.09086353, -0.0193944, -0.018936265
2023-07-18_20:45:20.058, Accelerometer: 1.8361923, 0.69302124, 9.332606
2023-07-18_20:45:20.068, Gyroscope: -0.16111097, -0.008399149, -0.021379653
2023-07-18_20:45:20.068, Accelerometer: 1.8242282, 0.59940207, 9.516854
2023-07-18_20:45:20.077, Accelerometer: 1.8122641, 0.506082, 9.689436
2023-07-18_20:45:20.078, Gyroscope: -0.20325942, 0.015423893, -0.020768806
2023-07-18_20:45:20.087, Accelerometer: 1.7186449, 0.37208393, 9.801899
2023-07-18_20:45:20.087, Gyroscope: -0.21914145, 0.036803547, -0.017103724
2023-07-18_20:45:20.098, Accelerometer: 1.5774683, 0.24047868, 9.859326
```

Ln 1, Col 1 100% Unix (LF) UTF-8

Figure 3.2: Example text file with the data collected for the rest tremor test of the right arm using the mobile app.

evaluation for these types of tremor, shown in figure 2.2, that help isolate the type of tremor being evaluated. For the rest tremor test the subject rests the hand of the side being evaluated on their leg, with the palm facing up. For the postural tremor test, the subject holds their arm up against gravity. Lastly, for the kinetic tremor test, the subject points their finger and moves the arm from the postural tremor test position to their nose, repeating this movement during the data collection.

The writing/drawing and the raising cup tests intend to recreate the movements of ADL. For the writing/drawing test the subject freely draws or writes and for the raising a cup test the subject either raises a cup or recreates that movement while holding the smartphone.

While the sensor data is recorded, the patient performs each test for thirty seconds, followed by an unrecorded fifteen second break. Data is collected from both upper extremities three times a day, one in the morning, one around lunch, and one at night. Each of the three data collections can be done at any time during a defined interval. The morning data collection can be done between 8:00 a.m. and 11:00 a.m., the lunch data collection can be between 12:00 p.m. and 3:00 p.m., and the night collection can be performed between 8:00 p.m. and 11:00 p.m.

3.2 Questionnaires

The data from the questionnaires was collected at the same times of the day as the sensor data. The questions that the subject is asked vary depending on the time the questionnaire takes place.

For the morning questionnaire, the patient is asked about their sleep on the previous night and how rested they are. The questions in the lunch questionnaire regard where, with whom, and what the patient did during the morning. As well as questions regarding possible difficulties that may be related to PD, like walking, standing, or speaking. The patient is also asked about how they experienced symptoms, like tremor, and about their medication. Lastly, in the night questionnaire, the patient answers the same questions about activities, difficulties, symptoms, and medication that were asked at lunch, but for the afternoon/evening period. Furthermore, the patient is asked how he felt during the day if he had any particular difficulty in any daily activity, and if any substances that could affect the signal were consumed. Thus these questions are important to contextualize the sensor data used to evaluate tremor. Additionally, these questions provide context and information for, not only tremor but other symptoms as well, giving information that, in the future, can be applied to the evaluation of PD in other perspectives. The full questionnaires are in appendix A.1, in Portuguese. It is also important to note that these questions are loosely inspired by the questions asked in the dataset collected in [25].

The questions can contain multiple choice answers, like in the first question below, or 7 point Likert scale answers, where the patient chooses the point of the scale he feels better represents the correct answer, like in the second question below. Both the questions presented here belong to the Lunch questionnaire.

1. Where were you during the morning?
    - Home
    - Work
    - Travel
    - House of friends/family
    - In public
    - Other
2. I experienced tremor.
    - Strongly disagree
    - Disagree
    - Somewhat disagree
    - Neutral
    - Somewhat agree
    - Agree
    - Strongly agree

The Likert scale was developed in 1932 by Likert. This scale is made up of statements for a real or hypothetical situation to which the subject is asked to show their level of agreement on a metric scale. The agreement ranges from strongly disagree to strongly agree and the number of possible answers, typically called points, can vary [26].

The Likert scale can have multiple numbers of points, even though the most common are the 5 point and 7 point Likert scales. Research has shown that the 7 point scale may perform better than the 5 point scale [26].

When the 5 point Likert scale is applied by an administrator or in paper the subject, sometimes, tries to choose between two points in the scale, like trying to choose a value of 3.5 when neither 3 or 4 correctly represent their intended answer. This may lead to rounding the value or to the addition of decimal values by the administrator, leading to a loss of information due to interpolation. When administered electronically the subjects are obligated to conform to an item that does not reflect their true response [27]. Furthermore, when obligated to choose an item that does not fully reflect their answer on repeated administration of this scale, the chosen item tends to vary, even if the subject's answer hasn't changed [26]. All of this reflects the insensitivity associated with the 5 point Likert scale[26, 27].

On the other hand, the 7 point scale has more options, increasing the probability that one of the points correctly reflects the subject's intended answer, which might eliminate the problem of having to choose between two undesirable answers. This reduces the ambiguity in the answers [26].

Another study found that the increase in the number of points leads to an increase in reliability. However, this increase begins to plateau at around 7 points and there is no significant increase in reliability above 11 points. Moreover, in one publication, the authors studied Likert scales of 2 point, 5 point, 7 point, 9 point, 11 point, 12 point and 100 point (percentage). They found that the 7 point scale presented the best accuracy and usability. The authors also found that the 5 point scale performed worse in all criteria when compared to the 7 point scale [27].

In conclusion, a 7 point Likert scale is more likely to reflect the respondent's true subjective answer. This number of points seems to be sensitive enough to minimize interpolations and compact enough to efficiently be answered. For these reasons, a 7 point Likert scale was chosen for the questionnaires.

3.3 Implementation

The mobile app and the database where the collected data is stored were implemented by a student of a bachelor's degree in Computer Science and Engineering at the University of Beira Interior for the project curricular unit. Furthermore, the student also helped idealize some of the features of this mobile app.

Firstly, on the implemented mobile app, the patient signs up, defining a password and an email address. Whenever the subjects open the app they log in and, afterward the page shown in figure 3.3 on the left appears. On this page, the patient selects the questionnaire they want to answer. This leads to the page shown in fig 3.3 on the right, containing the questions of the selected questionnaire.

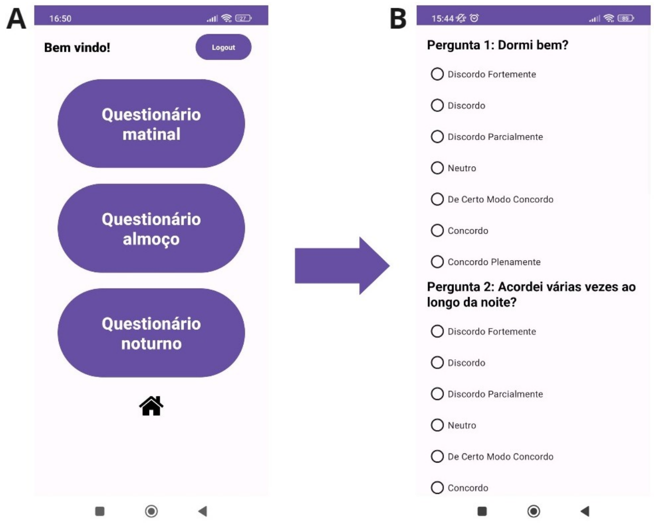


Figure 3.3: (A) Page in the mobile app where the subject selects the questionnaire they want to answer; (B) Following page where the selected questionnaire is answered.

The list of tests is shown after the questionnaire is answered. Additionally, there may be a need to access the tests without performing the questionnaire and, since initially, the plan is to use the app to collect data with the presence and assistance of someone involved in the study, the test list can be accessed using the home button on the page where the subject selects the questionnaire. Similarly, on the list of tests, shown in figure 3.4, the questionnaires can be accessed using the home button.

When using the app, on the list of tests the subject can select the test they want to perform. After selecting the test they choose the arm from which the data will be collected. Lastly, they start the recording and perform the respective movement for the selected test. The interface implemented for this process is shown in figure 3.4.

Data collected for the same test or questionnaire on different days or at different times of the same day will be saved on the same file as data previously collected for that test or questionnaire. Thus the timestamps are important to differentiate the time of day the data belongs to. Furthermore, when the test is not performed correctly and needs to be repeated it is important to take note of this occurrence since the timestamps from both collections are around the same time. Then, with the indication of the time at which the correct recording was initiated, the incorrectly collected data can be discarded from the file without losing information or including incorrect data in the dataset.

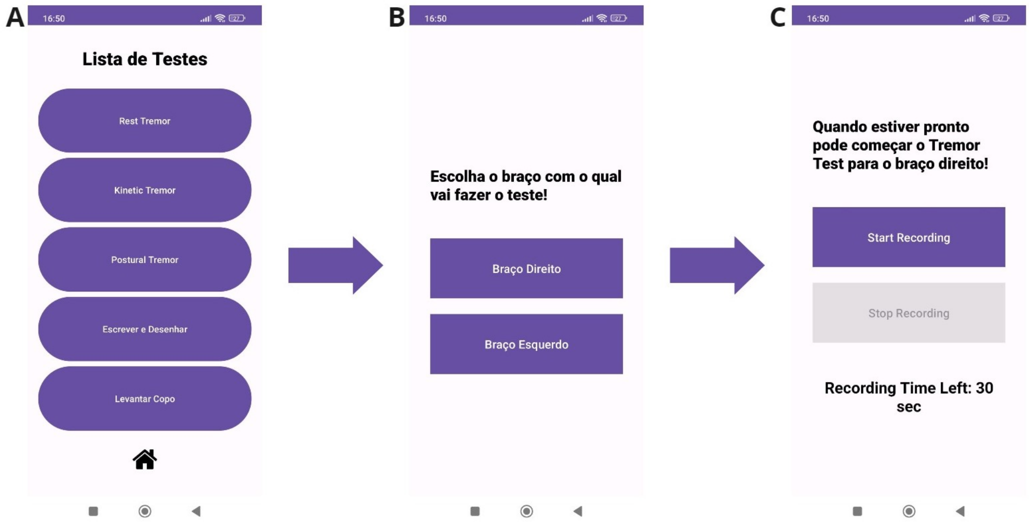


Figure 3.4: (A) Page in the mobile app where the subject selects the test they want to perform; (B) Following page where the subject selects the arm that they will use to perform the test; (C) Following page where the test can be initiated and the countdown until the end of the test is displayed.

3.4 Context

This mobile app, despite not being used for data collection in this thesis, can be implemented for the collection in the future.

Furthermore, the data collection using this mobile app from subjects without PD disease was approved by the ethics commission of the University of Beira Interior. This data hasn't been collected but will be in the future. The data will be collected from subjects without PD in different retirement homes. This data will be used as a control dataset to test differences between PD patients and people without PD. Furthermore, the collection of data using this mobile app will help test its applicability.

In the future, it would be beneficial to use this app to collect data from PD patients. However, this was not proposed and implemented for this thesis due to bigger limitations. Moreover, even though the app as it is proposed here collects data in isolated instances, a sensor device could be used with and connected to the app to collect data continuously during the daily activities of the patients.

