

BRNO UNIVERSITY OF TECHNOLOGY

Faculty of Electrical Engineering  
and Communication

DOCTORAL THESIS

Brno, 2023

Justyna Skibińska





**BRNO UNIVERSITY OF TECHNOLOGY**

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION  
DEPARTMENT OF TELECOMMUNICATIONS



**TAMPERE UNIVERSITY**

FACULTY OF INFORMATION TECHNOLOGY AND COMMUNICATION SCIENCES  
UNIT OF ELECTRICAL ENGINEERING

**MACHINE LEARNING-AIDED MONITORING AND  
PREDICTION OF RESPIRATORY AND  
NEURODEGENERATIVE DISEASES USING WEARABLES**

**DOCTORAL THESIS**

**AUTHOR**                      **M.Sc. Eng. Justyna Skibińska**

**ADVISORS**                      **Assoc. Prof. Jiri Hosek**  
**Assoc. Prof. Radim Burget**  
**Prof. Evgeny Kucheryavy**

**BRNO 2023**

# **MACHINE LEARNING-AIDED MONITORING AND PREDICTION OF RESPIRATORY AND NEURODEGENERATIVE DISEASES USING WEARABLES**

**This thesis has been completed in a joint/double doctoral  
degree programme at Brno University of Technology,  
Czech Republic and Tampere University, Finland.**



**DPAD-EIT Electronics and Information Technologies (Double-Degree)**  
Brno University of Technology



**DPAWE Doctoral Programme in Dynamic Wearable Applications with  
Privacy Constraints**  
Tampere University



This dissertation is funded by the European Union's Horizon 2020 Research and Innovation programme under the Marie Skłodowska-Curie grant agreement No. 813278, A-WEAR.



## ABSTRACT

This thesis focuses on wearables for health status monitoring, covering applications aimed at emergency solutions to the COVID-19 pandemic and aging society. The methods of ambient assisted living (AAL) are presented for the neurodegenerative disease Parkinson's disease (PD), facilitating 'aging in place' thanks to machine learning and around wearables - solutions of mHealth. Furthermore, the approaches using machine learning and wearables are discussed for early-stage COVID-19 detection, with encouraging accuracy.

Firstly, a publicly available dataset containing COVID-19, influenza, and healthy control data was reused for research purposes. The solution presented in this thesis is considering the classification problem and outperformed the state-of-the-art methods, whereas the original paper introduced just anomaly detection and not shown the specificity of the created models. The proposed model in the thesis for early detection of COVID-19 achieved 78 % for the k-NN classifier. Moreover, a second dataset available on request was utilized for recognition between COVID-19 cases and two types of influenza. The scrutinisation in the form of the classification between the COVID-19 and Influenza groups is proposed as the extension to the research presented in the original paper [1] illustrating the foundation for this study - statistical analysis of the dataset. Differences between the COVID-19 and Influenza cases in duration and intensity of the disease occur likewise manifest in heart rhythm. The accuracy of the distinction between COVID-19 cases and influenza in the middle of the pandemic (data were gathered from 03.2020 to 05.2020) was equal to 73 % thanks to the k-NN. Furthermore, the contribution as the classification model of two aforementioned combined datasets was provided, and COVID-19 cases were able to be distinguished from healthy controls with 73 % accuracy thanks to XGBoost algorithm. The undeniable advantage of the illustrated approaches is taking into consideration the incubation period and contagiousness of the disease likewise presenting the methodologies dedicated to data gathered by the Fitbit device. Furthermore, the parallel analysis of various types of Influenza, COVID-19, and healthy control is novel and has not been thoroughly investigated yet.

In addition, some solutions for the detection of the aforementioned aging society phenomenon are presented. This study explores the possibility of fusing computerised analysis of hypomimia and hypokinetic dysarthria for the spectrum of Czech speech exercises. The introduced dataset is unique in this field because of its diversity and myriad of speech exercises. The aim is to introduce a new techniques of PD diagnosis that could be easily integrated into mHealth systems. A classifier based on XGBoost was used, and SHAP values were used to ensure interpretability. The presented interpretability allows for the identification of clinically valuable biomarkers. Moreover, the fusion of video and audio modalities increased the balanced accuracy to 83 %. This methodology pointed out the most indicative speech exercise – tongue twister from the clinical point of view. Furthermore, this work belongs to just a few studies which tackle the subject of utilising multimodality for PD and this approach was profitable in contrast with a single modality. Another study, presented in this thesis, investigated the possibility of detecting Parkinson's disease by observing changes in emotion expression during difficult-to-pronounce speech exercises. The obtained model with XGBoost achieved 69 % accuracy for a tongue twister. The usage of facial features, emotion recognition, and computational analysis of tongue twister was proved to be successful in PD detection, which is the key novelty and contribution of this study. Additionally, the unique overview of potential methodologies suitable for the detection of PD based on sleep disorders was depicted.

## KEYWORDS

aging society; artificial intelligence; COVID-19; machine learning; Parkinson's disease; signal processing; wearables







# Author's Declaration

**Author:** M.Sc. Eng. Justyna Skibinska  
**Author's ID:** 222671  
**Paper type:** Doctoral thesis  
**Academic year:** 2022/23  
**Topic:** Machine Learning-Aided Monitoring and Prediction of Respiratory and Neurodegenerative Diseases Using Wearables

I declare that I have written this thesis independently, under the guidance of the advisors and using exclusively the technical references and other sources of information cited in the thesis and listed in the comprehensive bibliography at the end of the thesis.

As the author, I furthermore declare that, with respect to the creation of this thesis, I have not infringed any copyright or violated anyone's personal and/or ownership rights. In this context, I am fully aware of the consequences of breaking Regulation § 11 of the Copyright Act No. 121/2000 Coll. of the Czech Republic, as amended, and of any breach of rights related to intellectual property or introduced within amendments to relevant Acts such as the Intellectual Property Act or the Criminal Code, Act No. 40/2009 Coll. of the Czech Republic, Section 2, Head VI, Part 4.

Brno .....

.....  
author's signature\*

---

\*The author signs only in the printed version.



## ACKNOWLEDGEMENT

First of all, I would like to thank European Union's Horizon 2020 Research and Innovation programme under the Marie Skłodowska Curie grant agreement No. 813278 (A-WEAR: A network for dynamic wearable applications with privacy constraints, <http://www.a-wear.eu/>). This grant allowed me to conduct my research as part of a joint degree program at the Department of Telecommunications, Faculty of Electrical Engineering and Communication, the Brno University of Technology and Unit of Electrical Engineering, Faculty of Information Technology and Communication Sciences, Tampere University. The research was carried out between 2019 and 2022.

I want to express my deepest gratitude to my supervisors from the Brno University of Technology Assoc. Prof. Jiri Hosek likewise Assoc. Prof. Radim Burget for their support, valuable comments, ideas, and advice. I would like also to thank supervisor Prof. Evgeny Kucheryavy from Tampere University for advising.

Additionally, I truly appreciate the effort and support of Prof. Simona Elena Lohan and Dr. Aleksandr Ometov who coordinated the A-WEAR project and advised us. I want to express additionally my sincere thanks to all my colleagues from the A-WEAR project, especially to Asma Channa and Prof. Nirvana Popescu with whom I have the opportunity to cooperate. Additionally, I would like to thank Salwa Saafi, Raul Casanova Marques, and Jiri Pokorny with whom I have spent a great time together at the Brno University of Technology.

I have been also fortunate to cooperate with Dr. Jiri Mekyska and the Brain Diseases Analysis Laboratory (BDALab) from the Brno University of Technology. It has been a pleasure to discuss research topics and worked on the data shared by BDALab.

Finally, taking this opportunity, I would like to thank my family, my parents, and my friends for all their support and encouragement.

I would like also to mention and thank the Evidation 2019-2020 Participatory ILL Surveillance Program developed by Evidation Health and described in Synapse DOI: 10.7303/syn22891469 for the possibility to use the dataset shared by this initiative.



# Contents

Symbols and abbreviations	19
<b>1 Introduction</b>	<b>25</b>
1.1 Research Motivation . . . . .	25
1.2 Research objectives and methodology . . . . .	29
1.3 Dissertation Scope and Research Tasks . . . . .	31
1.4 Dissertation Outline and Main Results . . . . .	33
<b>2 Background</b>	<b>35</b>
2.1 COVID-19 Pandemic and Possibility for Detecting Disease . . . . .	35
2.2 Parkinson's Disease and the Methods of its Detection . . . . .	37
2.3 EEG, Time Series, and Methods for its Detection . . . . .	42
2.4 Conclusion . . . . .	44
<b>3 The state-of-the-art overview</b>	<b>45</b>
3.1 COVID-19 Detection with the Usage of Wearables and Machine Learning . . . . .	45
3.2 Parkinson's Disease Detection based on Symptoms . . . . .	53
3.2.1 Hypomimia . . . . .	53
3.2.2 Dysarthria . . . . .	60
3.2.3 Methods for Parkinson's Disease Detection based on Sleep Disorders and Actigraphy . . . . .	63
3.3 Deep learning and Time Series Analysis for EEG Analysis . . . . .	72
3.3.1 Literature Overview . . . . .	72
3.3.2 Casual Inference from Neural Time Series Data . . . . .	73
3.3.3 Possibility of Reducing the Number of Training Samples and Increasing Accuracy thanks to the ODE . . . . .	74
3.4 Conclusion . . . . .	75
<b>4 Wearables for COVID-19 detection - Practical Solutions</b>	<b>77</b>
4.1 COVID-19 Diagnosis at Early Stage Based on Smartwatches and Machine Learning Techniques . . . . .	78
4.1.1 Data Characterization . . . . .	78
4.1.2 Feature Extraction and Machine Learning . . . . .	79
4.1.3 Results . . . . .	81
4.1.4 Discussion and Summarisation . . . . .	86
4.2 The Distinction between COVID-19 Cases and Two Types of Influenza with Wearable Devices and Machine Learning . . . . .	89

4.2.1	Data Characterization . . . . .	90
4.2.2	Feature Extraction and Machine Learning . . . . .	91
4.2.3	Results . . . . .	91
4.2.4	Discussion and Summarisation . . . . .	93
4.3	Wearable Analytics and Early Diagnostic of COVID-19 Based on Two Cohorts . . . . .	95
4.3.1	Dataset Characterization . . . . .	96
4.3.2	Feature Extraction and Machine Learning . . . . .	97
4.3.3	Results . . . . .	98
4.3.4	Discussion and Summarisation . . . . .	101
4.4	Conclusion . . . . .	103
<b>5</b>	<b>mHealth Dedicated Solutions for Parkinson’s Detection</b>	<b>109</b>
5.1	Parkinson’s Disease Detection based on Changes of Emotions during Speech . . . . .	109
5.1.1	Dataset Characterization . . . . .	110
5.1.2	Feature extraction and Machine Learning . . . . .	111
5.1.3	Results . . . . .	111
5.1.4	Discussion and Summarisation . . . . .	114
5.2	Multimodal Detection of Parkinson Disease . . . . .	115
5.2.1	Dataset Characterization . . . . .	115
5.2.2	Feature Extraction . . . . .	118
5.2.3	Statistical Evaluation with Machine Learning . . . . .	124
5.2.4	Results . . . . .	124
5.2.5	Discussion and Summarisation . . . . .	130
5.3	Conclusion . . . . .	136
<b>6</b>	<b>Final Conclusion</b>	<b>139</b>
	<b>Bibliography</b>	<b>145</b>