Chapter 4

Computation Framework

In this chapter, a computation framework to evaluate the tremor associated with PD is proposed. The framework is schematized in figure 4.1. First, the dataset used is explored. Then the preprocessing and feature extracting steps are examined. Lastly, for the differentiation between PwPD and HC and the estimation of UPDRS tremor scores classifiers and methods based on the extracted features were implemented. Moreover, a short analysis of the periodograms resulting from the application of Welch's method was done. This computation framework was implemented in Python.

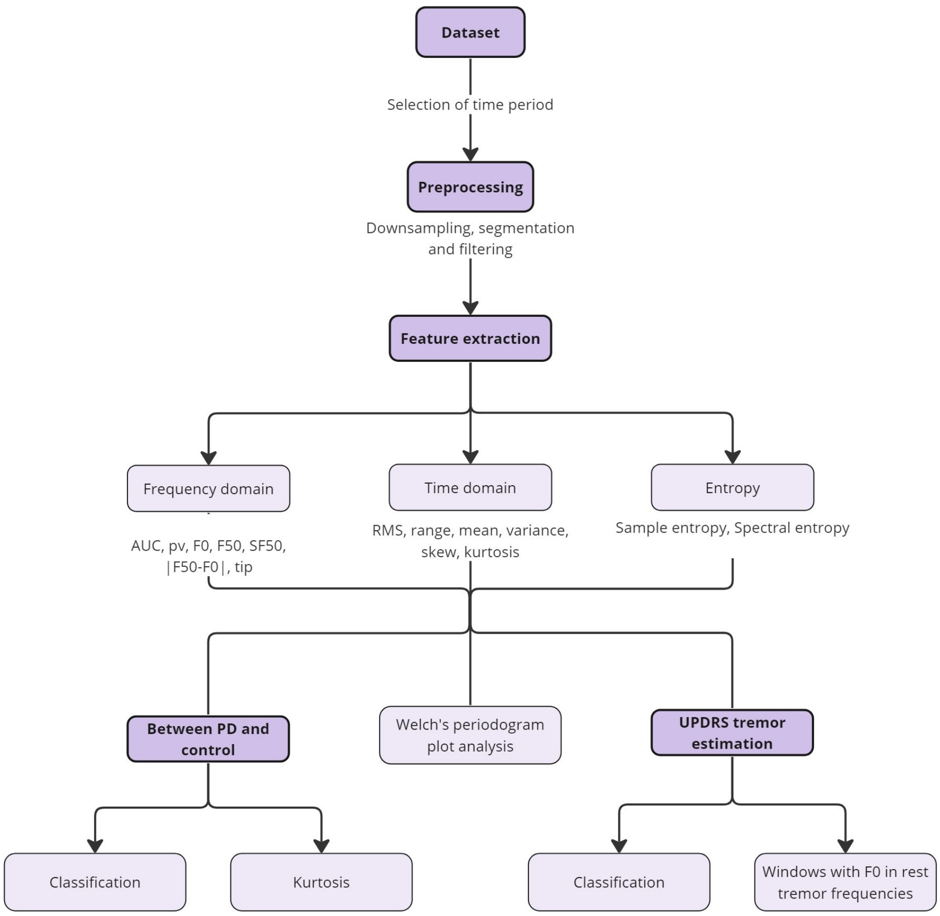


Figure 4.1: Schematic representation of the proposed computation framework.

4.1 Dataset

There is a lack of established protocols for the collection of tremor data from PwPD. Therefore, it is common for authors to collect their datasets according to their research needs. The collection protocols for these datasets vary and consequently, the datasets and ideal frameworks for the assessment of tremor associated with PD also tend to vary. Furthermore, not all of these datasets are ideal for each type of study and most of them are not open source. Therefore, ideally the dataset used should be collected specifically for the study, in order to control the collection conditions. Due to the lack of time and resources needed to collect a dataset of tremor data from PwPD, two available datasets were explored for the implementation of this work.

The first dataset explored is provided by the authors of [25], who also provide the code used in their computation framework. This dataset includes data from 20 PD patients, monitored during their daily lives with accelerometers and gyroscopes. Each pair of inertial sensors was placed on the chest, left or right wrists of the patient, and data was collected continuously throughout 14 days at a 200 Hz sampling frequency. During that time patients also used subjective electronic diaries, with the aim of capturing motor and non-motor symptom fluctuations and contextual information during the day. The data from the electronic diaries is collected through three types of questionnaires. The first is a "beep" questionnaire, available seven times a day with an accompanying notification, the "beep" sound. The second is the daily morning questionnaire, which the patient can complete between 6:00 a.m. and 12:00 p.m. The third is a daily evening questionnaire that can be completed by the patient between 8:00 p.m. and 3:00 a.m. This dataset does not contain data collected from healthy controls or UPDRS evaluation scores.

The second dataset, collected by the authors of [6], contains accelerometer data from 17 PD patients and 17 healthy controls. The data was acquired using 5 lightweight MC 10 BioStamp RC sensors placed on the trunk, left and right anterior thighs, and left and right anterior forearms. Data was collected for around two days at a sampling rate of 31.25 Hz. Moreover, the recording was initiated during a UPDRS evaluation by a trained professional, and annotations from this evaluation are provided. Therefore, the beginning of the data files were not collected in the wild. The dataset also contains demographic and clinical assessment data, including the evaluation for rest tremor.

In this thesis, the second dataset was used in the computation framework. This choice was made because the first dataset did not contain data from HC or the UPDRS evaluations on tremor. Furthermore, the only value attributed to the intensity of tremor in the first dataset is in the form of the question "I experienced tremor" in the patient questionnaires. This question is answered on a 7 point Likert scale and thus is, not only subjective to the patient but also does not correlate to the typically attributed UPDRS scores for tremor. The second database, due to having data from HC controls and UPDRS scores for rest tremor implementation of methods to distinguish between PwPD and HC, and to estimate the rest tremor scores.

The BioStamp RC device used to collect this dataset is not described by the authors. However, a different publication by different authors, [28], used this device and determined that it effectively captures limb motion. In this publication, the BioStamp Research Connect (BioStamp RC), shown in figure 4.2, is described as a wireless, comfortable, thin, skin-adherent, wearable-sensor patch containing accelerometer and gyroscope elements. A software application on a mobile device allows the user to configure and control the device. In addition to the accelerometer and gyroscope, the device also contains an electrocardiogram (ECG) and EMG sensors, and the data can be collected by a singular sensor or several sensors at the same time.

Figure 4.2: Top and bottom views of the BioStamp RC. Adapted from [28].

For this study, only data from the anterior forearm sensors was used. It is also important to note that individuals with ID 7 and 60 only had data files from the sensor on the right forearm. The annotations from the evaluations were discarded, because these annotations were not clear in terms of when each task started and finished being evaluated. Since the tasks performed during the UPDRS evaluation tend to make the type of tremor being evaluated the dominant type of tremor, the lack of clear annotation could lead to an overlap in the evaluation of the tremor intended to be evaluated with the evaluation tasks performed for other tremor types. For these reasons, this study was focused on the evaluation of rest tremor during free living. The rest tremor was the only evaluated due to the UPDRS scores in the dataset being only for this tremor type. Furthermore, the clinical assessment was performed during the ON and OFF states of the medication. The clinical data of each individual is present in table 4.1.

A time interval of four hours from each of the files containing data from each forearm of the subjects was selected. The selected interval starts 5 minutes after the indicated end of the evaluation. This way all data collected during the UPDRS tasks was discarded, giving a 5 minute time margin to guarantee no data collected while performing the collection tasks is included. This is possible by using the last timestamp of the clinical evaluation annotations. Since the evaluation ended on the ON medication state, the patient is considered to be in that state and used the rest tremor scores given for that same state of medication.

That is also the state for the rest tremor scores in table 4.1.

Table 4.1: Demographic and clinical data of individuals in the dataset.

ID	Group	Gender	Age	RT score	
				left	right
5	Control	F	74	0	0
6	PD	M	73	2	0
7	Control	F	52	0	0
8	Control	F	77	0	0
10	PD	F	72	0	2
12	PD	F	64	1	1
13	PD	F	60	0	0
14	Control	F	56	0	0
15	PD	M	65	0	1
16	Control	F	62	0	0
17	PD	M	74	0	3
18	Control	F	66	0	0
20	Control	F	68	0	0
22	Control	M	68	0	0
23	PD	F	68	0	2
24	PD	M	62	0	0
25	PD	F	72	0	1
27	Control	F	54	0	0
30	Control	F	68	0	0
33	PD	M	46	0	0
35	PD	M	67	2	0
36	PD	M	69	3	0
38	PD	M	78	1	0
39	Control	F	74	0	0
40	PD	F	75	1	0
41	Control	M	75	0	0
42	PD	M	84	0	0
43	Control	F	69	0	0
44	PD	F	63	0	0
45	Control	M	64	0	0
58	Control	M	39	0	0
60	Control	F	65	0	0
62	Control	F	56	0	0
63	PD	M	37	0	1

4.2 Preprocessing

Firstly, the magnitude of the accelerometer signal for the selected 4 hour time period is calculated with the square root of the sum of squares of each axis, as shown in equation 4.1, using the same equation as the authors of [7].

$$\text{Magnitude} = \sqrt{x^2 + y^2 + z^2} \quad (4.1)$$

According to [29], a sampling frequency of 100 Hz is sufficient to measure motor features related to PD, since the frequency of tremor in the upper extremities is lower than 13 Hz. Some authors opted to use higher sampling frequencies instead, like in some of the studies explored in the SLR, in section 2.2. However, many studies consider a sampling frequency of 50 Hz adequate for the detection of human activity with these sensors [11, 12, 19]. A downsampling step was sometimes implemented in the literature to reduce the sampling frequency to 50 or 100 Hz in the cases where the sampling frequency of the dataset used is higher than these values. Since the sampling frequency of the dataset is 31.25 Hz, there was no need to implement a downsampling step.

Afterward, a 4^{th} order high pass Butterworth filter with a cutoff frequency of 0.5 Hz was used to remove the effects of gravitational acceleration and low-frequency noise from the axes and the magnitude vector. This cut-off frequency is the same as in [7], where the features extracted from the time domain are also the same as in this thesis. Despite this publication not specifying the type of filter used, from the literature review it was concluded that the most common filter for the processing of accelerometer signals to evaluate tremor was a Butterworth filter, varying in order.

Lastly, the axes and magnitude vector of the 4 hour signal were segmented into 10 second windows, with no overlap. This interval was chosen, instead of a smaller one, to allow for the averaging of the periodograms of segments from these windows, performed on the feature extraction step for the Welch's method. The lack of overlap between windows is also important for the correct implementation of the Welch's method, due to the averaging of periodograms, mentioned before, which is typical of this method.

4.3 Feature Extraction

The features extracted from each window of the signals axes and magnitude vector were from both frequency and time domains and are shown in table 4.2. The extraction of the sample entropy and spectral entropy was implemented but these features were not extracted due to greatly increasing the computation time.

Table 4.2: Features extracted from the frequency and time domains.

Domain	**Features**
Frequency	AUC, pv, F0, F50, SF50, $\|F50 - F0\|$, and tip
Time	RMS, range, mean, variance, skewness, and kurtosis

The frequency domain features were extracted from the PSD, estimated using the Welch's method. This method determines the power contained in the signal's frequency components [12]. For that, the data was divided into four 2.5 second segments with 50% overlap, with the periodogram for each segment being calculated. Lastly, the periodograms in each 10 second window are averaged [30]. The periodograms are briefly explored in section 5.2.