# List of Figures

1.1	The COVID-19 development [2]. . . . .	27
2.1	The COVID-19 stages and potential therapies [3]. . . . .	36
2.2	The motor and non-motor symptoms of <i>Parkinson's disease</i> (PD). . . . .	38
2.3	The domains in which <i>artificial intelligence</i> (AI) finds application in PD [4]. . . . .	40
2.4	The parameters measured during <i>polysomnography</i> (PSG). . . . .	42
2.5	The applications of <i>electroencephalography</i> (EEG). . . . .	43
3.1	The illness duration for COVID-19, Influenza before the main pandemic and during the pandemic [1]. . . . .	47
3.2	The scheme of the neural network CDNet [5]. . . . .	50
3.3	The process flow of the approach in [6]. . . . .	55
3.4	The flow of the experiment from [7]. . . . .	58
3.5	The scheme of the used transfer learning in [8]. . . . .	62
3.6	The bidirectional LSTM [9]. . . . .	67
3.7	The architecture for real-time detection of sleep with the usage of LSTM [9]. . . . .	67
3.8	The scheme of the <i>convolutional neural network</i> (CNN) used in [10]. . . . .	69
3.9	The scheme of the architecture used in [11]. . . . .	73
4.1	Experiment scheme. . . . .	79
4.2	An outline of the feature extraction process for cases of HC and COVID-19. . . . .	80
4.3	The flow of the algorithm . . . . .	90
4.4	Flow of the algorithm. . . . .	96
4.5	Feature extraction. . . . .	97
5.1	The scheme of the experiment. . . . .	110
5.2	Emotional recognition based on the face image. (The image has been blurred for privacy reasons). . . . .	112
5.3	An analysis of the changes in emotions during a speech exercise. . . . .	112
5.4	The values of SHAP were derived from the XGBoost tongue twister model. . . . .	114
5.5	Flow of the algorithm. . . . .	116
5.6	Kernel Density Estimation of level of UPDRS III. . . . .	117
5.7	Kernel Density Estimation of duration of PD. . . . .	118
5.8	Flow of the facial features extraction. . . . .	119
5.9	Illustration of facial features [12]. . . . .	119
5.10	SHAP values for the best 10 features from the video modality. . . . .	126
5.11	SHAP values for the best 10 features from the audio approach. . . . .	127

5.12 SHAP values for the best 10 features from the multimodality. . . . .	128
5.13 SHAP values for the optimal video approach (TSK39). . . . .	131
5.14 SHAP values for the optimal audio approach (TSK7). . . . .	131
5.15 SHAP values for the optimal multimodal approach (TSK41). . . . .	132



# List of Tables

3.1	An overview of the methods of COVID-19 detection and pandemic development with the wearables and AI. [2]. . . . .	52
3.2	Overview of the works analysing hypomimia in patients suffering from PD. . . . .	59
3.3	Potential methodologies for detecting PD based on sleep disorders [13].	71
4.1	The scenario in which the experiment was conducted. . . . .	82
4.2	Statistical analysis of COVID-19 cases vs. HC based on Mann-Whitney U test with FDR correction. . . . .	82
4.3	Statistical analysis of COVID-19 cases, Influenza vs. HC based on Mann-Whitney U test with FDR correction. . . . .	83
4.4	COVID-19 disease detection results for 5-day windows (cohorts: 27 HC, 27 COV). . . . .	84
4.5	Detection of COVID-19 disease and the presence of Influenza cases within a 5-day window (cohorts: 34 HC, 27 COV, and Influenza 7). . . . .	85
4.6	COVID-19 detection results for 7-day windows (cohort: 26 HC, 26 COV). . . . .	85
4.7	The results of the COVID-19 disease detection for 10-day windows (cohorts: 24 HC, 24 COV). . . . .	86
4.8	Detection of COVID-19 disease as well as Influenza cases within a 7-day window (cohort: 33 HC, 26 COV, Influenza 7). . . . .	86
4.9	Detection results for COVID-19 disease and Influenza cases for 10-day windows (cohort: 31 HC, 24 COV, Influenza 7). . . . .	86
4.10	Identifying Influenza cases during the middle of the pandemic and cases of COVID-19. . . . .	92
4.11	Identifying Influenza cases before and during the pandemic. . . . .	92
4.12	Identifying COVID-19 cases and Influenza cases prior to the main pandemic. . . . .	93
4.13	Result of multiclass classification for COVID-19, Influenza cases prior to the main pandemic, and Influenza during the main pandemic. . . . .	93
4.14	The combinations of datasets and classes for each experiment. . . . .	95
4.15	Dataset mixtures and sample numbers for each experiment. . . . .	97
4.16	Analysis of COVID-19 cases, and HC cases using the Mann-Whitney U test and FDR correction. . . . .	99
4.17	Analysis of ill and HC cases using the Mann-Whitney U test and FDR correction. . . . .	99
4.18	The outcome of distinction COVID-19 (A and B dataset) from HC (for all 36 features). . . . .	100

4.19	The outcome of distinction COVID-19 (A and B dataset) from HC (for selected 20 features).	100
4.20	The outcome of distinction between COVID-19, Influenza vs. HC based on dataset A.	100
4.21	The outcome of distinction COVID-19 (A and B dataset) vs. Influenza (all analysed cases, from A and B dataset)	101
4.22	The outcome of distinction of COVID-19 (A and B dataset), Influenza (for all cases, from A and B dataset) vs. HC.	101
4.23	The outcome of multiclass classification of COVID-19, Influenza and HC for A and B datasets.	102
4.24	The overview of the achieved outcomes in Sections 4.1, 4.2, and 4.3.	105
5.1	Statistical analysis of the created features among PD and HC.	113
5.2	The prediction for tongue twister based on the various classifiers.	113
5.3	XGBoost predictions for tongue twister and reading text exercise.	114
5.4	Clinical and demographical information of the dataset.	117
5.5	Features extraction explanation.	120
5.6	Description of the acoustic features. The details of the features implementation are provided in [14].	121
5.7	Meaning of the part of the exercises in Czech and English.	121
5.8	Definition of the acoustic features in detailed.	122
5.9	Carried-out vocal tasks.	123
5.10	The statistical analysis of audio features.	124
5.11	The statistical analysis of video features.	125
5.12	Accuracy of PD detection from different modalities.	125
5.13	Video-based speech exercise accuracy.	128
5.14	An assessment of the accuracy of the best speech exercises based on audio recordings.	129
5.15	Evaluation of the accuracy of multimodal speech exercises.	129
5.16	An analysis of the results obtained from multimodal and video approaches.	130

# Symbols and abbreviations

<b>3D</b>	three dimensional
<b>1-D</b>	one dimensional
<b>ACE-R</b>	Addenbrooke's Cognitive Examination-Revised
<b>AKV</b>	absolute kinematic velocity
<b>AU</b>	Action Unit
<b>AAL</b>	Ambient Assisted Living
<b>ae</b>	approximate entropy
<b>AUC</b>	Area Under the Curve
<b>AUROC</b>	area under the receiver operating characteristic
<b>ADHD</b>	attention-deficit hyperactivity
<b>AE</b>	autoencoder
<b>AI</b>	artificial intelligence
<b>BDI</b>	Beck Depression Inventory
<b>BLSTM</b>	Bidirectional Long Short-Term Memory
<b>BVP</b>	Blood Volume Pulse
<b>BLE</b>	Bluetooth Low Energy
<b>BMI</b>	Body Mass Index
<b>BCI</b>	brain-computer interface
<b>CT</b>	computer tomography
<b>CNN</b>	convolutional neural network
<b>CuSum</b>	cumulative sum
<b>CDI</b>	decompensation index
<b>DL</b>	deep learning
<b>DNN</b>	deep neural networks

<b>DX</b>	Dysarthria Index
<b>DBN</b>	deep belief network
<b>ECG</b>	electrocardiography
<b>EEG</b>	electroencephalography
<b>EMG</b>	electromyography
<b>EOG</b>	electrooculography
<b>eHealth</b>	Electronic Health
<b>ECDF</b>	empirical Cumulative Distribution Function
<b>EU</b>	European Union
<b>EDS</b>	Excessive Daytime Sleepiness
<b>XAI</b>	explainable artificial intelligence
<b>FEC</b>	Face expression recognition
<b>FACS</b>	facial action coding system
<b>fEMG</b>	facial electromyography
<b>FECF</b>	facial expression change factor
<b>FEF</b>	facial expression factor
<b>FER</b>	facial expression recognition
<b>FDR</b>	false discovery rate
<b>FFT</b>	Fast Fourier Transform
<b>FI</b>	fragmentation index
<b>FPS</b>	frame per second
<b>FOG</b>	Freezing of Gait
<b>FC</b>	fully-connected layers
<b>GSR</b>	galvanic skin response
<b>GRU</b>	Gated Recurrent Units

<b>GLVQ</b>	Generalised Learning Vector Quantisation
<b>GAN</b>	generative adversarial network
<b>GRAD-CAM</b>	Gradient-weighted Class Activation Mapping
<b>HC</b>	healthy control
<b>HROS</b>	ratio of heart rate to the number of steps taken
<b>HROS-AD</b>	heart rate over steps anomaly detection
<b>HR</b>	heart rate
<b>HRV</b>	Heart Rate Variability
<b>HBNN-C</b>	Hierarchical Bayesian neural network
<b>HOG</b>	histogram of oriented gradients
<b>HD</b>	hypokinetic dysarthria
<b>iRBD</b>	idiopathic rapid eye movement sleep behavior disorder
<b>ICT</b>	Information and Communication Technology
<b>IBI</b>	inter-beat interval
<b>IoT</b>	Internet of Things
<b>IoMT</b>	Internet of Medical Things
<b>IQR</b>	interquartile range
<b>k-NN</b>	k-Nearest Neighbour
<b>LED</b>	Levodopa Equivalent Dose
<b>LDA</b>	Linear Discriminant Analysis
<b>LPCC</b>	linear prediction cepstral coefficients
<b>LSTM</b>	Long-Short Term Memory Network
<b>LSTM-VAE</b>	Long Short-term Memory Variational Autoencoder
<b>ML</b>	machine learning
<b>MRI</b>	magnetic resonance imaging

<b>MCC</b>	Matthews correlation coefficient
<b>MAX</b>	maximally discriminative facial movement coding systems
<b>max</b>	maximum
<b>mRMR</b>	Maximum Relevance Minimum Redundancy
<b>MAE</b>	Mean Absolute Error
<b>MFCC</b>	Mel-frequency cepstral coefficients
<b>MET</b>	metabolic equivalents
<b>MMSE</b>	Mini-Mental State Examination
<b>min</b>	minimum
<b>mHealth</b>	mobile health
<b>MASS</b>	Montreal Archive of Sleep Studies
<b>MTCNN</b>	Multitask Cascaded Convolutional Networks
<b>NMSS</b>	Non-Motor Symptoms Scale
<b>NN</b>	neural network
<b>NREM</b>	non-Rapid Eye Movement
<b>OPTN</b>	Ordinal Partition Transition Networks
<b>ODE</b>	ordinary differential equation
<b>OC-SVM</b>	One Class-Support Vector Machine
<b>pVal</b>	p-value
<b>PD</b>	Parkinson's disease
<b>PPG</b>	photoplethysmography
<b>PSG</b>	polysomnography
<b>PET</b>	positron emission tomography
<b>PCA</b>	Principal Component Analysis
<b>PCS</b>	Progressive Confidence Strategy

<b>PIM</b>	proportional-integral mode
<b>RF</b>	radiofrequency
<b>RBD</b>	Rapid Eye Movement Behavioral Disorder
<b>REM</b>	rapid eye movement
<b>RGB</b>	red, green, blue
<b>RNN</b>	recurrent neural network
<b>RLR</b>	Regularized logistic regression
<b>rsd</b>	relative standard deviation
<b>RBDSQ</b>	REM sleep behavior disorder screening questionnaire
<b>RR</b>	respiration rate
<b>RHR</b>	Resting Heart Rate
<b>RHR-Diff</b>	RHR Difference
<b>RLS</b>	Restless Leg Syndrome
<b>RT-PCR</b>	reverse transcription-polymerase chain reaction
<b>RMS</b>	root mean square
<b>RMSSD</b>	root mean square of successive differences
<b>SFHR-NET</b>	Semantic Feature based Hypomimia Recognition Network
<b>SF-C</b>	Semantic Feature Classifier
<b>SJTU</b>	Shanghai Jiao Tong University
<b>se</b>	Shannon entropy
<b>SHAP</b>	SHapley Additive exPlanations
<b>STFT</b>	short-time Fourier transform
<b>SNR</b>	signal-to-noise ratio
<b>SE</b>	sleep efficiency
<b>SOL</b>	sleep onset latency

<b>std</b>	standard deviation
<b>SGD</b>	Stochastic gradient descent
<b>SMOTE</b>	Synthetic Minority Oversampling Technique
<b>SSDA</b>	stacked sparse denoising autoencoder
<b>SVM</b>	Support Vector Machines
<b>TAT</b>	time above threshold
<b>TICC</b>	Time-aware Toeplitz Inverse Covariance-based Clustering
<b>UPDRS</b>	Unified Parkinson's Disease Rating Scale
<b>USA</b>	United States of America
<b>var</b>	variance
<b>VSG</b>	videosomnography
<b>VGG</b>	Visual Geometry Group
<b>WASO</b>	wake after sleep onset
<b>WHT</b>	wearable health technology
<b>WHO</b>	World Health Organization
<b>ZCM</b>	zero-crossing mode



# 1 Introduction

The COVID-19 pandemic and the aging society are considered the biggest issues which Europe faces [15, 16]. They are emergency problems on a large scale that need fast taking care of them - developing methodologies destined for diagnosis diseases such as COVID-19 and PD. The COVID-19 pandemic has caused a plethora of deaths and this disease has a high contagiousness rate [17]. Screening tests of society are highly desirable and they should be easily approachable and destined to be used in the early stage to limit the spreading of the disease. The screening tests are defined as the procedure used to get to know if the examined person has a disease before it will manifest visible symptoms. Moreover, the aging society carries neurodegenerative diseases and PD belongs to this group of diseases [15]. The most accurate test is the *positron emission tomography* (PET), *magnetic resonance imaging* (MRI) and *computer tomography* (CT), however, those methods are used in the advanced stage of the disease and are expensive [18, 19]. The usage of wearables and solutions based on *mobile health* (mHealth) and *Electronic Health* (eHealth) concepts seems to be justified for PD detection, but also for COVID-19 recognition. Wearables are electronic devices which are relatively inexpensive and accessible [20, 21]. Furthermore, the utility of *machine learning* (ML) allows for creating the support system methodologies which could predict the occurrence of the illness [2, 22]. Additionally, the usage of novel *explainable artificial intelligence* (XAI) can provide the clinical interpretability of the created models [23].

## 1.1 Research Motivation

The wearable devices are electronic devices which could sense, gather and upload the data [21]. The wearables can be classified as on-body (e.g., smartwatches), in-body (implants [24]), and also around-body wearables (mobile phones, smart-cards) [25, 26]. It is expected that the wearable market will be growing increasingly. This market is characterised by a high pace of progression and exponential growth is predicted in the coming years. The price of the market was estimated at 71.91 billion \$ in 2023 [27, 28]. The interesting niche is the *wearable health technology* (WHT) global market. The WHT global market achieved 16 billion \$ in 2021 [29]. Nevertheless, the WHT is an atypical market because it characterises the conditions of two markets, not just one, i.e., the healthcare and technology market. This multidisciplinary is generating an unique opportunity for the implementation of new approaches. Moreover, the existing neurodegenerative and chronic diseases occurring among the elderly people are creating the need for solutions from WHT market [30]. Additionally, WHT offers a big potential in fighting with many diseases including

COVID-19 disease. The support system methodologies trained with the usage of ML and data gathered by the wearables could serve as an extra diagnostic/monitoring tool to determine the occurrence of the diseases or their progress [31, 32]. Furthermore, the Universal Design File is going to be commonly used for wearables [33] which allows using wearable technology together with artificial intelligence as support methodologies widely in society. This concept aims to facilitate the usage of the considered item to all people of all sizes, ages, and in all health conditions, and it is destined to all buildings [34]. Some of the principles of Universal Design are, for instance: flexibility in use, simple and intuitive use, and low physical effort. Those rules are making the wearable an attractive device for the consumer. Additionally, the low-cost and easy accessibility for the population makes wearables a potentially valuable screening tool [24].

The data acquired by wearable devices are predominantly of the time series type [35]. This kind of data requires appropriate architectures to extract information from them. Based on the data, there could be created support decision methodologies that serve in classification, forecasting, or anomaly detection problems [36, 37]. For this purpose, ML algorithms are commonly utilised [38]. Additionally, the step of the data pre-processing is equally important as the further steps [39]. Depending on the data type - structured or unstructured, the approaches of ML methodologies could be appropriately selected [40]. To use some groups of the ML algorithms for the detection (for instance disease) purpose such as *Support Vector Machines* (SVM), XGBoost, *k-Nearest Neighbour* (k-NN), Random Forest, Decision Tree [41–43], the data for the input should be provided in the structured form. Moreover, the usage of neural networks is appropriate for the raw time series, i.e., unstructured data. They are especially suitable because of the possibility to learn and use long-term dependencies. The examples of the neural networks which are commonly used for the time series are *one dimensional* (1-D) CNN, *Long-Short Term Memory Network* (LSTM), *Gated Recurrent Units* (GRU), *Bidirectional Long Short-Term Memory* (BLSTM). Additionally, the utility of transfer learning allows for achieving the state-of-the-art-results [44–50]. With the usage of the aforementioned solutions, it is feasible to create support system methodologies, including those for healthcare, that are even highly accurate. However, that is not all. To fully understand the broader applications of umbrella terms of healthcare systems, wearables and ML, the following existing concepts need to be introduced first: eHealth, mHealth, and *Internet of Medical Things* (IoMT).

eHealth and mHealth are using wearables, multimedia (electronically distributed mix of media containing audio, images, video, and text), and communication systems technologies for providing health technologies [51]. The following key advantages of eHealth and mHealth could be listed: ease of use, lifetime monitoring, cost reduction

of the healthcare system, and data analytics. IoMT is referred to as a network of smart things - medical devices, people, sharing healthcare data. IoMT could sense, process the data, network, and communicate [52].

These promising methods and concepts (eHealth, mHealth, and IoMT) could find applications for specific health problems which the population is dealing with nowadays and probably will need solutions in the future [51].

Official confirmation of Coronavirus has been announced on 29 December 2019 by the *World Health Organization* (WHO) in China. This disease has a relatively high value of the basic reproduction ( $R_0$ ) range. It is estimated between 2.6 to 4.71, which in combination with the relatively high death rate is the reason why the disease has become one of the most deadly pandemics in history [17]. Thereby, screening and testing of COVID-19-positive people are nowadays considered to be one of the most effective ways how to stop or limit the further spreading of the infection likewise eliminate the danger of renovating the high passed so far state of emergency. Wearables open doors to completely new ways of how the health status can be monitored and possibly how to recognise the disease in its early stage. In the case of COVID-19, the early detection of this illness is of high importance, since the disease is communicable approximately two days before the first symptoms. The development of the disease is illustrated in Fig. 1.1. Some recent works report that wearable devices which are already available on the market, in particular the sensors of which they are equipped, can be efficiently used for monitoring the diseases such as influenza.

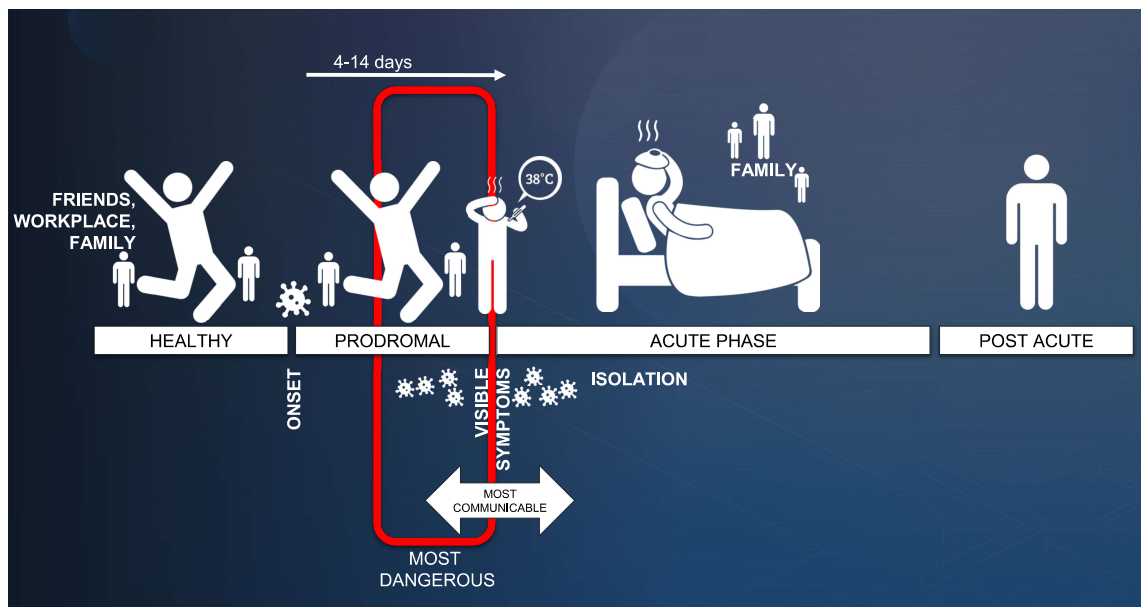


Fig. 1.1: The COVID-19 development [2].

The question is where else the wearable together with ML could find broader and needed applications. One of the concepts deserves extra attention. *Ambient Assisted Living* (AAL) is defined as the usage of *Internet of Things* (IoT) and *Information and Communication Technology* (ICT) for home healthcare. The idea is of high importance in the view of the aging society. The standard medical practice is creating a big burden for the economy of the healthcare system. By the same token, less expensive and more approachable technologies need to be introduced. The usage of wearables and digital technologies could support ‘enabling aging in place’. The elders could live in their domestic environments, with the community to which they are accustomed. This approach is supported by society and has a positive influence on the elders. Additionally, the maintenance costs of the wearables-aided healthcare systems are decreased [53]. Furthermore, wearables could prevent emergencies and mortality rates, which will be greatly appreciated by patients, their families, and medical staff [54]. Those aforementioned cases are regarded as highly demanded especially in the still-developing countries because of the limited number of institutions to care for the elders and lacking financial resources [55]. The common examples of the utilisation of ML in AAL applications are: human activity recognition, monitoring and forecasting diseases, and indoor and outdoor localization of elders [56]. The scope of the algorithms used for this purpose is really broad: from k-NN, SVM, Naive Bayes, Random Forest, SVM, to neural networks like CNN, LSTM, and others [56]. The special target of this thesis will be PD. This illness is the second most common neurodegenerative disorder with a prevalence of 2 % for people over the age of 65 years [57].

This thesis is focused on the usage of ML techniques together with the wearables for COVID-19 detection, and creating methods of AAL dedicated to recognise PD. These techniques represent a big promise for new innovative solutions in the mHealth and eHealth areas and have the potential to form the future of health care. To develop ML models and train them, three datasets are utilised for creating support system methodologies, mHealth and eHealth solutions. Two of them represent the records of COVID-19 cases, Influenza and *healthy control* (HC) group. HC is regarded in clinical studies as a person who does not have the illness or disorder being studied, however, this person could suffer from other diseases [58]. Those data were collected thanks to the Fitbit device and contains records of the heart rate and activity of the person - the number of steps taken. The third dataset represents the records of PD patients and HC. The dataset contains video and audio records. The symptoms of PD - hypomimia and *hypokinetic dysarthria* (HD) are computationally analysed.

The conducted COVID-19 detections consider the character of the disease, i.e., the contagiousness of the disease and incubation period. Taking into account those

parameters allows for the detection of the disease in the early stage. In addition, the distinction between the diseases, i.e., COVID-19 and Influenza, is possible thanks to the existing representation of the Influenza cases in the dataset. The target of the practical part of the thesis is also not only to design the support system methodologies but also to present the clinical interpretability of the models. They are provided for COVID-19 and PD detection thanks to the statistical analysis and usage of *SHapley Additive exPlanations* (SHAP) values. In addition, the creation of several models is the scope of the thesis to identify the most accurate of them and to determine the parameters of predictions inter alia such as accuracy, sensitivity, and specificity.

Moreover, the aim of this work is to analyse the possibility to detect PD detection based on hypomimia and HD motor symptoms. The aforementioned dataset contains video and audio records. 43 unique clinical speech exercises are used to detect PD. The utility of the whole spectrum of speech exercises allows for the identification of the most suitable task for automatic PD detection in clinical practice. Furthermore, the multimodality approach of PD detection is explored, i.e., the combination of audio and video modality. Additionally, the prediction models generated thanks to the single modality are compared to those created for the multimodal approach to identify if the combinations of selected modalities could achieve better results. Moreover, the possibility of PD detection is evaluated for emotion recognition tasks between the groups. It is justified by the fact that PD patients manifest impairments in expressing emotions. Furthermore, the thesis provides the theoretical basements of the conducted experiments, likewise describes the transferable methodologies used for a spectrum of diseases, and which are suitable for PD recognition based on sleep disorders symptoms. Moreover, the work presents the approaches destined for EEG analysis. The characteristic of EEG signals and the application of EEG in diagnosis were presented. There are illustrated approaches with the usage of deep learning methods likewise the novelty: neural *ordinary differential equation* (ODE). ODEs are regarded as neural networks having big potential and they could be applied to wearable-related data.