The features extracted from the averaged periodogram were the area under the curve (AUC) between the frequencies of 3 and 6 Hz, peak value (pv), fundamental frequency (F0), central frequency (F50), frequency dispersion (SF50) and the absolute value of the difference between F50 and F0 ($|F50 - F0|$), like in [12]. Additionally, the tremor intensity parameter (tip), suggested in the same publication, was also calculated. The AUC was calculated with the scikit-learn library, which computes the AUC using the trapezoidal rule, like in [4]. Additionally, this feature was calculated between the frequencies of 3 and 6 Hz to determine the power of the signal in the interval corresponding to the rest tremor frequencies. The features extracted from the frequency domain are described in table 4.3.

Table 4.3: Features extracted from the periodograms and theirs description [12].

Feature	**Description**
AUC	Total power of the signal.
pv	Maximum value of the PSD.
F0	Frequency of maximum power. Can be used to determine the dominant tremor type in a window, according to the type of tremor frequency band it belongs to.
F50	Central point where the periodograms are divided into two equal parts.
SF50	Width of the frequency band centered in F50 that contains 68% of the total power.
$\|F50 - F0\|$	Absolute value of the difference between F50 and F0.
tip	Calculated by dividing pv by SF50.

In regards to the time domain, the features extracted were the same as in [7]. Those being, the root mean square (RMS), range, mean, variance, skewness, and kurtosis. From these features, the skewness measures the lack of symmetry and the kurtosis describes the length of the peak [31]. The authors of this publication also provided their code, which was used as a reference for the implementation.

4.4 Distinction between PD and HC

4.4.1 Classification

In this step, a classifier was implemented to differentiate between the PwPD group and the control group. Five classifiers were tested for this task and these were decision tree (DT), support vector machine (SVM), k-nearest neighbors (kNN), multi-layer perceptron (MLP), and Bagging Tree. All the classifiers used were supervised learning methods and capable of both binary and multiclass classification. The DT classifier is a model that predicts a target variable with decision rules that it has learned from the data features, this is accomplished by grouping samples with the same labels [32]. For the SVM a hyper plane or a set of hyper-planes is constructed and achieves a better separation between classes the larger the functional margin. This margin is the distance to the nearest training points of any class and a larger functional margin is associated with a lower generalization error for the classifier. The samples on the margin boundaries, also called support vectors, can be within the margin boundaries when the data can not be linearly separated [33].

The kNN classifier is a non-generalized learning method, as such it stores instances of the training data instead of constructing a general internal model. This classifier predicts a label for a new sample based on the labels of a predefined number of samples, k, closest in distance to the new sample [34]. The MLP classifier is a type of neural network that, given a set of features and a target, learns a function approximator for the classification. This model contains an input layer, consisting of a set of neurons that represent the input features, an output layer, that transforms the values from the last layer into the output values, and one or more layers in between, also called hidden layers. Each neuron of each layer receives the value from all the neurons in the last layer and sends that processed value to all the neurons in the next layer. Between each layer, the value is transformed with a weighted summation, a bias, and an activation function [35]. A Bagging Tree classifier is an ensemble method, these methods combine the predictions of several base estimators to improve the generalizability and robustness of a single estimator. The Bagging Tree classifier trains several decision tree classifiers with random subsets of the original training dataset and then forms a final prediction by averaging their predictions. This method reduces overfitting and the variance of the base estimator, through the introduction of randomization [36].

The parameters used for each classifier are described in table 4.4 and the library used for the implementation was scikit-learn [37]. The classifiers were tested using 80% of the dataset for training and 20% for testing. The classifier with the best recall for the control and PD groups was chosen.

Table 4.4: Parametrization used for the tested classifiers.

Classifiers	**Classifier type**	**Classifier parameter**
DT	sklearn.tree.DecisionTreeClassifier	criterion = 'gini', max_leaf_nodes = 7, max_features = "auto"
SVM	sklearn.svm.SVC	kernel = 'rbf', C = 1
kNN	sklearn.neighbors.KNeighborsClassifier	n_neighbors = 3
MLP	sklearn.neural_network.MLPClassifier	solver = "sgd", alpha = 1, max_iter = 1000, learning_rate = 'adaptive'
Bagging Tree	sklearn.ensemble.BaggingClassifier	estimator = DecisionTreeClassifier(), n_estimators = 30, max_features = 39, random_state = 5

These classifiers used 39 features, which were 13 features from each of the X, Y, and Z axes. Afterwards, the classifier with the best results was chosen and was implemented using the same 39 features. The features from the magnitude, despite being extracted, were left out. Furthermore, the leave one patient out validation was used, leaving one patient out of the training, and then applying that classifier to that same patient, obtaining the result for that patient. This process was repeated for each patient until the results were obtained for all the patients.

The classifier returns the percentage of windows classified as PD for each person in the dataset. A percentage of 50% was considered inconclusive, a percentage under 50% was considered non-PD, and a percentage above 50 % was considered as PD. The classification was implemented for both forearms together and individually. A patient was classified as PD when one of the forearms was classified as PD.

4.4.2 Kurtosis

The data of all windows from each feature of each axis was either summed or averaged for all patients. For the F0, F50, SF50, and $|F50 - F0|$ features, the values were averaged, since these features represent frequencies and would lose their significance if the values were summed. For the rest of the features, due to their small values for each window, the values were summed. Afterward, the correlations between the resulting values for each feature in each axis, the rest tremor scores, and the distinction between individuals with PD and healthy controls were calculated. From this, it was determined that the feature with the strongest correlation to the distinction between individuals with PD and HC was the kurtosis on the Y axis.

Afterwards, the mean and standard deviation for that feature were calculated for both groups of the dataset and the values obtained were used to define the typical intervals for the PD and control groups. The sum value that was calculated for the kurtosis on the Y axis was classified as part of the PD or HC groups, according to those intervals. If either forearm of a patient was classified as PD, the patient was also classified as PD. The prediction for any value outside the typical intervals is considered inconclusive.

4.5 UPDRS rest tremor score estimation

4.5.1 Classification

For the prediction of the tremor scores using a classifier, the dataset was first balanced due to the low number of samples with UPDRS scores 1, 2, and 3 when compared with score 0. Since no samples of the data had a score of 4, this score was excluded from the classification.

The data was sampled using the Synthetic Minority Over-sampling Technique (SMOTE). In this technique, the minority class is over-sampled through the creation of synthetic samples based on a predefined number of nearest neighbors, k, of the same class [16, 38]. The SMOTE was implemented with the imbalanced-learn library using the parameters sampling_strategy = 'auto' and k_neighbors = 2.

Lastly, the same classifier implemented for the distinction between PD and HC was implemented, using the same parameters as when it was tested for that task, as described in subsection 4.4.1. Additionally, like for the aforementioned classification task, the features extracted from the magnitude were left out from the training of the classifier, and leave one out validation was used in the same way.

4.5.2 Fundamental frequency within rest tremor frequency

Firstly, the windows where the F0 was between 3 and 6 Hz were selected and saved in new files for each patient, since rest tremor is typically observed in this frequency band and the available UPDRS scores in the dataset are for this type of tremor. Consequently, these files will only contain the features from windows of the 4 hour signal previously selected where the peak value is in the rest tremor frequency band, thus selecting windows with this type of tremor as the dominant tremor.

The data was analyzed in regards to the number of windows selected for each 4 hour signal, correlating this value to the UPDRS score for rest tremor. After counting the number of windows with the F0 between 3 and 6 Hz from each axis and the magnitude of both forearms of each patient, these values were organized in a single file. Then, the mean of the three axes was calculated for each forearm of each patient.

Additionally, the same analysis was performed for intervals of 2 and 1 hour by dividing the 4 hour interval into two intervals of 2 hours and four intervals of 1 hour, as such the number of predictions was doubled and quadrupled, respectively.

Values for the number of windows corresponding to each score were defined and adapted for the shorter signal intervals of 2 hours and 1 hour. The values established for the analysis of the different signal intervals are presented in table 4.5. It is important to note that since, in this dataset, no rest tremor score of 4 was given for the UPDRS evaluation, this score was not included in this analysis.

Table 4.5: UPDRS scores and the corresponding number of windows for the 4,2, and 1 hour intervals.

UPDRS score	Number of windows		
	4 hours	2 hours	1 hour
0	< 200	< 100	< 50
1	200 - 400	100 - 200	50 - 100
2	400 - 600	200 - 300	100 - 150
3	> 600	300 - 400	150 - 200

To get the final UPDRS score prediction the number of windows in each axis, in the magnitude and the mean was divided by 200, 100, and 50 for the 4 hour, 2 hour, and 1 hour intervals, respectively, since these are the intervals established for each 0 these time segments, as shown in table 4.5. The values obtained were then rounded down to get the final score prediction.

4.6 Performance evaluation metrics

The metrics used for the evaluation of the performance of the proposed methods were the accuracy, the recall, and the success rate.

The accuracy and recall for both groups were used to compare the classifiers tested for the distinction between PD and HC. In this case, the accuracy shows how often the classifier correctly predicts that a sample belongs to the PD or HC groups and the recall shows the amount of samples from the group that are correctly predicted. Equations 4.2, 4.3, and 4.4 show how these metrics were calculated for this comparison.

$$\text{Accuracy} = \frac{\text{PD predicted as PD} + \text{HC predicted as HC}}{\text{PD predicted as PD and HC} + \text{HC predicted as HC and PD}} \tag{4.2}$$

$$\text{Recall HC} = \frac{\text{HC predicted as HC}}{\text{HC predicted as HC} + \text{HC predicted as PD}} \tag{4.3}$$

$$\text{Recall PD} = \frac{\text{PD predicted as PD}}{\text{PD predicted as PD} + \text{PD predicted as HC}} \tag{4.4}$$

After the classifier with the best performance was chosen, the classifier was trained, as described in subsection 4.4.1, by training on all patients except for one and, afterward applying the classifier to the patient that was left out. Classifying the windows of data from that patient as PD or HC. For the proposed method the percentage of windows classified as PD for each patient needs to be calculated. Since the classifier is applied to only one patient all windows belong to the same group, either PD or HC. However, if each patient the classifier is applied to is labeled as part of the PD group, the recall, calculated as shown in equation 4.5, will be the percentage of windows classified as PD.

$$\text{Recall} = \frac{\text{Windows predicted as PD}}{\text{Windows predicted as PD} + \text{Windows predicted as HC}} \tag{4.5}$$

Lastly, the success rate was used to evaluate the performance of the proposed methods and was calculated by dividing the amount patients classified correctly by the total amount of patients.

Chapter 5

Results and Discussion

5.1 Dataset

Firstly, it is important to discuss how the characteristics of this dataset might influence the results obtained. The dataset used mostly contains individuals with an absent or slight rest tremor, represented by the scores 0 and 1, respectively. Having only four instances where rest tremor was evaluated as mild, with a score of 2, and two instances where it was evaluated as moderate, with a score of 3. Consequently, the individuals, who experience tremor, mostly experience it infrequently [10]. Furthermore, it is common to have patients with a score of higher than 0 for one forearm, while having 0 for the other. This could indicate that the patients are experiencing one sided tremor, which is typically observed on the onset of the disease [10].