## 1.2 Research objectives and methodology

The subject of this thesis is correlated to the usage of machine learning and wearables for the detection and monitoring diseases thanks to the usage of machine learning and wearables. The particularly considered topics are COVID-19 detection and neurodegenerative diseases like PD likewise EEG analysis. Thereby, the following seven main Research Objectives (ROx) in this thesis have been identified with their

related Research Questions (RQx). The achieved research tasks are depicted in Section 1.3. They are the answers to the Research Objectives and Research Tasks.

**RO1. Classification of COVID-19 cases thanks to the wearable-related data: heart rate and number of steps taken**

- **RQ1.1.** Is it possible to improve the results presented in [59] based on this same dataset and proposed classification model instead of just anomaly detection? How accurate would be achieved the machine learning model?
- **RQ1.2.** What kind of features would be the most appropriate for COVID-19 detection?
- **RQ1.3.** How to develop a successful methodology for early COVID-19 detection concerning the nature of the disease?

**RO2. Differentiation COVID-19 patients from Influenza cases based on wearable data**

- **RQ2.1.** What will be the accuracy of detecting COVID-19 cases and Influenza cases that happened before the main pandemic or in the middle of the pandemic?
- **RQ2.2.** How accurate will be the distinction between Influenza cases before the main pandemic versus Influenza in the middle of the pandemic?
- **RQ2.3.** How to design an algorithm for support methodology to detect COVID-19?

**RO3. The distinction of COVID-19 patients from Influenza cases and HC thanks to the wearable data and two datasets**

- **RQ3.1.** How to develop a suitable support methodology to detect COVID-19 based on two different datasets?
- **RQ3.2.** How accurate will the prediction of COVID-19 cases and ill cases be based on two various datasets?
- **RQ3.3.** Which features will be the most beneficial to detect COVID-19 cases?

**RO4. Recognition of PD hinged on facial expression impairments and classification of emotions**

- **RQ4.1.** Which speech exercise: tongue twister or reading poem will be more suitable for PD detection based on hypomimia and classification of emotions?
- **RQ4.2.** Which features based on emotion will be the most valuable for PD detection?
- **RQ4.3.** How accurate will the PD detection model be based on hypomimia and impairments in expressing emotions?

**RO5. Recognition of PD thanks to the multimodality – audio and video**

- **RQ5.1.** Which biomarkers are the most significant for PD detection?
- **RQ5.2.** Which Czech speech exercise is the most powerful for PD detection?

- **RQ5.3.** What will be the highest obtained accuracy for PD detection? What are the components of this solution?

**RO6. The review of the transferable methodologies of detection of sleep disorders thanks to the actigraphy device for Parkinson’s disease detection**

- **RQ6.1.** Which methodologies used for sleep disorders and other illnesses could be applied for PD detection based on sleep abnormalities and actigraph device?
- **RQ6.2.** What kind of diseases could have suitable biomarkers appropriate for PD detection based on sleep disorders and actigraph?
- **RQ6.3.** What parameters could characterise sleep disorders in PD?

**RO7. The review of the application of deep learning techniques in the EEG analysis**

- **RQ7.1.** What characterises the EEG?
- **RQ7.2.** Which deep learning methods are suitable for EEG analysis?
- **RQ7.3.** What kind of methods will be suitable for EEG analysis based on wearable data?

## 1.3 Dissertation Scope and Research Tasks

Considering the discussion in subsection 1.1 the potential of the AAL solution is profound. Parallely and unexpectedly the COVID-19 pandemic became also one of the biggest issues in the current world. Emerging technologies like ML and commonly available wearables could serve as enablers to effectively deal with the problems with which the world is struggling nowadays. Moreover, the combination of them could be regarded as a paradigm shift in the field of eHealth solutions. Additionally, those technologies open the doors for faster screening of society and creating the support methodologies which could serve doctors as the assisting tools and facilitate the lives of the patients and infected people thanks to the limitation of necessary appointments at physicians, and hospitals, likewise more approachable screening for the ill person. Furthermore, the created solutions increase turnover for the sector of the silver economy and healthcare [60].

The only way to evaluate the considered support methodologies is the training and testing the proposed models with retaining the principles of the ML techniques. The crucial part is also the gathering and usage of suitable datasets. Those tasks were discussed in this thesis from theoretical as well as experimental points of view.

To emphasise, this thesis provides an overview of created support methodologies for the ML-based detection of COVID-19, and PD. The developed models are distinguished into five subsequent studies. Furthermore, the theoretical discussion of

the transferable methodologies of detection of sleep disorders thanks to the actigraphy device for PD detection likewise the application of deep learning techniques in the EEG analysis were conducted. The following research tasks based on ML and answers to Research Questions and Objectives from 1.3 were carried out:

- **COVID-19 detection based on records of heart rate and activity** – the proposed classification models were trained on a publicly available dataset introduced in [59]. The models were optimised to detect the differences between the healthy and ill states of the participants thanks to the analysis of spectral, frequency, and statistic features. Moreover, the statistical analysis of the proposed features was provided. A few ML models were tested and the outcomes of the prediction were illustrated [2].
- **COVID-19 distinction from Influenza cases** – the introduced ML models classified the cases between COVID-19 cases and two types of Influenza considering the *heart rate* (HR) and personal activity. The analysis was conducted based on the claim from [1] about occurring differences between the illnesses [61].
- **COVID-19 differentiation from HC and Influenza based on two datasets** – the introduced models were malleable on two datasets ( [1, 59]) to obtain a bigger cohort and to consider Influenza in the classification tasks. One of the datasets was undersampled to combine both of them. The statistical analysis is presented and several classification models were depicted to distinguish COVID-19 cases from HC, likewise a few types of Influenza from COVID-19, and ill cases from HC [62]. Furthermore, the multiclass classification was considered.
- **PD detection based on hypomimia and emotion recognition** – Thanks to the automatic analysis of changes in emotion during doing Czech speech exercises by participants, the ML models were developed for PD detection. The two exercises which were analysed were difficult-to-pronounce tongue twister likewise reading poem. A few classifiers were tested for this purpose. The interpretability of the best model was illustrated by SHAP values. Additionally, the statistical analysis of the generated features was presented [63].
- **PD detection based on multimodal approach (video and audio)** – the interpretable support methodology of PD detection was created based on hypomimia and HD symptoms. The tasks used for the evaluation of the disease were 43 Czech speech exercises. The created facial features based on detected facial landmarks together with the audio features were used for the model trained by the XGBoost classifier. Additionally, both the video and audio features were examined to check their statistical significance. The interpretability of the model was given thanks to the usage of SHAP values.



## 1.4 Dissertation Outline and Main Results

This doctoral thesis is the descriptive outcome of the study funded from European Union's Horizon 2020 Research and Innovation programme under the Marie Skłodowska Curie grant agreement No. 813278 (A-WEAR: A network for dynamic wearable applications with privacy constraints, <http://www.a-wear.eu/>). The described research was conducted between 2019 and 2022. The most important part of the research is included in the chapters 2 - 6 based on the published conference and journal papers.

The structure of the thesis is starting from the introduction, the theoretical background of the research, the state-of-the-art related to the topics analysed and described in the thesis likewise the instances of the conducted research about COVID-19 and PD with the usage of ML and wearables.

The first chapter is the introduction of the thesis (Chapter 1), whereas Chapter 2 shows the theoretical basement of the discussed problem in the thesis, such as COVID-19 detection, PD recognition, and EEG analysis as the time series concept.

Furthermore, the Chapter 3 considers the current state-of-the-art issues dealt within the scope of this thesis. The modern solutions for COVID-19 detection utilising wearables and ML are depicted. Moreover, PD detections based on hypomimia, HD, and sleep disorders are presented. Additionally, the possible translations of recognition of sleep disorders based on actigraphy records and ML for PD detection are illustrated. Furthermore, the problem of analysis of EEG time series is broadly described with emphasis on the deep learning methods for this aim.

Chapter 4 and Chapter 5 introduce the practical solutions for the defined issues together with a logical presentation of the carried out research. The Chapter 4 describes three support methodologies developed for the detection of COVID-19 using data from wearables. The first of them is recognising the COVID-19 cases from HC based on HR measurements and personal activity. While the second solution of ML is differentiating the COVID-19 cases from two types of Influenza. The third set of solutions is the models created thanks to the combination of two previously utilised models.

The Chapter 5 considers the automatic detection and interpretation of the presented ML models for PD recognition. The first task is dealing with the automatic recognition of emotions among PD patients and HC thanks to the utilised neural network and based on the variability of their occurrence during the assigned speech exercises, the model can detect PD disease. The second research in this area takes under analysis the symptoms of hypomimia and HD among PD patients. The multimodal solution is provided, which is analysing the facial and voice features.

The final Chapter 6 summarises the whole thesis and provides the highlights of

the main results.

## 2 Background

This chapter contains the theoretical background described in this thesis. The topics relate to diseases such as COVID-19, PD likewise their detections. Moreover, the EEG character and analysis, the utility of this biosignal for diseases recognitions likewise the application of neural networks for EEG. Firstly, the nature of COVID-19 and the symptoms of this illness are mentioned. Additionally, the methods dedicated to its recognition including wearable devices are briefly depicted. Furthermore, the PD is similarly described, i.e., the character of this disease together with the signs of this illness are listed. Some symptoms of PD: hypomimia, HD, and sleep disorders in PD are broadly illustrated. Moreover, the importance of mHealth for PD is explained.

### 2.1 COVID-19 Pandemic and Possibility for Detecting Disease

The pandemic of COVID-19 began in December 2019 [16, 62]. On 11 March 2020, the WHO officially announced the outbreak of the pandemic [20]. Two words are used in the definition of pandemic, pan and demos. Demos refers to the people, and pan refers to everyone. The world faces up to a pandemic, which causes a state of emergency, numerous infections, and deaths likewise the occurring obstacles in everyday life. The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the reason for this illness [61, 64]. Coronavirus belongs to the Coronavirinae and Torovirinae subfamilies of the Coronavirinae family. It is known that Coronaviruses are capable of infecting a wide variety of organisms, including rodents, birds, mammals, and humans [2, 17].

There is a wide range of symptoms associated with COVID-19. The symptoms range from coughing, fever, hoarseness, shortness of breath, chest pain, or abdominal pain [65], as well as the rare loss of smell and taste [66]. Additionally, examining the wearable records revealed that there were changes in HR around the time of onset of symptoms [1]. According to [3], the authors have identified the three stages of COVID-19: the early stage (stage I), the pulmonary phase (stage II), and the hyperinflammation phase (stage III). If the disease is detected at the prodromal stage (stage I), it will have the greatest impact since it represents a time when a person feels healthy but is already infectious, which is resulting in social contact and the spread of the disease to others. As mentioned by the authors [3], this phase is characterized by fever, dry cough, and mild constitutional symptoms. Detection in the stage I prevents further complications, and the duration of the illness is

reduced [3]. The three stages of the disease likewise potential therapies are illustrated in Fig. 2.1.