This may also lead to the conclusion that the methods used if good results are achieved for the distinction between the PD and HC groups, could be used in the support of diagnosis on the onset of the disease and/or for low rest tremor scores. Nonetheless, testing with individuals suffering from PD at a more advanced stage and/or with a higher rest tremor score could help testify the proposed methods, as well as reveal new approaches for the tasks in this study.

5.2 Welch's periodograms

The PSD plots obtained through Welch's method were briefly visually analyzed, however this analysis was not used for the estimation of rest tremor scores or to differentiate individuals from the PD and the control groups.

The graph in figure 5.1 shows the Welch's periodogram from the left forearm of patient 5. This patient is part of the HC group and was given a UPDRS rest tremor score of 0 for both forearms. In this figure, several plots are visible each representing the periodogram of one window from the 4 hour time period evaluated. Furthermore, in these periodograms, the peaks are mostly between 0 and 3 Hz. According to [12], dyskinesia is observed in this frequency band, however, in [11], this band is said to contain normal human movement. Nonetheless, this tremor band and the peaks in this band don't represent tremor associated with PD. It is important to note that the periodograms of subjects from the control and PD groups are sometimes similar, with these similarities mostly occurring with subjects from the PD group with an UPDRS score of 0.

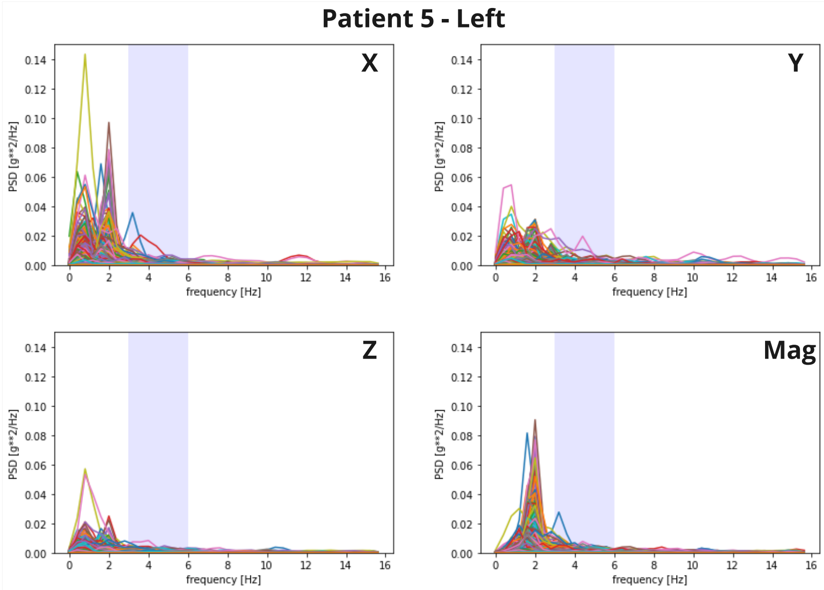


Figure 5.1: Welch's periodogram for the X, Y, and Z axes and magnitude (Mag) from the left forearm of patient 5.

Patient 12 has a tremor score of 1 for both forearms. On the periodograms extracted from the data from the right forearm, shown in figure 5.2, particularly on the Y axis more plots with peaks on the rest tremor frequency band, between 3 and 6 Hz, are visible.

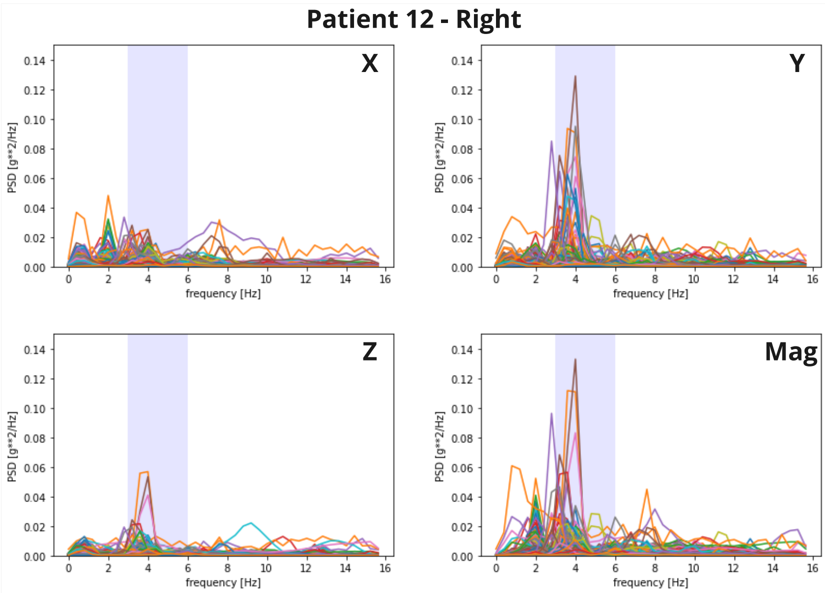


Figure 5.2: Welch's periodogram for the X, Y, and Z axes and magnitude (Mag) from the right forearm of patient 12.

From the analysis, it was noted that both plots of the Welch's periodogram of a signal where the forearm was given a score of 3, had distinctive peaks in the frequency bands corresponding to rest, postural, and kinetic tremors, as shown in figures 5.3 and 5.4. In the literature reviewed the frequency bands used by different authors vary, despite being similar. As there is not a single universally used frequency band for the rest, postural and kinetic tremors, the frequency bands for each tremor considered in this study were the ones defined in [12]. Those frequency bands are between 3 and 6 Hz for rest tremor, between 6 and 9 Hz for postural tremor, and between 9 and 12 Hz for kinetic tremor.

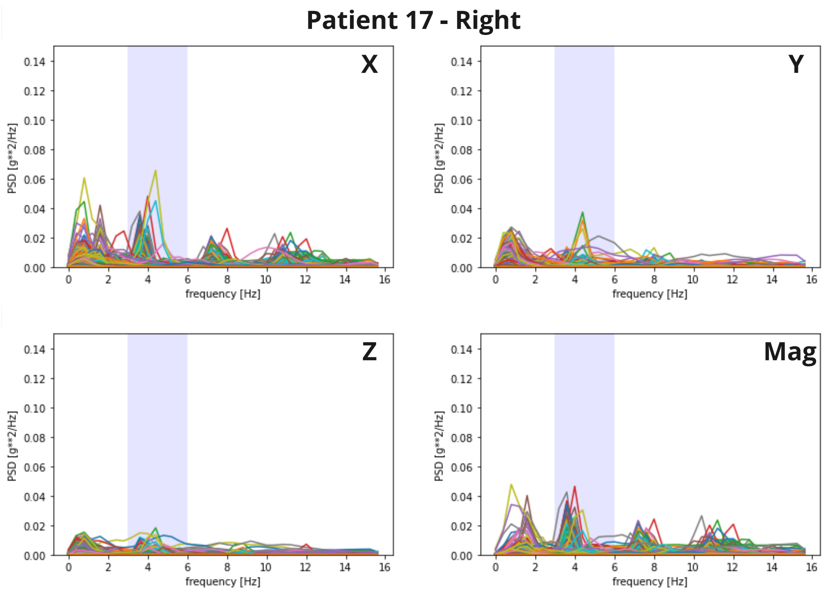


Figure 5.3: Welch's periodogram for the X, Y, and Z axes and magnitude (Mag) from the right forearm of patient 17.

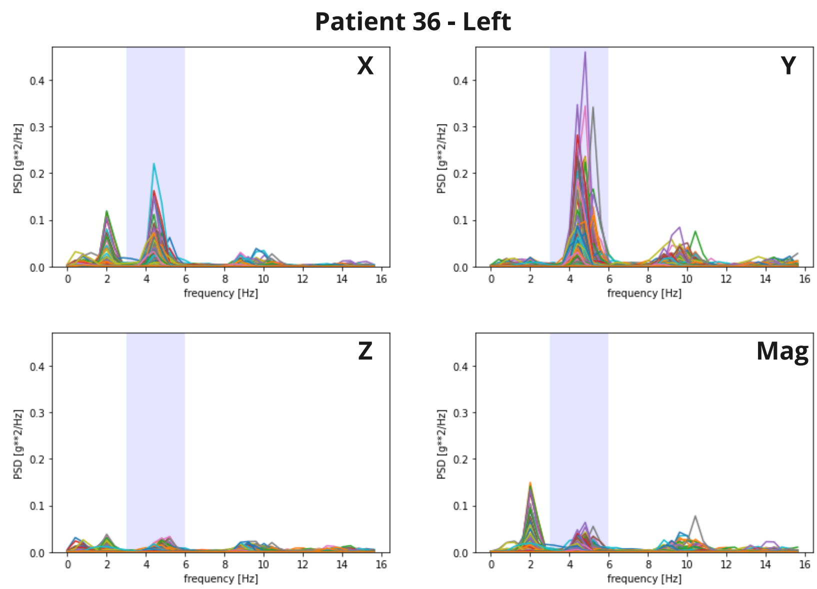


Figure 5.4: Welch's periodogram for the X, Y, and Z axes and magnitude (Mag) from the left forearm of patient 36.

Additionally, some signals with a tremor score of 2 presented similar peaks in the tremor frequency bands, like the one shown in figure 5.5. No periodogram extracted from the HC group's signals had similar peaks. Thus, these types of plots seem to be associated with PwPD and a higher score tremor.

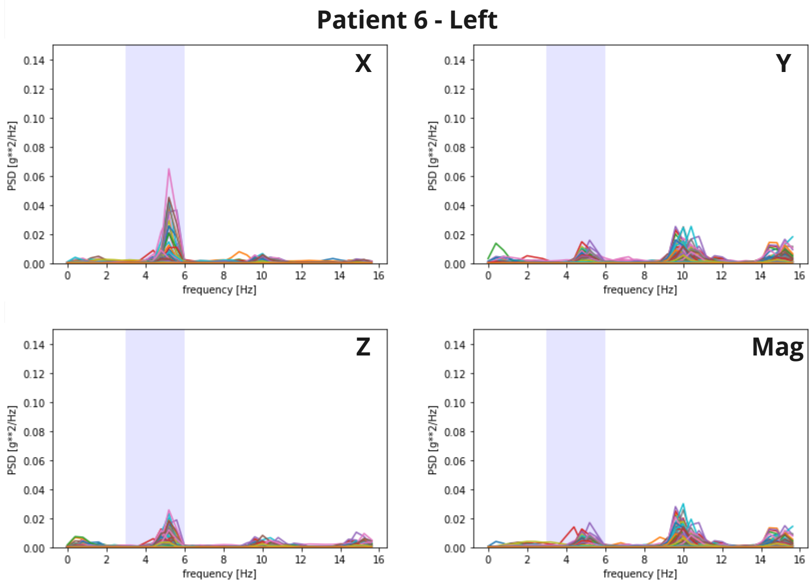


Figure 5.5: Welch's periodogram for the X, Y, and Z axes and magnitude (Mag) from the left forearm of patient 6.

The scale of the PSD axis for the periodograms of patient 36, shown in figure 5.4, is different from the scale of that axis for the other periodograms presented in this subsection. All graphs except for this one were presented with a smaller scale since the plots were not as visible with the same scale as for patient 36, whose periodograms for the left forearm had peaks with higher PSD values, particularly in the Y axis. All of the Welch's periodograms in this subsection are presented in appendix A.2 with the same scale for the PSD axis as patient 36.