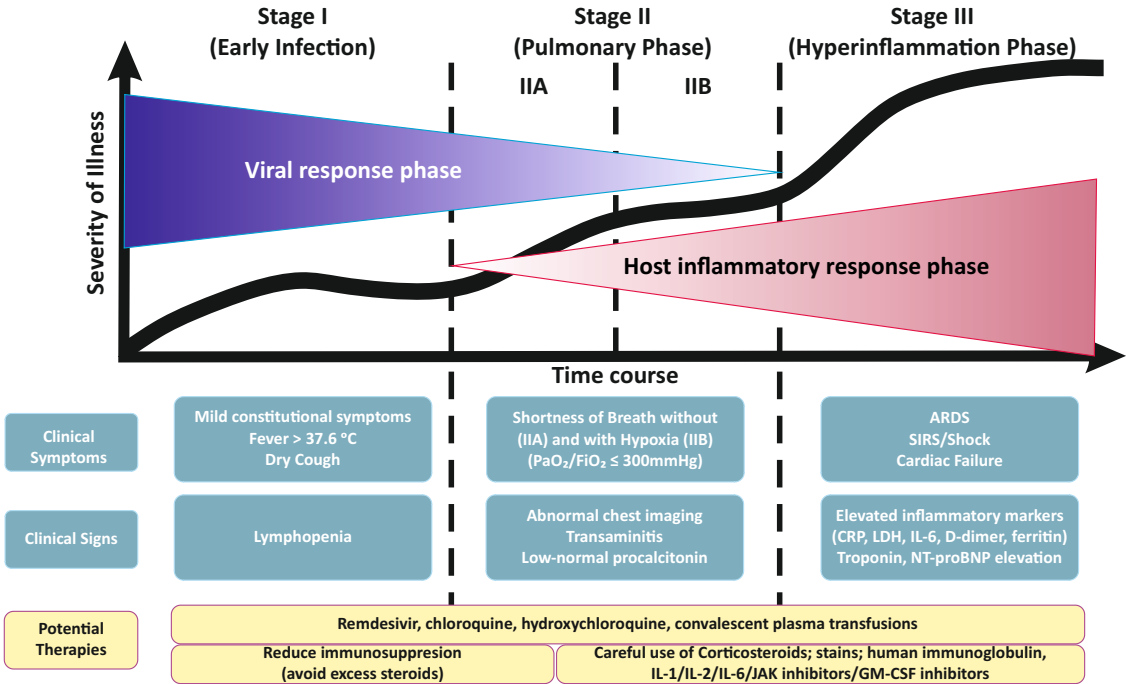


Fig. 2.1: The COVID-19 stages and potential therapies [3].

Additionally, long-term complications have been identified such as cardiovascular, respiratory, as well as neurological problems, in addition to many others that have not yet been fully described [67–69]. There are several risk factors associated with this disease, including civilization diseases, old age, renal dysfunction, and hepatic dysfunction [2,70] In order to limit the spread of the disease, it is ideal to detect the disease before the highest contagious period, which is considered to manifest 2 days prior to the visible onset of the symptoms until 1 day after the onset [71].

It is especially notable that variations in HR are present in COVID-19 cases, and that they persist for a longer period than common influenza. The resting HR is elevated nearly the time when symptoms have started [1]. Moreover, COVID-19 [72] also showed variations from norms during sleep. It was reported that deep sleep was associated with raised *respiration rate* (RR), while *non-Rapid Eye Movement* (NREM) sleep was associated with increased HR at night. In contrast, the *root mean square of successive differences* (RMSSD), as well as the Shannon entropy for nocturnal RR series declined. An analysis of Z-values and 1257 participants wearing Fitbit devices led to this conclusion [72].

The virus is primarily transmitted through social contact, namely face-to-face contact, coughing, talking, or sneezing [73]. This disease has a relatively high value of the basic reproduction ( $R_0$ ) range. It is estimated between 2.6 to 4.71, which in combination with a relatively high death rate, is why the disease became one of the most deadly pandemics in history [17]. The possible solution for controlling social contact to limit the spreading of the virus is the usage of tracing apps, mobile phones and wearables [20]. There are several ways in which the screening test data can be collected. Imaging technologies offer the most accurate diagnostics, even approaching 100 percent in some cases [74]. Furthermore, *reverse transcription-polymerase chain reaction* (RT-PCR) is the most widespread diagnostic, despite being relatively accurate, these methods are typically used after the onset of disease in order to confirm the diagnosis. Nevertheless, wearable devices [20] appear to be the most inexpensive and fastest method of screening a large population. They appear in the population quite widely which makes them a good candidate to be used as a screening test. It is essential that the COVID-19 should be detected two days before onset. During those two days (on average), people are unaware that they are infected, which makes it easy for them to spread the disease. Due to the difficulty of identifying those symptoms, it is not an easy task. Wearable sensors can be used to analyse a variety of physiological parameters, including activity levels, temperature, cardiovascular strain, blood pressure, sleep parameters, respirations variable, sound monitoring, coughing, SpO<sub>2</sub> level, humidity sensors [20,75,76]. Analysing the data is the final step in creating the support methodology - assisting technology for clinical purposes. To sum up, ML holds great promise for the analysis of COVID-19-related data [2,62,76].

## 2.2 Parkinson's Disease and the Methods of its Detection

PD is one of the most prevalent neurodegenerative diseases in society. This disease occurs in 2-3 % of society in the *European Union* (EU) beyond 65 years old [57]. A major challenge that the EU will have to deal with within the next 30 years is the aging of society. This issue is associated with neurodegenerative diseases, and one of them is PD [15]. A distinction can be made between the motor and non-motor symptoms of PD. The manifesting main motor symptoms depend on the progressive loss of dopaminergic neurons in substantia nigra pars compacta [77–79]. Among the motor symptoms could be distinguished the following signs: hypomimia, HD, the *Freezing of Gait* (FOG), bradykinesia, tremor, PD dysgraphia, dyskinesia, dysphagia [80–87]. Sleep disorders, hallucinations, depression, anxiety,

constipation, cognitive deficits, urinary symptoms, etc. belong to the non-motor symptoms [77,88–90]. The motor and non-motor symptoms are listed in Fig. 2.2.

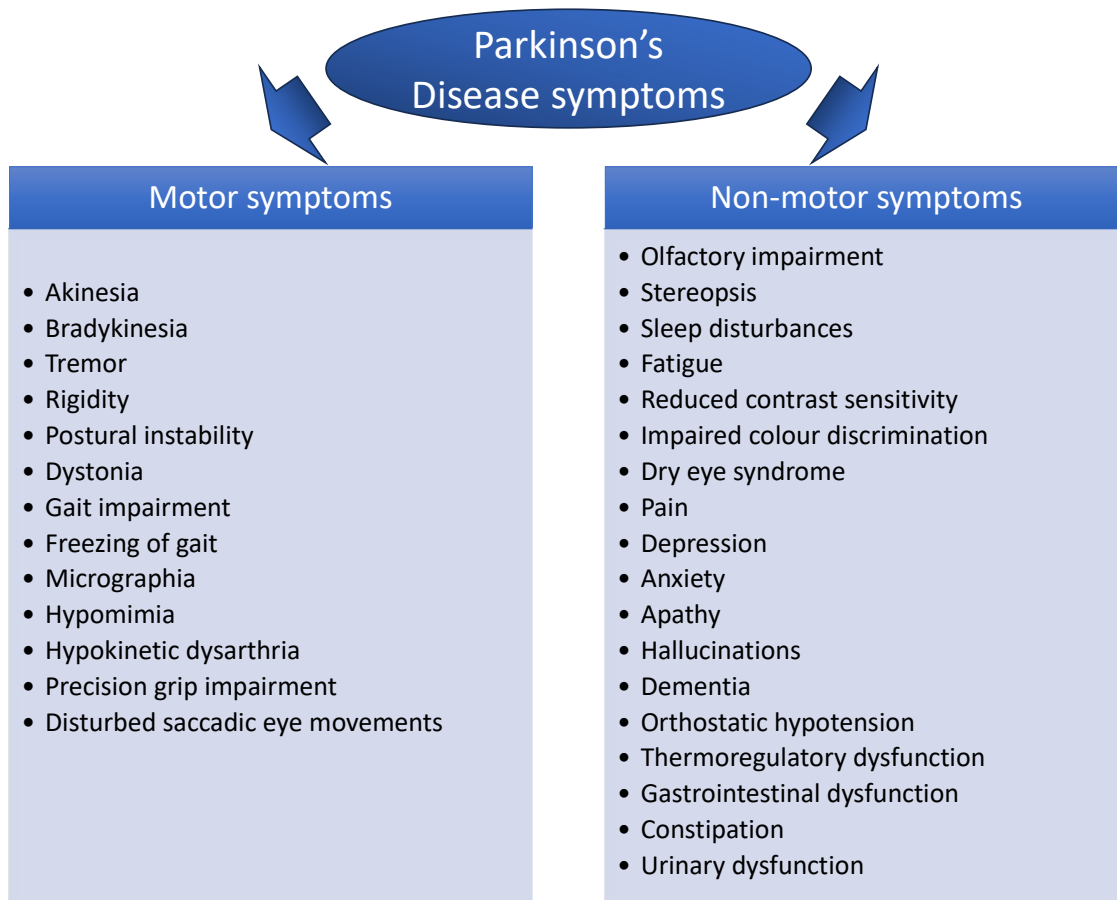


Fig. 2.2: The motor and non-motor symptoms of PD.

If the disease is detected in the early stage, the deterioration of the health is minor because the treatment is implemented. Because of this fact, early detection is especially desirable [91]. The aging of society forces the demand of creating new technologies for detecting neurodegenerative diseases in their early stage. Recently, the novelty in technology allows for utilising them for purpose of PD detection [63,92].

Nevertheless, the serious issue is that the early symptoms of the disease are lowly apparent. Moreover, the detection of cognitive decline is not a simple task because of the diversity of intelligence in the population, and the variation in education length. Furthermore, PD occurs more frequently in the male group [18,19]. The methods which are the most accurate, however, are thereby expensive including MRI, CT, and PET. Those methods are rather used in more advanced stages and the hospital environment. The limitation is also the price of the examination. Because of those

factors, there is a need for relatively cheaper and more approachable solutions for patients [18, 19].

This disease cannot be cured, however, the process of development could be inhibited. There are applied methodologies such as neurostimulation or pharmacotherapy [93, 94]. The patients visit the hospital several times per year to maintain relatively good health, nevertheless, they could meet the Hawthorne effect or this amount of appointments per year is not sufficient [95]. Patients with PD can experience sudden neurodegeneration, side effects such as levodopa-induced dyskinesia, or numerous fluctuations in motor function. To prevent the deterioration of health quality, such incidents require immediate intervention. Telemedicine solutions, which are capable of addressing those issues, are in the research interest of many scientists. The application of mobile phone usage for PD detection and monitoring of the progress of the disease seem to be especially interesting. Examples of the application of mobile smartphones for healthcare belong among the mHealth systems [96–100]. In order to minimise the possibility of manifested damages, the methods of PD recognition will be targeted particularly for early detection. Early detection refers to the recognition of a disease at an early stage of its progression [101].

PD management is challenging despite the achievements in treatment approaches [4]. One of the most desirable targets is PD detection, especially in the early stage. The used tools for this purpose are based on artificial intelligence. AI in the healthcare domains is a paradigm shift, to detect, predict, and manage PD. The seven categories were distinguished and are presented in Fig. 2.3 where the AI finds application to deal with PD. The most common analysing modalities with ML are speech, gait, sleep pattern with actigraphs, hypomimia, handwriting, and tremor. Not only the detection of the PD is the aim of the researchers, but also the prediction of wearing-off state [102]. Additionally, finding the clinical interpretable biomarkers is in the circle of researcher interest. Furthermore, the monitoring of brain lipidomics and the monitoring of dysregulated gut microbiome status among PD patients with the usage of AI is a captivating domain for scientists. Next, the AI is explored for smart gait and monitoring nanorobots in treatment and diagnostic likewise boosting the potential of telemedicine. Further, the submittal of assistance in an advanced stage of PD is a topic where the AI solutions will aid. For instance, the metaverse applications could support PD patients with cognitive decline. The technologies that could also support the solutions for PD patients in advanced stages are IoT, wearables, sensors, and mHealth. Additionally, AI could serve in neurosurgery to strengthen for example the process of decision-making during surgery. Moreover, AI could facilitate the drug discovery for this disease and development process [4].