5.3 Distinction between PD and HC

5.3.1 Classification

The results for the five classifiers that were tested to differentiate between the PD and HC groups are shown in table 5.1.

Table 5.1: Results in percentage of the accuracy and recall for the five tested classifiers.

Classifier	Acc	Recall HC	Recall PD
DecisionTree	62%	62%	62%
SVM	62%	62%	62%
KNN	67%	64%	70%
MLP	68%	65%	70%
BaggingTree	72%	73%	70%

Based on these results, the chosen classifier was the Bagging Tree, due to, not only having the best accuracy (Acc) but also having the best recall, above 70%, for both groups.

Applying the Bagging tree classifier, with the method described in subsection 4.4.1, a success rate of 85.3% was achieved for the distinction between subjects from the PD and HC groups.

The method used for the classifier is based on the notion that the resting tremor is an occasional symptom for PD patients and people without this disease can also experience tremor, not associated with PD and caused by other factors [10]. Since this study is focused on tremor and the distinction between the PD and HC groups is also based on this symptom, windows from subjects of the PD group can have no tremor and thus be predicted to belong to the HC group and windows from subjects of the HC group can contain tremor not associated with PD or noise and thus be predicted to belong to the PD group. As such, the classification was implemented and its results were evaluated based on the percentage of windows for each group that is classified as PD.

For this, leave one patient out was implemented, as described in 4.4.1, training data with all patients except one and then applying the trained classifier to the patient that was left out. Then, if all windows, for both PD and HC groups, patients are considered to belong to the positive class and are thus PD, the recall can be used to calculate the percentage of windows that are predicted to be part of the PD group for each patient. In this case, the recall is calculated by dividing the amount of windows predicted to be part of the PD group by the sum of windows predicted to as part of the PD and HC control groups.

Figure 5.6 contains a graphic representation of the percentage results obtained from the classification, where for each patient the Recall(1) represents the recall for the classifier trained and tested for both forearms with the positive class as PD, the Recall(1)_left represents the recall for the left forearm, and Recall(1)_right represents the recall for the right forearm. As such, the percentages show the amount of windows for each forearm and each patient predicted to belong to the PD group.

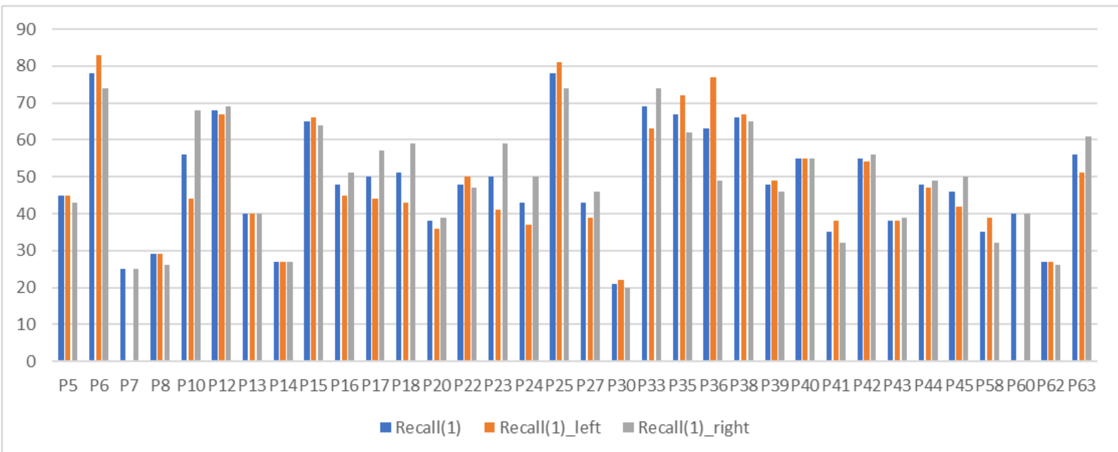


Figure 5.6: Classification results with, where Recall(1) is the percentage with both forearms and Recall(1)_left and Recall(1)_right are the classification for the left and right forearms respectively.

Despite the proposed method being based on the analysis of the individual results from both forearms, the results of the classifier with both forearms simultaneously were also analyzed, as shown in figure 5.6 and explored in this paragraph. There was an inconclusive result for the left forearm of patient 22, the right forearm of patient 24, and the right forearm of patient 45. In the case of patient 22, the right side is 47%, consequently, the patient is classified as control, thus being given the correct classification. On the other hand, the right side of patient 24 is given a 37%, also being classified as control, receiving the wrong classification. For patient 45, the left side has a percentage of 42%, being correctly classified as control. Patients 13 and 44 were classified as control, despite being part of the PD group, nonetheless, these patients have rest tremor scores of 0 for both hands. This indicates that tremor was absent during the clinical evaluation, since the presence and intensity of tremor are not constant and this state of absent tremor could occur during the evaluated signal or tremor can be infrequently present during the 4 hours, this could explain the incorrect prediction. For patient 16, with the classification for both forearms together the patient is given the correct classification of control. However, for the individual forearm classification, the right side is classified as PD and thus the patient is classified as part of the PD group. Patients 17 and 23 were classified as inconclusive and patient 24 was incorrectly classified as control by the classifier trained and applied to both forearms together. Lastly, patients 13 and 18 were given the wrong classification for classification based on both forearms together and individually.

Some incorrect predictions may be due to some windows from subjects of the PD group not containing tremor and some windows from the HC group containing either tremor caused by a different factor or, instead containing noise generated by living activities. Moreover, the method could be tested with patients on the OFF state of medication, possibly improving the distinctions between both groups. This method could also be tested with data collected in the clinic, which could reduce the noise associated with living activities. Furthermore, the classifier results could be improved by applying feature selection and, after this step, a different classifier may provide better results.

5.3.2 Kurtosis

The highest correlation, among all the features, with the distinction between individuals being part of the PD or control group, was the kurtosis of the Y axis, with a value of -0.71. This means that this feature, in the Y axis, is the best in distinguishing individuals with and without PD.

The mean and standard deviation (SD) of the values of these features were calculated for the PD and HC groups, as described in subsection 4.4.2. Then, the standard deviation was summed and subtracted from the mean. The intervals obtained, shown in table 5.2, represent the typical values of kurtosis for the two groups.

There is a gap between the typical value intervals of the two groups. The prediction for any

Table 5.2: Mean and standard deviation of kurtosis, in the Y axis, values for the PD and control groups.

Group	Mean	SD	Interval
PD	4.47	2.36	2.11-6.83
Control	9.31	2.45	6.86-11.76

mean value in this gap, as outside the typical intervals, would be considered inconclusive, however all previously calculated mean values for the kurtosis from all data windows of the subject, for the right and left sides, where within one of the typical intervals defined for both groups.

A success rate of 83.3% was achieved for the distinction between people with PD and healthy controls. As such, despite both methods obtaining promising results, this method achieves a worse success rate than the Bagging Tree classifier. However, this method also needs less processing power and computation time, which could be used as an argument for its implementation. These results could be further improved by testing different arguments in the kurtosis.

The task of making this distinction is extremely important, especially for patients at the onset of the disease, since early detection can help slow down the progression of the disease. In addition, it is pertinent that the diagnosis is correct since a prolonged levodopa medication intake by people not affected by PD can create other serious health problems [10].

Both methods tested for this task could be implemented to support the diagnosis of PD. Furthermore, by applying the kurtosis to the forearms of patients where the Bagging Tree classifier was inconclusive a success rate of 88.2% is achieved, due to the right forearm of patient 24 being classified as PD by the kurtosis, and consequently, this patient being correctly classified as part of the PD group. Hence, the conjugation of both methods may be a good approach for this task.

5.4 UPDRS rest tremor estimation

5.4.1 Classification

For this task, the classifier returns the number of windows it classifies as belonging to each class, those classes being the UPDRS scores 0 to 3. Then the percentage of windows belonging to each class was calculated.

For all patients, the highest percentage of windows was classified as having a score of 0. Thus, it was concluded that this classifier could not be implemented to predict UPDRS scores by identifying the tremor score detected in most windows of the 4 hour time period. Nonetheless, the classifier may have had better results if more data for scores 2 and 3 were available in the dataset since they are associated with a more frequent tremor.

Despite this, when the real UPDRS score is not 0, the percentage of windows with a predicted score of 0 is lower than when the real UPDRS score is 0. For the windows with a real score of 0, the percentage of windows classified with the same score is around 90%, however for all windows with a different score the percentage of windows classified with a score of 0 is lower than 80%.

More windows with a real score of 1 were predicted to have a score of 2 than their real score. For windows with a real score of 2, the percentage of windows predicted as score 2 is lower than the percentage of windows predicted as that same score for windows with a real score of 1 and 3. Moreover, windows with a real score of 2 have a lower percentage of windows predicted to have this score than windows predicted to have scores 1 and 3. Other than the percentage of windows predicted to be 0, windows with a real score of 2, have more windows predicted to have a score of 1 than any other score.

Windows with a real score of 3, despite having the highest rest tremor score included in this study, have a higher percentage of windows classified with a score of 0 than windows with a real score of 2. This contradicts the characterization of rest tremor with a real score of 3 in [10], shown in table 2.2, as being present most of the time. Additionally, windows with a real score of 3 also have a lower percentage of windows predicted to have a score of 3 than windows with a real score of 1 and 2. The percentage of windows predicted to represent each score for each one of the real scores are shown in table 5.3.

Table 5.3: Predicted percentage of windows with each UPDRS rest tremor score for subjects' forearms with each UPDRS rest tremor score.

Real score	**Predicted score**			
	0	1	2	3
0	90.09%	6.01%	2.56%	1.21%
1	79.57%	8.76%	9.34%	2.32%
2	61.86%	19.24%	8.13%	10.76%
3	66.96%	11.10%	19.90%	2.05%

Some of the inconsistencies detected, especially for windows with real scores 2 and 3, might be due to those being the classes with more synthetic data, with only four subjects having a score of 2 in one of their forearms and two subjects having a score of 3 in one of their forearms. Noise in some signals might also have an influence on these results. Moreover, these inconsistencies might also be the result of an incorrect score being given in the clinic or tremor not being consistently present or not consistently having the same score. As such, since the scores given to each forearm of the patient were considered as the real score for all windows of that forearm data, some windows had an incorrect real score as a label. Thus, an unsupervised or semi-supervised machine learning method might have better results for this task, since the classifier would not train with incorrect labels. A semi-supervised machine learning method could be implemented using data from clinical evaluations with correct tremor score labels given by a trained professional for that data during the evaluation and continuous unlabelled data.

The percentage of windows classified with each score for each of the PD and HC groups is shown in table 5.4. As expected, the HC group had a higher percentage of windows classified with a score of 0 and the PD group had a higher percentage of windows classified as 1, 2, or 3.

Table 5.4: Predicted percentage of windows classified with each UPDRS rest tremor score for the PD and HC groups.

Group	Percentage of windows			
	0	1	2	3
PD	80.68%	10.06%	6.21%	2.86%
HC	92.82%	4.28%	1.95%	0.94%

5.4.2 Fundamental frequency within rest tremor frequency

The association of the number of windows with F0 within 3 and 6 Hz and the UPDRS tremor scores is feasible because the scores are not only related to the intensity of the tremor but also to the frequency of the tremor and duration of its occurrence. Since a higher frequency of tremor and longer duration will lead to more windows in the signal with the presence of tremor, the number of windows should be associated with the UPDRS tremor score. The correlation between the features extracted and rest tremor scores was also tested but no strong correlations were found.

The results for the prediction of the UPDRS rest tremor scores according to the number of windows with F0 between 3 and 6, excluding the rounding of the prediction values, for patient 6 are shown in table 5.5.

Table 5.5: Analysis and prediction performed on the data on both forearms of patient 6 for the 4 hour time period.

ID	Number of windows					Prediction				
	X	Y	Z	Magnitude	Mean	X	Y	Z	Magnitude	Mean
P6_left	537	297	488	436	440.67	2.68	1.48	2.44	2.18	2.20
P6_right	263	213	239	317	238.33	1.32	1.065	1.195	1.585	1.19

Table 5.6 shows the UPDRS rest tremor scores given during the clinical assessment and the rounded down predictions of these scores for patient 6 in the 4 hours period. For the patient shown in this table, all predictions were incorrect for the right forearm and the predictions obtained from the Y axis were also incorrect for the left forearm. It should be noted that tremor is not constant and its presence and intensity vary during the day, as stated previously. A rest tremor with a UPDRS score of 0, is characterized as absent or normal, a tremor with a score of 1 is characterized as infrequently present for a short duration, and a tremor with a score of 2 is characterized as infrequently present for a considerably long duration, as stated in 2.2 and [10]. Consequently, it is possible that for the right forearm, no tremor was present during the clinical evaluation and it was present

during the 4 hour time period that was evaluated since for rest tremor scores 1 and 2 this symptom is not frequently present. Furthermore, noise in the signal might also be the cause for these incorrect predictions.

Table 5.6: Predicted and real UPDRS rest tremor scores for patient 6 in a 4 hour time period.

ID	UPDRS score	Prediction				
		X	Y	Z	Magnitude	Mean
P6_left	2	2	1	2	2	2
P6_right	0	1	1	1	1	1

After rounding down the prediction values this method's success rate was determined based on the percentage of times the prediction was correct. The success rates for the different axes, as well as the magnitude and mean, in the different time periods are presented in table 5.7. From these results it was concluded that the predictions using the magnitude vector failed the majority of the time, thus not having a good success rate for any of the time periods. On the contrary, the mean number of windows in all axes has a better performance than the magnitude for any time period, with the period of 4 hours having the best success rate at 83.33% and the period of 1 hour having the worst at 74.24%. However, except for the Y axis which had the second worst performance in all time periods after the magnitude, both the X and Z axes performed better than the mean for the 2 and 1 hour time periods, with the X axis also performing better in the 4 hour time period. The X axis had the highest success rates for the 4 and 2 hour periods and the Y axis had the highest success rate for the 1 hour period. However, for the later time period, the difference between the success rate for the X and Z axes is only 0.38%. As such, it can be concluded that for this dataset and the signals evaluated the best success rate for the prediction of UPDRS rest tremor scores using this method is achieved with the X axis. In addition, the best success rate overall is 87.88%, for the 4 hour time period with the X axis.

Table 5.7: Success rate for all axes, the magnitude and the mean of the number of windows in the 4, 2, and 1 hour intervals.

Interval	Success rate				
	X	Y	Z	Magnitude	Mean
4	**87.88%**	71.21%	81.82%	34.85%	83.33%
2	82.58%	59.09%	81.82%	37.88%	78.79%
1	77.27%	59.09%	77.65%	39.39%	74.24%

Despite having rounded down the prediction values to obtain the final score prediction, the values with decimals allow us to know if the prediction was close to the limit between scores and by how much it might have failed. For the X axis in the 4 hour time period, that had the best success rate, if all the windows that failed the prediction by less than 0.1 are discarded, the success rate increases to 92.06%. Nonetheless, since to apply this method in a real setting it is not known which predictions are correct or incorrect, all the windows that are within 0.1 of the limit between scores were discarded. For example, windows with values between 0.9 and 1.1, including these values, were discarded. By discarding these windows and marking their predictions as inconclusive between the two scores they are closest to, which would be 0 and 1 in the previous example, the success rate achieved would be 91.23%.

Chapter 6

Conclusions and Future Work

6.1 Conclusion

Parkinson's Disease is the second most common neurodegenerative disorder, affecting a large percentage of the elderly population. This study is focused on the tremor associated with this disorder, and its evaluation using wearable sensors. This dissertation can be divided into three parts, the SLR, the proposed mobile app, and the computation framework.

The SLR provides an overview of some of the methods implemented for the evaluation of the tremor associated with PD using wearable sensors. Moreover, the discussion of this review provides a more in depth explanation of the datasets, technology, experimental design, and computation framework used. Thus, the SLR contains a summary of all steps involved in the analysis of PD tremor data, from the data collection to its processing, giving the context needed for the creation of a complete evaluation model including all these steps. Nonetheless, not many articles were included in this review and it could be expanded by reviewing the exclusion criteria. This would allow for more studies to be included, providing a better understanding of the most common methods.

The mobile app was proposed for the purpose of collecting data from accelerometer and gyroscope sensors during 5 different tasks for the evaluation of the tremor associated with PD. Furthermore, this app also collects the answers of users to three types of questionnaires, in order to contextualize the sensor data. The app was to be used in a clinical or home environment, but first, it needs to be tested with data collection performed while the UPDRS evaluations are performed by an examiner and the processing of that data. Furthermore, wearable sensors could be incorporated into the data collection by sending the sensor data to the app. This makes the collection of continuous sensor data possible. In the future, the methods for the evaluation of the tremor could be incorporated into the app, turning it into a system capable of collecting and processing data, as well as showing results to a subject or doctor. This could support the diagnosis of PD or help in follow-up evaluations by providing information on the tremor through time.

Data has not been collected using the mobile app. However, the data collection using this mobile app from subjects without PD disease was approved by the ethics commission of the University of Beira Interior. The data will be collected from subjects without PD in different retirement homes. This data will be used can be used as a control dataset to test differences between PD patients and people without PD. Furthermore, the collection of data using this mobile app will help test its applicability. In the Future, it would be beneficial to use this app to collect data from PD patients. However, this is not proposed and implemented for this thesis due to bigger limitations in the organization of the data collection from PD patients.

In the computation framework, the analysis of the methods proposed can be divided into methods to distinguish between PD and HC and methods to estimate UPDRS rest tremor scores. For the distinction between PD and HC, one method consists of the implementation of a Bagging tree classifier and the other of the analysis of the kurtosis. Despite the classifier performing slightly better for this task, an approach using both methods, where kurtosis is used for inconclusive predictions by the classifier, had the best performance. Both these methods, implemented individually or together could help support a PD diagnosis. On the other hand, the Bagging Tree classifier implemented for the estimation of UPDRS rest tremor scores did not have promising results. However, the estimation of these scores based on the number of F0 windows in the rest tremor frequency range had a better success rate and could help monitor tremor and the progression of the disease by providing data for follow-up consultations.

In conclusion, the methods and mobile app proposed in this study could, either individually or combined, help support both diagnosis and monitoring of tremor and its progression as the disease progresses. Nonetheless, the methods proposed need to be tested using other datasets in order to verify their results and their results could be improved with steps like feature selection for the classifier and different arguments in the kurtosis. Furthermore, by calculating the correlations between the UPDRS rest tremor scores and the features with F0 between in the rest tremor frequency band a new approach for this task could be developed.

6.2 Future Work

In this dissertation, it was not possible to delve deeper into each aspect of the study, as such more testing and development will be part of the future work. Future tasks for the further development of the mobile app and processing methods proposed are the following.

- Implement a wearable sensor device connected to the app;
- Test the applicability of the mobile app for at home data collections.
- Collect data using different sensors, like electromyography, which could provide more information regarding the tremor and/or other symptoms;
- Collect sensor data in clinical settings, while the UPDRS evaluation or tests proposed for the mobile app are performed, and with a higher sampling rate of 50 or 100 Hz;

- Test the computation framework with data collected during UPDRS clinical evaluations, helping validate this method to be applied in both a controlled environment and during free living;
- For the distinction between the PD and HC groups, feature selection and selecting specific axes or the magnitude for the classifier could improve its results;
- The success rate obtained by implementing the kurtosis method could be improved by testing different arguments of this feature;
- For the estimation of UPDRS rest tremor scores using a classifier, test unsupervised and semi-supervised machine learning methods and test the classifier using only data from windows with F0 in the rest tremor frequency band;
- Test the correlations between the features in windows where F0 is in the rest tremor frequency band and the UPDRS rest tremor scores, which could reveal a new approach for the estimation of the scores.