One of the symptoms which could be utilised for PD detection is hypomimia. The hypomimia manifests in an expressionless face with no or little sense of animation,

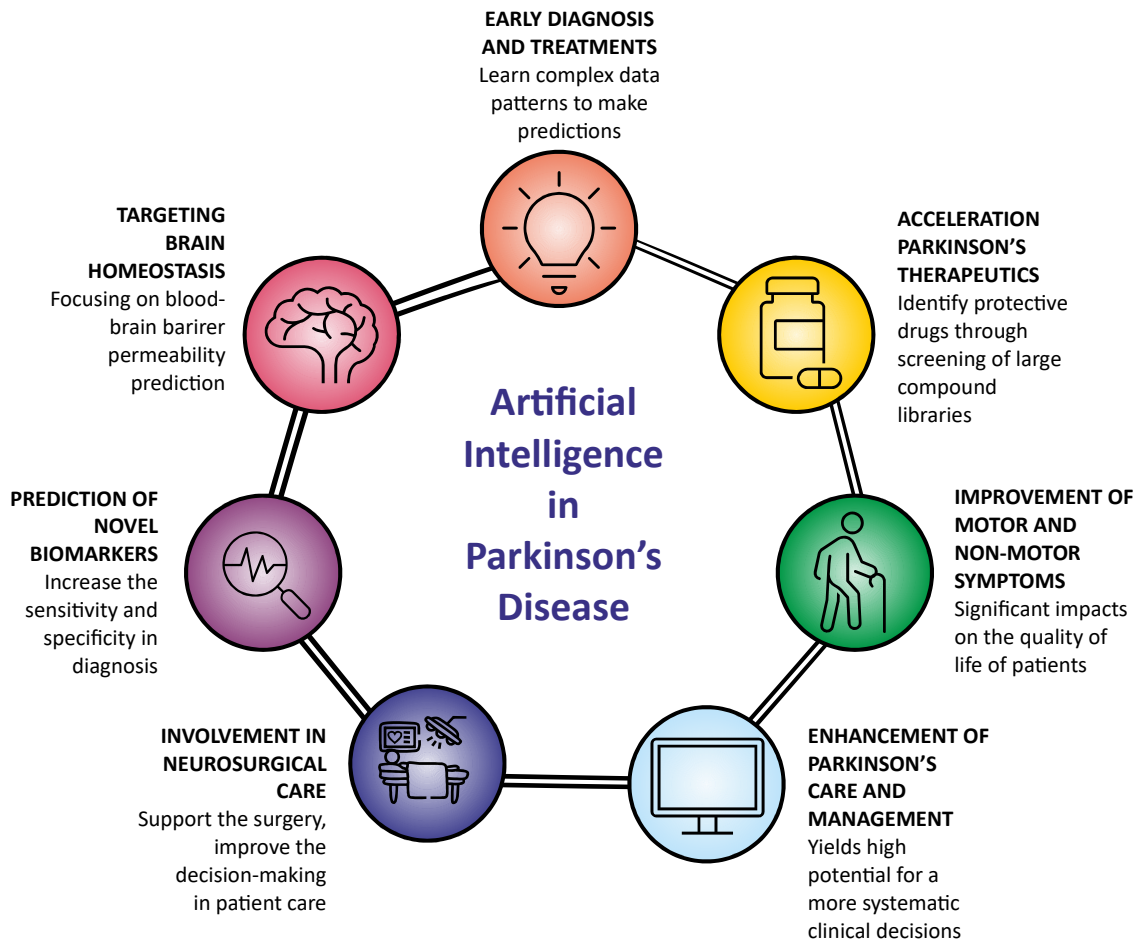


Fig. 2.3: The domains in which AI finds application in PD [4].

the reduction of facial expression [103], the slowness and limitation of facial motion (facial bradykinesia) [104]. Moreover, there is observed a stiffness of muscles, the issue with orofacial movements, i.e., the slower speed of the jaw lips [105–107], decreased blinking rate [108], unconsciously opened mouth [109], flattened nasolabial folds [109], occurring asymmetry in the face [110], decreased ability to raise eyebrows [88], etc. It is considered that PD patients recollect a so-called ‘poker face’. Furthermore, expressing emotions by them is a challenging task [88]. Additionally, Parkinsonians have an impaired ability to recognise human emotions in comparison to the healthy group [111].

The communication skills of PD patients are also affected because of occurring dysarthria among them [89] likewise the impairment of cognitive skills [112]. Communications serve as the way of expressing the emotions, feelings, information, and ideas by people [113]. Social well-being is disrupted in PD because of a decrease in



communication skills [112]. A good indicator of the level of communication skills and progress of the disease could be regarded as the tongue twister, i.e., validated speech exercise. Pronunciation in this speech exercise is challenging and could reveal the PD level because of difficulty in the appropriate usage of tongue and mouth. Furthermore, dysarthria affects articulators and their debility may manifest particularly during the performance of exercises such as the tongue twister [89].

Another early motor symptom of PD is HD [114,115]. This mark occurs parallelly with hypomimia [80]. HD manifests in 90 % of PD patients [116] and this speech disorder occurs because of a basal ganglia control circuit pathology [80]. HD occurs in the field of phonation, prosody, articulation, and respiration. The exact difficulties are apparent in irregular pitch fluctuations, breathy and harsh voice quality, monoloudness, reduced loudness, airflow insufficiency, unnatural speech rate, imprecise articulation, monopitch, improper pausing, etc. The detailed description of HD is attached in [84,117–119].

Moreover, sleep problems are considered signs of PD. Among them could be distinguished following symptoms: insomnia, *Excessive Daytime Sleepiness* (EDS), *Rapid Eye Movement Behavioral Disorder* (RBD), *Restless Leg Syndrome* (RLS), and breathing difficulties. Those symptoms manifest in the early stage of illness [87,120–122].

The quality and quantity of sleep influence people's health. The measurements describing sleep could be indicators of illnesses. Wearable devices (including smartwatches) can be used to evaluate sleep disorders as well as sleep diaries, WiFi-based, bed sensors, PSG, *videosomnography* (VSG), *radiofrequency* (RF), and EEG [123]. However, the PSG is considered to be the gold standard. The records of respiration, *electrocardiography* (ECG), EEG, *electromyography* (EMG), *electrooculography* (EOG), oximetry, and body position are collected during conducting PSG [124,125] (see Fig. 2.4). Unfortunately, this procedure is carried out in the hospital environment. Moreover, there are also tests like VSG, accelerometry, and Continuous Positive Air Pressure [123]. The drawback of the PSG is that it is often executed in the later stage, not in the highly demanded early stage, and it is performed in hospitals.

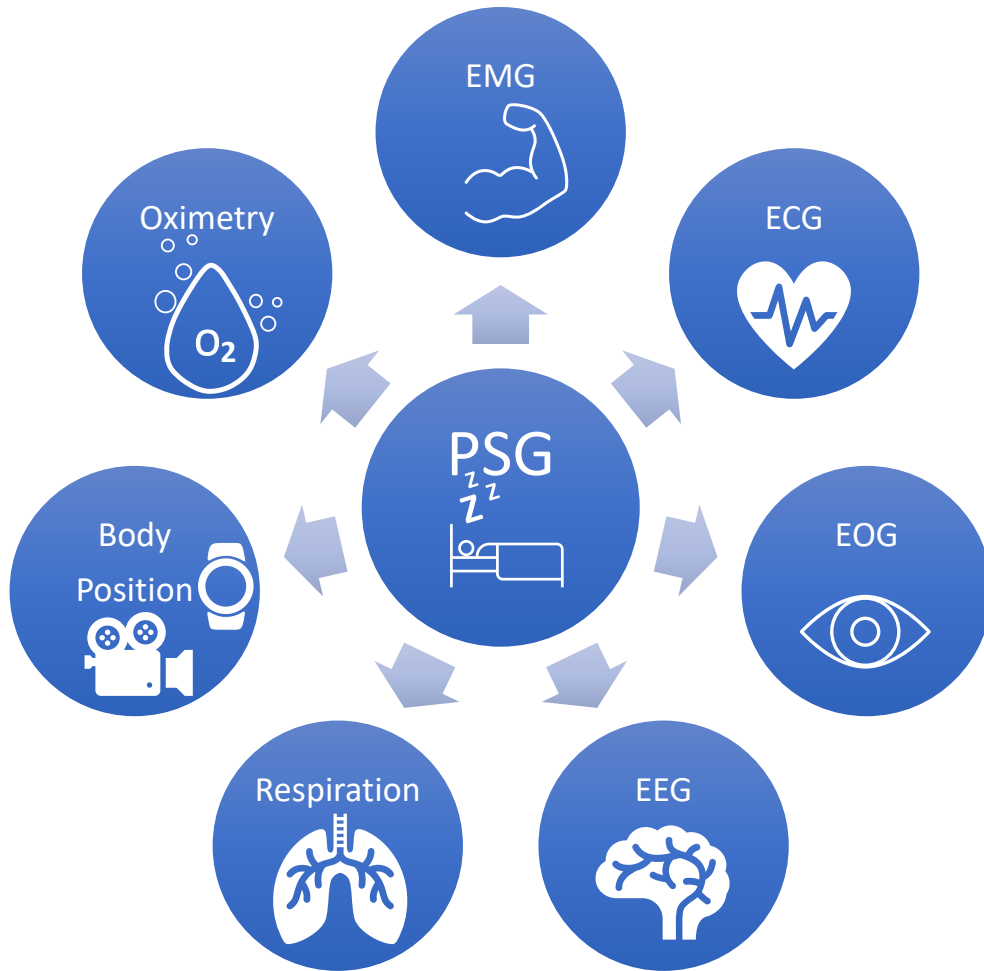


Fig. 2.4: The parameters measured during PSG.

## 2.3 EEG, Time Series, and Methods for its Detection

The non-intrusive diagnostic technique for analysing the bioelectrical brain function is so-called EEG. Firstly, the electrodes are situated on the head skin surface, then the signals of variance in potential likewise the changes between the different parts of the brain in potential are gathered. Those signals are finally amplified to generate a record of them. This record is the so-called electroencephalogram.

The aim is to catch the flow of the waves which have characteristic forms and typical frequency bands. 1-100 HZ is the range of frequencies, whereas amplitude is equal to 5 to several hundred  $\mu V$ .

The experiments with EEG in humans date back to 1924. It was used for the recognition of the pathologies. The analysis of emotions, *brain-computer interface* (BCI), and mental workload represent the field where EEG found applications recently. Mental disorders could be analysed by EEG such as Alzheimer's disease,

schizophrenia, *attention-deficit hyperactivity* (ADHD), depression, and dementia. Moreover, the symptoms of a brain tumour or epilepsy could be found with the usage of EEG. Furthermore, this method is useful for analysing the sleeping pattern, human ability to concentrate, and profoundness of anesthesia [126,127]. The aforementioned applications of the EEG are depicted in Fig. 2.5.

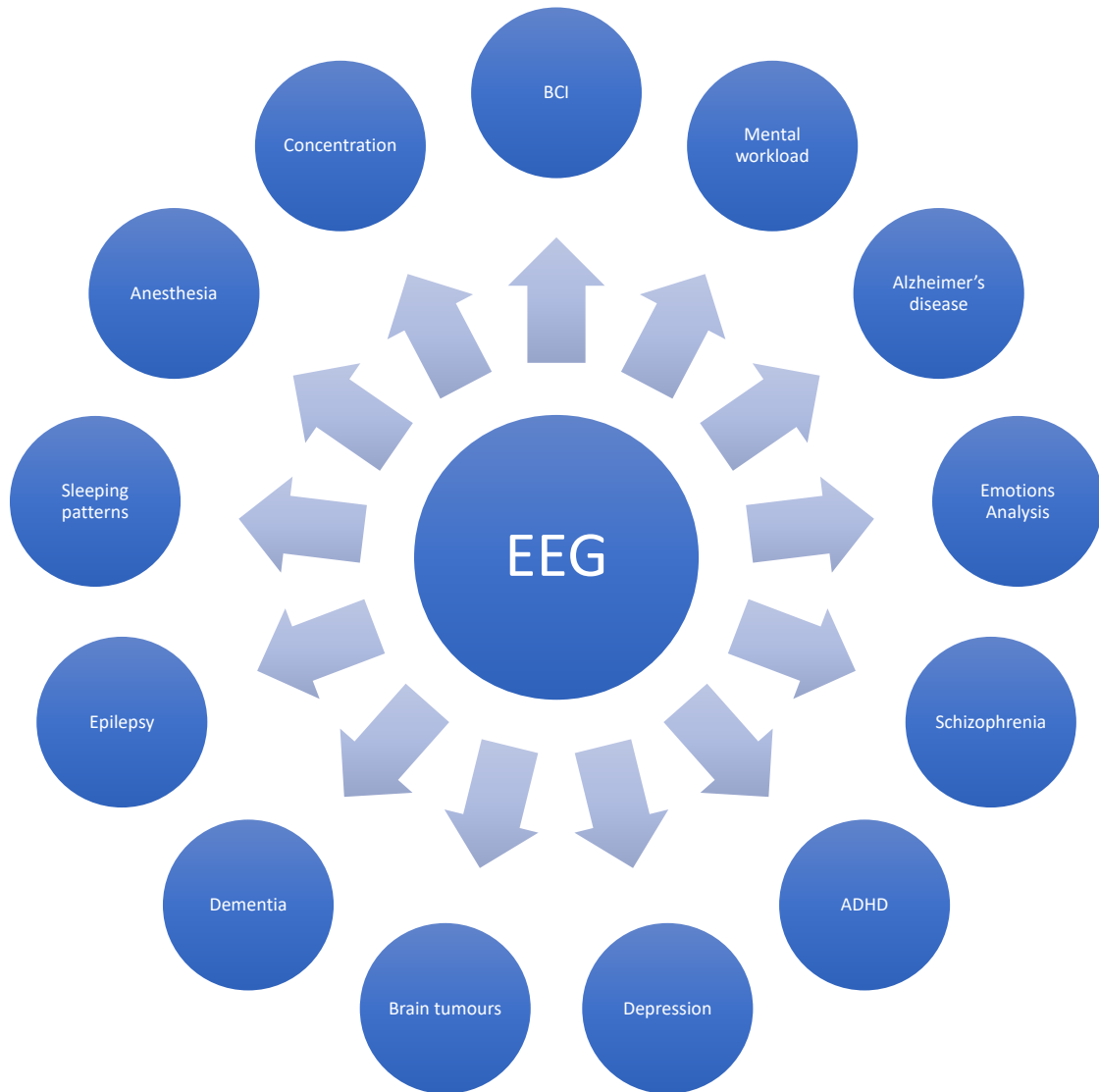


Fig. 2.5: The applications of EEG.