Bibliography

[1] A. de Oliveira Andrade, A. P. S. Paixao, A. M. Cabral, A. G. Rabelo, L. M. D. Luiz, V. C. Dionísio, M. F. Vieira, J. M. Pereira, A. Rueda, S. Krishnan *et al.*, "Task-specific tremor quantification in a clinical setting for parkinson's disease," *Journal of Medical and Biological Engineering*, vol. 40, no. 6, pp. 821–850, 2020. 1, 5, 6, 7, 8, 20

[2] W. Huo, P. Angeles, Y. F. Tai, N. Pavese, S. Wilson, M. T. Hu, and R. Vaidyanathan, "A heterogeneous sensing suite for multisymptom quantification of parkinson's disease," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 6, pp. 1397–1406, 2020. 1, 5, 6, 8, 9

[3] T. Zajki-Zechmeister, M. Kögl, K. Kalsberger, S. Franthal, N. Homayoon, P. Katschnig-Winter, K. Wenzel, L. Zajki-Zechmeister, and P. Schwingenschuh, "Quantification of tremor severity with a mobile tremor pen," *Heliyon*, vol. 6, no. 8, p. e04702, 2020. 1, 5, 6, 7, 8, 20

[4] A. Smid, J. W. J. Elting, J. M. C. van Dijk, B. Otten, D. M. Oterdoom, K. Tamasi, T. Heida, T. van Laar, and G. Drost, "Intraoperative quantification of mds-updrs tremor measurements using 3d accelerometry: A pilot study," *Journal of Clinical Medicine*, vol. 11, no. 9, p. 2275, 2022. 1, 6, 8, 9, 40

[5] M. Burq, E. Rainaldi, K. C. Ho, C. Chen, B. R. Bloem, L. J. W. Evers, R. C. Helmich, L. Myers, W. J. Marks, and R. Kapur, "Virtual exam for parkinson's disease enables frequent and reliable remote measurements of motor function," *npj Digital Medicine*, vol. 5, no. 1, 2022, cited by: 6; All Open Access, Gold Open Access, Green Open Access. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85130721603&doi=10.1038%2fs41746-022-00607-8&partnerID=40&md5=fe7b9b2b8e1d0c0725856cb250f6a3e4 1, 5, 7, 8, 15, 17, 19, 20, 21, 22, 25, 26

[6] J. L. Adams, K. Dinesh, C. W. Snyder, M. Xiong, C. G. Tarolli, S. Sharma, E. R. Dorsey, and G. Sharma, "A real-world study of wearable sensors in parkinson's disease," *npj Parkinson's Disease*, vol. 7, no. 1, p. 106, 2021. 1, 36

[7] N. Shawen, M. K. O'Brien, S. Venkatesan, L. Lonini, T. Simuni, J. L. Hamilton, R. Ghaffari, J. A. Rogers, and A. Jayaraman, "Role of data measurement characteristics in the accurate detection of parkinson's disease symptoms using wearable sensors," *Journal of neuroengineering and rehabilitation*, vol. 17, no. 1, pp. 1–14, 2020. 5, 6, 7, 8, 9, 20, 25, 38, 39, 40

[8] D. K. Simon, C. M. Tanner, and P. Brundin, "Parkinson disease epidemiology, pathology, genetics, and pathophysiology," *Clinics in geriatric medicine*, vol. 36, no. 1, pp. 1–12, 2020. 5, 6

[9] M. J. Armstrong and M. S. Okun, “Diagnosis and treatment of parkinson disease: a review,” *Jama*, vol. 323, no. 6, pp. 548–560, 2020. 5, 20

[10] B. Anbalagan, S. Karnam Anantha, and R. Kalpana, “Novel approach to prognosis parkinson’s disease with wireless technology using resting tremors,” *Wireless Personal Communications*, vol. 125, no. 4, pp. 2985–2999, 2022. 5, 6, 7, 8, 15, 17, 19, 20, 21, 22, 23, 26, 45, 49, 51, 52, 53

[11] R. San-Segundo, A. Zhang, A. Cebulla, S. Panev, G. Tabor, K. Stebbins, R. E. Massa, A. Whitford, F. De la Torre, and J. Hodgins, “Parkinson’s disease tremor detection in the wild using wearable accelerometers,” *Sensors*, vol. 20, no. 20, p. 5817, 2020. 5, 8, 9, 15, 17, 19, 20, 21, 22, 23, 25, 26, 27, 29, 39, 45

[12] E. Kuosmanen, F. Wolling, J. Vega, V. Kan, Y. Nishiyama, S. Harper, K. Van Laerhoven, S. Hosio, D. Ferreira *et al.*, “Smartphone-based monitoring of parkinson disease: quasi-experimental study to quantify hand tremor severity and medication effectiveness,” *JMIR mHealth and uHealth*, vol. 8, no. 11, p. e21543, 2020. 6, 15, 17, 19, 20, 21, 22, 23, 26, 29, 39, 40, 45, 47

[13] 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: official journal of the Movement Disorder Society*, vol. 23, no. 15, pp. 2129–2170, 2008. 7

[14] J. J. Kavanagh and H. B. Menz, “Accelerometry: a technique for quantifying movement patterns during walking,” *Gait & posture*, vol. 28, no. 1, pp. 1–15, 2008. 8

[15] 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,” *Movement disorders*, vol. 31, no. 9, pp. 1314–1326, 2016. 10, 11

[16] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “Smote: synthetic minority over-sampling technique,” *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002. 10, 42

[17] M. Ouzzani, H. Hammady, Z. Fedorowicz, and A. Elmagarmid, “Rayyan—a web and mobile app for systematic reviews,” *Systematic Reviews*, vol. 5, no. 1, p. 210, 2016. [Online]. Available: http://dx.doi.org/10.1186/s13643-016-0384-4 12

[18] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, R. Chou, J. Glanville, J. M. Grimshaw, A. Hróbjartsson, M. M. Lalu, T. Li, E. W. Loder, E. Mayo-Wilson, S. McDonald, L. A. McGuinness, L. A. Stewart, J. Thomas, A. C. Tricco, V. A. Welch, P. Whiting, and D. Moher, “The prisma 2020 statement: an updated guideline

for reporting systematic reviews," *BMJ*, vol. 372, 2021. [Online]. Available: https://www.bmj.com/content/372/bmj.n71 13

[19] L. Sigcha, I. Pavón, N. Costa, S. Costa, M. Gago, P. Arezes, J. M. López, and G. De Arcas, "Automatic resting tremor assessment in parkinson's disease using smartwatches and multitask convolutional neural networks," *Sensors*, vol. 21, no. 1, p. 291, 2021. 15, 17, 19, 20, 21, 22, 23, 24, 25, 26, 27, 29, 39

[20] L. Chen, G. Cai, H. Weng, J. Yu, Y. Yang, X. Huang, X. Chen, and Q. Ye, "More sensitive identification for bradykinesia compared to tremors in parkinson's disease based on parkinson's kinetigraph (pkg)," *Frontiers in Aging Neuroscience*, vol. 12, p. 594701, 2020. 15, 17, 19, 20, 21, 22, 25, 26

[21] J. McNames, V. V. Shah, M. Mancini, C. Curtze, M. El-Gohary, M. Aboy, P. Carlson-Kuhta, J. G. Nutt, and F. Horak, "A two-stage tremor detection algorithm for wearable inertial sensors during normal daily activities," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2019, pp. 2535–2538. 16, 17, 19, 20, 21, 22, 23, 26

[22] B. Boroojerdi, R. Ghaffari, N. Mahadevan, M. Markowitz, K. Melton, B. Morey, C. Otoul, S. Patel, J. Phillips, E. Sen-Gupta *et al.*, "Clinical feasibility of a wearable, conformable sensor patch to monitor motor symptoms in parkinson's disease," *Parkinsonism & related disorders*, vol. 61, pp. 70–76, 2019. 16, 18, 19, 20, 21, 22, 26

[23] L. A. Sanchez-Perez, L. P. Sanchez-Fernandez, A. Shaout, J. M. Martinez-Hernandez, and M. J. Alvarez-Noriega, "Rest tremor quantification based on fuzzy inference systems and wearable sensors," *International journal of medical informatics*, vol. 114, pp. 6–17, 2018. 16, 18, 19, 20, 21, 22, 23, 26

[24] L. Klingelhoefer, A. Rizos, A. Sauerbier, S. McGregor, P. Martinez-Martin, H. Reichmann, M. Horne, and K. Chaudhuri, "Night-time sleep in parkinson's disease–the potential use of parkinson's kinetigraph: a prospective comparative study," *European journal of neurology*, vol. 23, no. 8, pp. 1275–1288, 2016. 21

[25] J. G. Habets, M. Heijmans, A. F. Leentjens, C. J. Simons, Y. Temel, M. L. Kuijf, P. L. Kubben, and C. Herff, "A long-term, real-life parkinson monitoring database combining unscripted objective and subjective recordings," *Data*, vol. 6, no. 2, p. 22, 2021. 31, 36

[26] A. Joshi, S. Kale, S. Chandel, and D. K. Pal, "Likert scale: Explored and explained," *British journal of applied science & technology*, vol. 7, no. 4, pp. 396–403, 2015. 32

[27] K. Finstad, "Response interpolation and scale sensitivity: Evidence against 5-point scales," *Journal of usability studies*, vol. 5, no. 3, pp. 104–110, 2010. 32

[28] K. R. Ammann, T. Ahamed, A. L. Sweedo, R. Ghaffari, Y. E. Weiner, R. C. Slepian, H. Jo, and M. J. Slepian, “Human motion component and envelope characterization via wireless wearable sensors,” *BMC Biomedical Engineering*, vol. 2, no. 1, pp. 1–15, 2020. 37

[29] A. Channa, R.-C. Ifrim, D. Popescu, and N. Popescu, “A-wear bracelet for detection of hand tremor and bradykinesia in parkinson’s patients,” *Sensors*, vol. 21, no. 3, p. 981, 2021. 39

[30] P. Welch, “The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms,” *IEEE Transactions on audio and electroacoustics*, vol. 15, no. 2, pp. 70–73, 1967. 39

[31] S. Kokoska and D. Zwillinger, *CRC standard probability and statistics tables and formulae*. Crc Press, 2000. 40

[32] Scikit-learn, “1.10. decision trees,” accessed: 2023-12-14. [Online]. Available: https://scikit-learn.org/stable/modules/tree.html 40

[33] ——, “1.4. support vector machines,” accessed: 2023-12-14. [Online]. Available: https://scikit-learn.org/stable/modules/svm.html 40

[34] ——, “1.6. nearest neighbors,” accessed: 2023-12-14. [Online]. Available: https://scikit-learn.org/stable/modules/neighbors.html 41

[35] ——, “1.17. neural network models (supervised),” accessed: 2023-12-14. [Online]. Available: https://scikit-learn.org/stable/modules/neural_networks_supervised.html 41

[36] ——, “1.11. ensembles: Gradient boosting, random forests, bagging, voting, stacking,” accessed: 2023-12-14. [Online]. Available: https://scikit-learn.org/stable/modules/ensemble.html 41

[37] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and Édouard Duchesnay, “Scikit-learn: Machine learning in python,” *Journal of Machine Learning Research*, vol. 12, no. 85, pp. 2825–2830, 2011. [Online]. Available: http://jmlr.org/papers/v12/pedregosa11a.html 41

[38] I. learn, “Smote,” accessed: 2023-12-23. [Online]. Available: https://imbalanced-learn.org/stable/references/generated/imblearn.over_sampling.SMOTE.html 42

Appendix A

Appendix

A.1 Questionnaires

A.1.1 Morning questionnaire

1. Dormi bem.
    - Discordo fortemente
    - Discordo Parcialmente
    - Neutro
    - De certo modo concordo
    - Concordo
    - Concordo plenamente
2. Acordei várias vezes ao longo da noite.
    - Discordo fortemente
    - Discordo Parcialmente
    - Neutro
    - De certo modo concordo
    - Concordo
    - Concordo plenamente
3. Sinto-me descansado.
    - Discordo fortemente
    - Discordo Parcialmente
    - Neutro
    - De certo modo concordo
    - Concordo
    - Concordo plenamente
4. Foi fisicamente difícil levantar-me.
    - Discordo fortemente

- Discordo Parcialmente
- Neutro
- De certo modo concordo
- Concordo
- Concordo plenamente

5. Foi mentalmente difícil levantar-me.
    - Discordo fortemente
    - Discordo Parcialmente
    - Neutro
    - De certo modo concordo
    - Concordo
    - Concordo plenamente

A.1.2 Lunch questionnaire

1. Onde esteve no decorrer da manhã?
    - Casa
    - Trabalho
    - Viajar
    - Casa de amigos/família
    - Em público
    - Outro
2. Com quem esteve de manhã?
    - Ninguém
    - Família
    - Companheiro
    - Colega
    - Amigo
    - Outro
3. O que esteve a fazer de manhã?
    - Trabalho
    - Descansar
    - Trabalho de casa/biscate

- Desporto
- Outro

4. Consegui andar com facilidade.
    - Discordo fortemente
    - Discordo Parcialmente
    - Neutro
    - De certo modo concordo
    - Concordo
    - Concordo plenamente
5. Consegui estar de pé com facilidade.
    - Discordo fortemente
    - Discordo Parcialmente
    - Neutro
    - De certo modo concordo
    - Concordo
    - Concordo plenamente
6. Consegui falar com facilidade.
    - Discordo fortemente
    - Discordo Parcialmente
    - Neutro
    - De certo modo concordo
    - Concordo
    - Concordo plenamente
7. Experienciei tremor.
    - Discordo fortemente
    - Discordo Parcialmente
    - Neutro
    - De certo modo concordo
    - Concordo
    - Concordo plenamente
8. Experienciei rigidez.
    - Discordo fortemente

- Discordo Parcialmente
- Neutro
- De certo modo concordo
- Concordo
- Concordo plenamente

9. Movimentei-me devagar.
    - Discordo fortemente
    - Discordo Parcialmente
    - Neutro
    - De certo modo concordo
    - Concordo
    - Concordo plenamente
10. Tomou medicação para Parkinson?
    - Sim
    - Não
11. Sentiu o efeito da medicação?
    - Discordo fortemente
    - Discordo Parcialmente
    - Neutro
    - De certo modo concordo
    - Concordo
    - Concordo plenamente

A.1.3 Night questionnaire

Perguntas referentes ao perído da tarde.