The main problem with EEG signals is the inferior *signal-to-noise ratio* (SNR). The filters are used to reduce the influence of noise on the signal, and threshold the outliers of the signal. The impact of the cardiac bioelectrical activity, human movements, the motion and variance in the tension of muscles, and eye movement are eliminated thanks to the denoising technique. Moreover, the analysis of EEG is

challenging because of its non-stationary character. Because of this, the classification of the EEG signal is challenging. Furthermore, the EEG signal varies interpersonally [127].

The methods suitable for EEG analysis are the modern deep learning techniques likewise transition signal into the frequency domain and extraction of the features. The deep learning techniques serve as the pre-processing stage, feature generation, regression, and classification. The types of neural networks, which were used for the analysis of the EEG, are CNN [128], LSTM [129], *recurrent neural network* (RNN) [130], restricted Boltzmann machines [131], *generative adversarial network* (GAN) [132], *autoencoder* (AE), *fully-connected layers* (FC) [133], *deep belief network* (DBN) [131], transfer learning [129,134]. The augmentation of the data is feasible thanks to the usage of transfer learning and GAN. This approach has a chance to increase the quality of classification tasks [135].

## 2.4 Conclusion

This chapter provided an introduction to the topics elaborated in this thesis. The nature of the diseases analysed in this work was described, i. e. COVID-19 and PD. Additionally, the symptoms of those illnesses were briefly illustrated. They are crucial foundations to understand assumptions of the developed methodologies presented in Chapters 4, and 5. The prelude of the diagnostic methods of COVID-19 and PD was provided in this chapter. The state-of-the-art methods of diseases recognition and prediction of the pandemic are broadly described in the next Chapter 3. That part of the thesis is focused on the presentation of the used methods of artificial intelligence, wearables, eHealth, and mHealth. Furthermore, the foundations of EEG analysis were presented (it is an answer to **RQ7.1.**). Additionally, the instances of artificial intelligence methods - commonly used neural networks suitable for EEG processing were portrayed which could be applied to wearable-related data (they are the answers to **RO7.**, **RQ7.2.**, **RQ7.3.**). They are: CNN [128], LSTM [129], RNN [130], restricted Boltzmann machines [131], GAN [132], AE, FC [133], DBN [131], transfer learning [129,134], likewise transfer learning and GAN. The applications of ML for EEG and neurodegenerative diseases are described also in the following Chapter 3.

## 3 The state-of-the-art overview

This chapter presents the state-of-the-art methodology for assisting in the diagnosis of COVID-19, PD and analysing EEG signals. The review of technology dedicated to the recognition of COVID-19 is focused on the personal level likewise the analysis of pandemic trends thanks to the usage of wearables and ML. Furthermore, the illustrated methods of PD diagnosis and evaluation of the progress of the illness are based on hypomimia, HD, and sleep disorders. Moreover, the approaches to the detection of different diseases, which would be potentially suitable for PD detection based on sleep disorders and actigraph records, are discussed. Moreover, the applications of deep learning for EEG analysis are depicted. Additionally, the examples of Granger Causality for EEG are illustrated. Finally, the operational rules of neural ODE are explained and the advantages of this neural network are presented.

### 3.1 COVID-19 Detection with the Usage of Wearables and Machine Learning

So far, there were introduced few solutions for the detection of COVID-19 with the usage of wearable devices together with ML approaches in the literature. A short review of them is presented in this section. If COVID-19 is detected early, the reproduction rate can be significantly reduced and the infection can be prevented from spreading. Nevertheless, the symptoms do not manifest approximately two days before the visible onset of the disease which supports the spreading of the virus.

The authors in [59] made an analysis of changes in heart rhythm and daily activity of COVID-19 cases based on records of HR and the number of steps taken during the day. The sampling rates were one per minute and one per day, respectively. Additionally, the sleep patterns were monitored, however, the data were incomplete. The records of the devices come from the Fitbit smartwatch. Stanford University performed an experiment. The target of the experiment was to detect anomalies in the prodromal stage of the disease. They obtained 32 COVID-19 cases, 15 Influenza, and 73 HC among 5300 participants. Three algorithms were developed: *Resting Heart Rate* (RHR) Difference anomaly detection, the *heart rate over steps anomaly detection* (HROS-AD), and *cumulative sum* (CuSum) [59]. Thanks to the CuSum algorithm, 63 % of COVID-19 cases were recognised positively. Nevertheless, the authors did not consider specificity [59]. *RHR Difference* (RHR-Diff) offline anomaly detection tried to find anomaly detection in HR thanks to the residuals standardization of RHR. 1 hour signal of RHR was standardised on the RHR average of 28

days. The time window - interval is considered to be an anomalous if the window is under the relevance of 0.05. The HROS-AD is an unsupervised learning anomaly detection algorithm [59]. The metric the *ratio of heart rate to the number of steps taken* (HROS) and Gaussian distribution analysis were used. The moving average, undersampling to one hour, and Z-score transformation were utilised for HROS-AD algorithm. The anomalies found by the Gaussian distribution analysis were recognised as outliers. The algorithm which was working in real time was CuSum. The deviations of residuals of RHR were summed and 28 days of records were taken into account during performing CuSum.

In another work [1], the Fitbit - wearable device was also utilised for COVID-19 analysis. COVID-19 cases and two types of Influenza were taken into consideration. The authors enrolled 7000 participants and gathered data for 41 COVID-19 cases, 85 Influenza during the pandemic, and 1265 Influenza before the main pandemic. The number of steps taken by human was collected together with RHR records. A longer median duration of COVID-19 cases (12 days) was observed than the spanning of Influenza before the main pandemic (7 days, Pre-COVID-19 Flue) and during the pandemic (9 days, Non-COVID-19 Flue). The self-reported illness duration is illustrated in Fig. 3.1. Thanks to the statistical analysis, it was proved that raised RHR manifests often nearly the onset of the disease. The authors also compared the RHR between COVID-19 and Influenza cases, and COVID-19 records characterise higher values of RHR. Additionally, symptoms such as shortness of breath, anosmia, and chest pain were typical for COVID-19 cases.

The data from [59] were also analysed in [136]. The PCovNet was proposed which is a *Long Short-term Memory Variational Autoencoder* (LSTM-VAE), to detect the anomalies in the early stage of the disease. This network has been trained on 25 COVID-19 cases analyses the RHR. The 0.946 precision and 0.234 recall was achieved in the experiment conducted by authors [136]. Furthermore, F-beta was computed with the result of 0.918. The usage of this parameter is nevertheless untrustworthy because the true negatives are omitted. Moreover, the authors outweighed the importance of precision over recall thanks to the  $\beta$  parameter. According to RHR and PCovNet, 100 % of the individuals with the disease were considered ill, however, these individuals were already infected, and as such, they should be placed in quarantine. The analysis of anomaly detection based on this same dataset [59] was provided in [137]. The *One Class-Support Vector Machine* (OC-SVM) was used to detect COVID-19 cases. The authors outperformed the results from [59]. The anomalies in RHR signal were detected 23,5 % - 40 % earlier in comparison to [59]. Moreover, the authors of [137] provided false positive rates. They established the optimal time window length of RHR as 300 and 350. 21 among 29 COVID-19 cases was the maximum anomalies number found by the

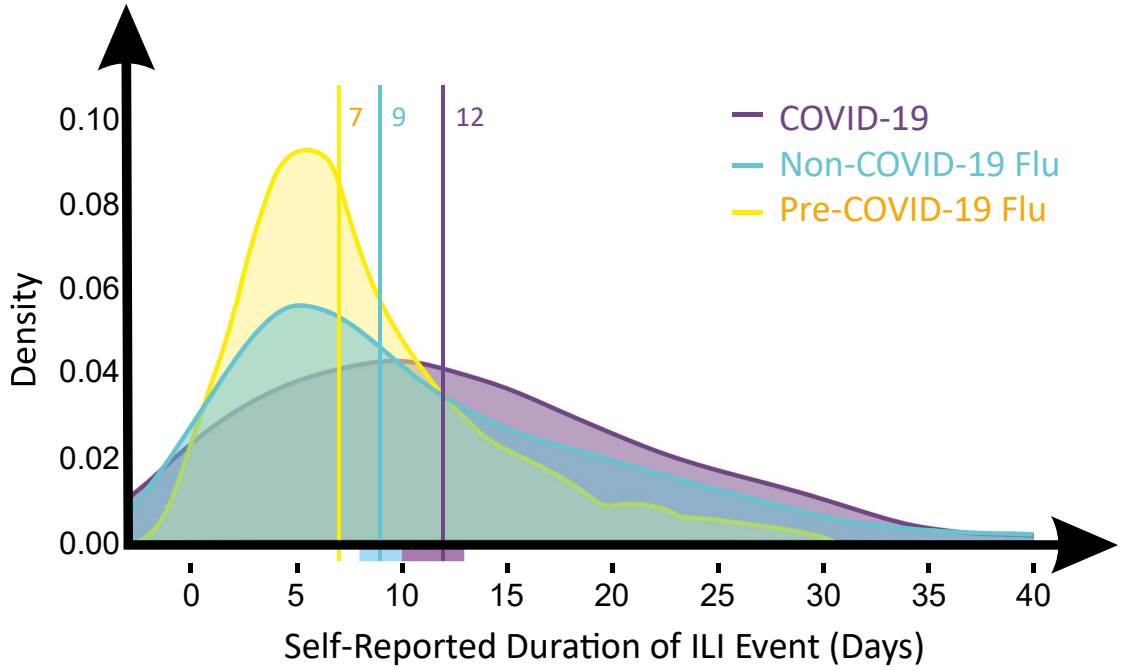


Fig. 3.1: The illness duration for COVID-19, Influenza before the main pandemic and during the pandemic [1].

OC-SVM based on RHR (RHR-OC-SVM). For HROS, the OC-SVM (HROS-OC-SVM) detected anomalies for 24 among 29 COVID-19 patients. The number of outliers for HC was 39.96 (maximum was 100) for the false positive rate thanks to the RHR-OC-SVM [62].

Another modality that was used for COVID-19 evaluation thanks to the wearable was temperature. In [138], the Oura ring was utilised in the experiment. This device is capable of gathering the values of temperature, respiratory rate, HR, and *Heart Rate Variability* (HRV). The temperature changes were observed before other symptoms in this research. This phenomenon was detected in 38 participants among 50. The nonparametric Kruskal Wallace test with Tukey-Kramer post hoc comparison was performed. Additionally, a strong correlation between cardiac rhythm and fever was observed in [5]. Elevation of HR was detected as 8.5 beats per minute on average per 1 °C. This relation is evident for RHR. This fact could be explained by demonstrating higher precision in gauging resting time by wearables. Unfortunately, this dependence is not special only for COVID-19, but also for influenza. Moreover, a rise in RHR is also characteristic of short sleep. Furthermore, the authors in [139] used also temperature as the parameter of characterisation of COVID-19. They proved that temperature has an influence on COVID-19 detection among physiological signals such as temperature, HR, HRV, RR, and *metabolic equivalents* (MET)

collected by the Oura Ring. For the classification task by analysing data coming from 73 people with COVID-19 disease, the researcher in [139] obtained for all of the modalities the *Area Under the Curve* (AUC) = 0.819 and after removing the temperature modality AUC = 0.770. The scientists tested Random Forest to detect COVID-19 patients.