1. Onde esteve de tarde?
    - Casa
    - Trabalho
    - Viajar
    - Casa de amigos/família
    - Em público

  - Outro

2. Com quem esteve de tarde?

  - Ninguém
  - Família
  - Companheiro
  - Colega
  - Amigo
  - Outro

3. O que esteve a fazer de tarde?

  - Trabalho
  - Descansar
  - Trabalho de casa/biscate
  - Desporto
  - Outro

4. Consegui andar com facilidade.

  - Discordo fortemente
  - Discordo Parcialmente
  - Neutro
  - De certo modo concordo
  - Concordo
  - Concordo plenamente

5. Consegui estar de pé com facilidade.

  - Discordo fortemente
  - Discordo Parcialmente
  - Neutro
  - De certo modo concordo
  - Concordo
  - Concordo plenamente

6. Consegui falar com facilidade.

  - Discordo fortemente
  - Discordo Parcialmente

- Neutro
- De certo modo concordo
- Concordo
- Concordo plenamente

7. Experienciei tremor.
   - Discordo fortemente
   - Discordo Parcialmente
   - Neutro
   - De certo modo concordo
   - Concordo
   - Concordo plenamente
8. Experienciei rigidez.
   - Discordo fortemente
   - Discordo Parcialmente
   - Neutro
   - De certo modo concordo
   - Concordo
   - Concordo plenamente
9. Movimentei-me de vagar.
   - Discordo fortemente
   - Discordo Parcialmente
   - Neutro
   - De certo modo concordo
   - Concordo
   - Concordo plenamente
10. Tomou medicação para Parkinson?
    - Sim
    - Não
11. Sentiu o efetito da medicação?
    - Discordo fortemente
    - Discordo Parcialmente

- Neutro
- De certo modo concordo
- Concordo
- Concordo plenamente

Perguntas referentes a todo o dia.

1. Hoje senti-me:
    - Bem
    - Em baixo
    - Com medo
    - Stressado/a
    - Cansado/a
    - Com sono
    - Alegre
    - Relaxado/a
    - Concentrado/a
    - Cansado/a
2. Tive dificuldade em:
    - Vestir-me
    - Comer/beber
    - Cuidado pessoal
    - Atividade doméstica
    - Nenhum
3. Consumiu alguma das seguintes substâncias ao longo do dia?
    - Café
    - Álcool
    - Tabaco
    - Nenhum

A.2 Welch's Periodograms

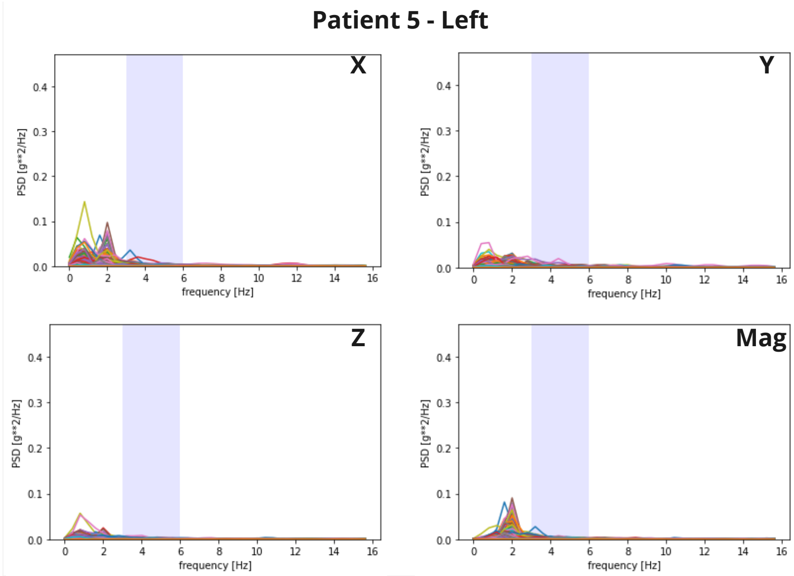


Figure A.1: Welch's periodogram for the X, Y, and Z axes and magnitude (Mag) from the left forearm of patient 5, with the scale of the PSD axis adjusted to 0.4 g^2/Hz.

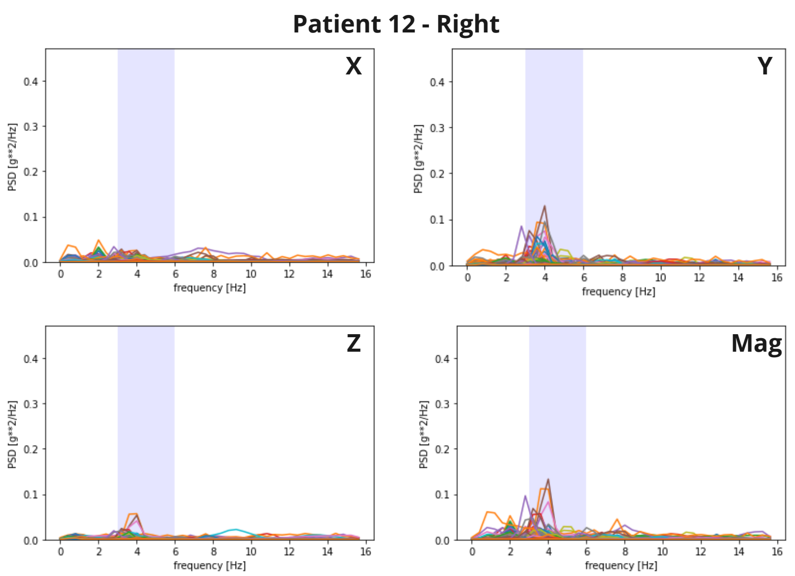


Figure A.2: Welch's periodogram for the X, Y, and Z axes and magnitude (Mag) from the right forearm of patient 12, with the scale of the PSD axis adjusted to 0.4 g^2/Hz.

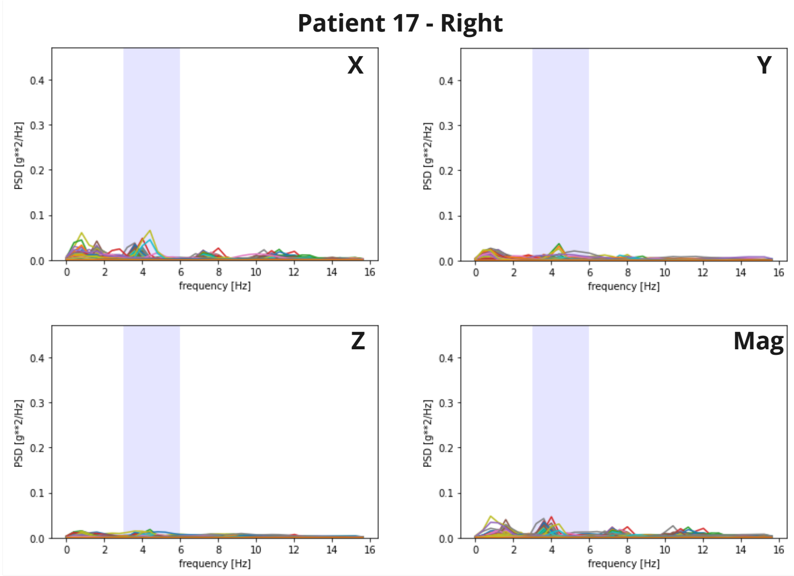


Figure A.3: Welch's periodogram for the X, Y, and Z axes and magnitude (Mag) from the right forearm of patient 17, with the scale of the PSD axis adjusted to 0.4 g^2/Hz.

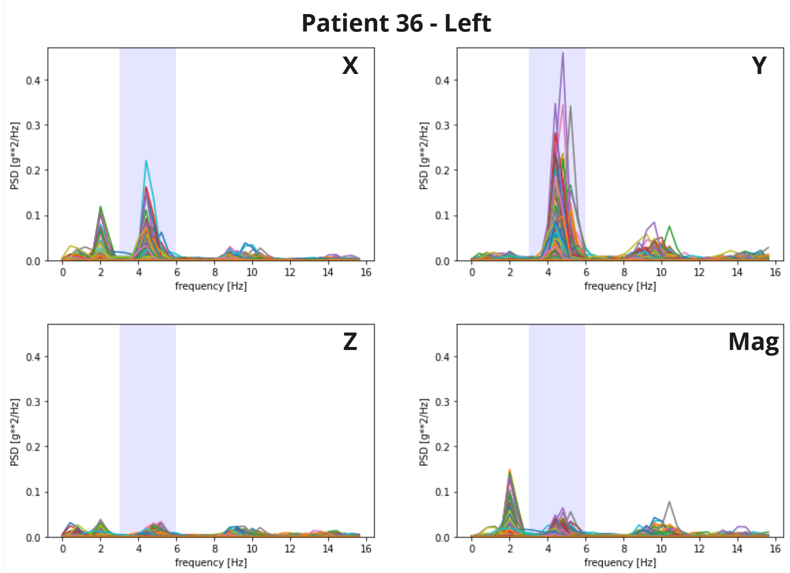


Figure A.4: Welch's periodogram for the X, Y, and Z axes and magnitude (Mag) from the left forearm of patient 36, with the scale of the PSD axis adjusted to 0.4 g^2/Hz.

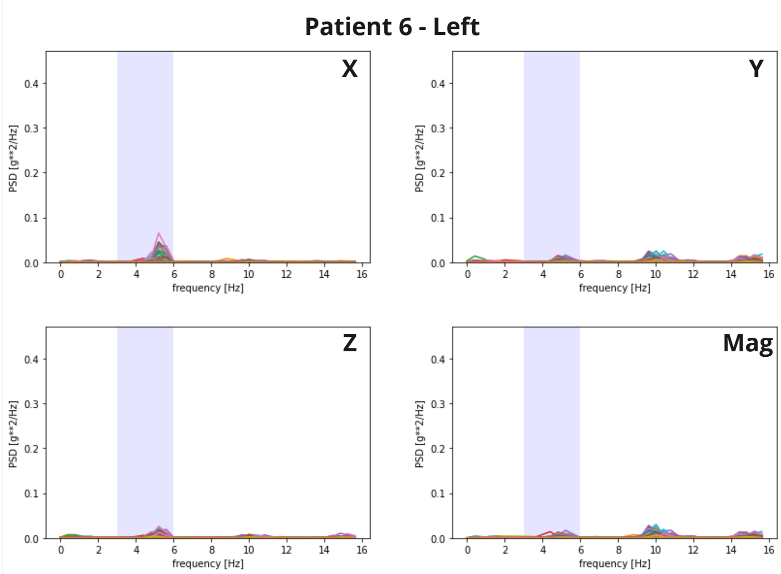


Figure A.5: Welch's periodogram for the X, Y, and Z axes and magnitude (Mag) from the left forearm of patient 6, with the scale of the PSD axis adjusted to 0.4 g^2/Hz.