The utilisation of medical device was presented in [140]. It was the smartwatch Empatica E4. The research considered the usage of a four-layer neural network – CovidDeep together with augmentation techniques. Several modalities were taken into account, including the *inter-beat interval* (IBI), pulse oximeter, skin temperature, blood pressure, and *galvanic skin response* (GSR). Moreover, the neural network analysed extra collected parameters such as weight, height, gender, age, addiction to drinking and smoking, and habits. The dataset contains records of 30 symptomatic COVID-19 cases, 27 asymptomatic COVID-19 patients, and 30 HC. The signals were portioned into 15 s windows. The CovidDeep contains data pre-processing, synthetic data generation with the TUTOR framework, architecture pre-training, grow-and-prune synthesis with a decision tree and random forest, and output generation through softmax. The best result of COVID-19 detection was 98.1 % of accuracy. The modalities which were taken into account were blood pressure, GSR, oxygen saturation, and a questionnaire. The use of the Empatica E4 is still challenging as a screening test due to its limited distribution in population and high cost.

The influence of oxygen saturation ( $SpO_2$ ) and RR were taken under analysis in [141]. 208 records of COVID-19 cases from smartwatches were selected for the study. The statistical analysis was conducted using chi-square distributions and independent t-tests. According to the chi-square distribution, there is no significant difference between IoT factors and gender. Range and coverage, compatibility, interoperability, performance, and secure connectivity with wearables belong to the IoT factors.

The respiration rate was used for COVID-19 detection [142]. The data were gathered by the WHOOP smartwatch. The dataset contains the record of 81 COVID-19 cases and 190 HC. The median RR per minute was gathered during the night and recognised as the respiration rate. This signal was analysed by the WHOOP strap algorithm. The study allows recognition of 20 % of COVID-19 cases two days before the visible onset of the disease and, 80% of COVID-19 cases were detected three days after the onset. The gradient boosting was used as a classifier. The detection of 20 % of COVID-19 cases is much more crucial because of the need for the detection of the disease in its early stage and the usage of it as a screening test. Nevertheless, the percentage rate is low for this purpose.

The changes in HR, HRV, and RR were examined in [72]. The records were



collected by Fitbit smartwatch. The number of COVID-19 cases was 1181 and HC was 13662. The applied CNN obtained  $AUC = 0.77 \pm 0.03$ , whereas the sensitivity was equal to 47 % and specificity was equal to 95 %. The researchers computed the following parameters: the estimated mean RR, the mean nocturnal HR during NREM sleep, and the Shannon entropy of the nocturnal RR series. As a method of normalization, the Z-score was applied. Moreover, during the training process, parameters such as gender, age, and *Body Mass Index* (BMI) were input into the neural network. Moreover, the elevation during disease in HR together with RR was observed, whereas deterioration in HRV was detected.

The above-mentioned works address the issue of detecting COVID-19 on a personal level through the use of wearable devices. Moreover, the literature reports on the studies of analysis of pandemic trends – the crowd-level analysis [2].

The analysis on the crowd level was presented in [5]. A system for alerting of anomalies in physiological signals was developed using deviations in sleep patterns and RHR calculation from *photoplethysmography* (PPG) wearable records. Thanks to the Huami devices, the records of 1.3 million participants were collected. The support methodology was created with the usage of the heterogeneous neural network CDNet. CDNet contains CatNN and DenNN. This network consists of dense numerical features (historically officially reported COVID-19 rates, historical physiological anomaly rate, active user density) and sparse categorical features (season, weather, and holiday activity). The schematic structure of the CDNet network is presented in Fig. 3.2. Using Pearson’s correlation, the COVID-19 infection rate was compared with the physiological anomaly rate. The regions of China such as South China, Central China, and North China were taken into account together with South-Central Europe. Foshan, a Chinese city with a correlation of 0.81, was observed to have the highest correlation. The average value across all cities was 0.68. However, local events might influence people’s common behavior patterns, just as individual variability may affect the model.

Furthermore, AI was used to analyse and manage data in [143]. The dataset consists of blood parameters collected from the laboratory in 2020. ML was applied to detect COVID-19. There were 80 patients with COVID-19 among 600 patients. A total of 18 features are included in the dataset. Classifiers such as Random Forest, Naïve Bayes, and SVM were applied. The normalization and feature selection was performed as the pre-processing step. The highest results were obtained for SVM: 95 % accuracy, 94 % F1, 95 % precision, 95 % recall, and 95 % AUC. Unfortunately, there is no provided information in which stage of COVID-19, the data were gathered.

Additionally, not only the stage of COVID-19 performed with the usage of ML but also the need for hospitalization in [144]. The authors developed the COVID-19

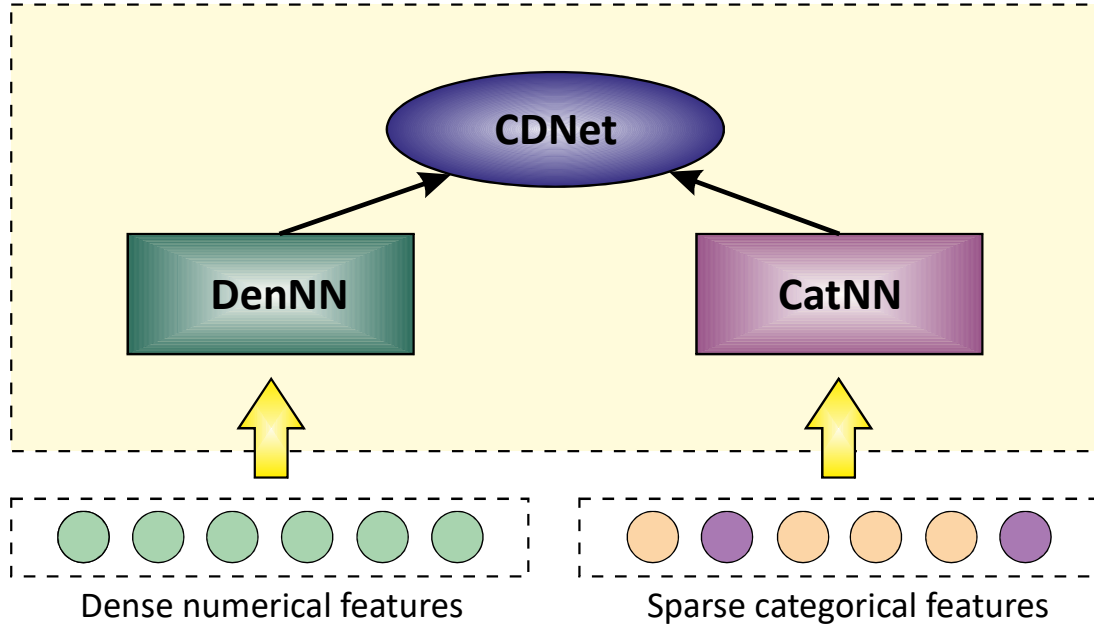


Fig. 3.2: The scheme of the neural network CDNet [5].

*decompensation index* (CDI). Continuous monitoring was done with the wearables situated on the chest - VitalConnect VitalPatch and finger pulse oximeter - Proactive, Protekt® Finger Pulse Oximeter 20110. There were gathered records of raw 125 Hz ECG, triaxial accelerometer, skin temperature, and oxygen saturation. There are 22 positive cases (requiring hospitalization) and 308 negative cases included in the dataset. AUC of 0.84 was obtained by gradient boosting.

The summary of the usage of ML methods with wearable for analysing the COVID-19-related data is presented in Table 3.1.

Furthermore, the platform for monitoring COVID-19 was introduced in [145]. It is known as the Biovitals Sentinel and the platform utilises armband biosensors (Everion). It is possible to utilize this platform as a source of data for supporting system methodology. A number of signals are collected, including RR, pulse rate, daily activity statistics, skin temperature, blood oxygen saturation, and blood pressure.

Furthermore, the utilization of wearable for measuring physiological signals is highly needed. There are a variety of options. In [146], the authors used the skin sensor placed on the throat to measure the accelerometer signal and temperature. Additionally, this wearable device could measure the cough frequency, duration, and intensity of cough with a wireless solution. The heart measurements were also performed. Unfortunately, this skin sensor and approach are not suitable for early-

stage detection, but for a more developed stage when the cough is occurring. A continuous measurement of the progression of the disease could be made using this device. Moreover, the situation of the wearables could be various. On-body sensors can be integrated into smart rings, headbands, camera clips, sociometric badges, smartwatches, and embedded in clothing [20, 62]. The use of smartwatches and smart masks was primarily used for the early detection of COVID-19 [147]. An interesting method for measuring  $SpO_2$  was presented in [148]. The researchers proposed a wearable equipped with a PPG sensor to measure oxygen saturation. This device exhibits the same level of accuracy as finger pulse oximetry as well as the ability to respond more quickly. This wearable could measure potential hypoxemia state. Further, the usage of a headset wearable was proposed for coughing detection in [149]. Nevertheless, this application is not suitable for early detection of COVID-19, because the cough is later a symptom of the disease.

Additionally, the ability of a smartwatch, such as Apple Watch, to collect ECGs can be used to monitor the COVID-19 disease [150].

There were presented a few solutions for the detection of COVID-19 thanks to the usage of wearables and ML. It depends on how data are collected and which modality is chosen to be able to be used in the early stage of the detection. There were introduced also solutions that are atypical like wearable placed in the ear or smart masks in comparison to common smartwatches for monitoring COVID-19. The application of Empatica E4 is unfortunately limited, because of the cost of the device and not spreaded widely in society, to be used as a screening test. The usage of a Fitbit device, Apple Watch, Garmica, or Oura Ring seems to be more suitable.

A review of the methodology of COVID-19 recognition and pandemic evolution with ML and wearable devices is presented in Table 3.1.

Tab. 3.1: An overview of the methods of COVID-19 detection and pandemic development with the wearables and AI. [2].

Citation	Main aim	Device	Kind of data gathered	Size of the dataset	Accuracy, efficacy	Machine learning method	Comments
[5]	Predicting the epidemic trend including anomaly detection with COVID-19 infection rate	Huami (ACC, PPG)	HR, sleep data	1.3 mln participants	The highest Pearson correlation for Chinese cities: Foshan 0.81, average 0.68	CDNet (CatNN, DenNN)	The simulation provided for North, Central, South China, and South-Central Europe.
[138]	Statistical analysis of daily temperature for COVID-19 disease and creating digital biomarkers	Oura ring	temperature	50 COVID-19 cases	38/50 patients exhibited some temperature anomalies before the onset of the disease	Threshold based on min/max temperature record after z-score, Statistical evaluation: nonparametric Kruskal Wallare test, with Tukey-Kramer post hoc comparison	More wearables should include temperature sensors.
[99]	Anomaly detection of COVID-19 disease	limited to Fitbit	HR, sleep disorders, number of steps	73 HC, 32 COVID-19 cases, 15 Influenza	63 % anomaly detection in COVID-19 cases	Developed algorithms: RHR-Diff, HROS-AD, CUSim	Anomaly detection evaluated on COVID-19 disease cases without considering classification problem.
[141]	Correlation of wearables related data with gender and IoT factors	Lack of detailed informations	RR, oxygen saturation	208 COVID-19 cases	no significant differences between IoT factors and gender	Chi-Square distribution and independent measures t-Test	There should be a difference of future created support system methodologies between the population according to the analysed factors.
[140]	Evaluation of COVID-19 disease based on Empatica E4	Empatica E4	GSR, BI, skin temperature, pulse oximeter, blood pressure questionnaire	30 HC, 57 COVID-19 cases (27 asymptomatic, 30 symptomatic)	98,1 % accuracy	CovidDeep	The data contains self-assessment done by patients, the pre-processing step is not clear. The results are obtained with the medical device – Empatica.
[142]	Detection of COVID-19 disease	WHOOP Strap	Respiration rate	81 COVID-19 cases, 190 HC	20 % COVID-19 subjects recognised before the onset, 80 % cases 3 days after onset	Gradient Boosting	80 % is well results of accuracy, however, the target is to detect disease before the clear onset.
[72]	Prediction of the COVID-19 disease based on RR, HR, HRV and also age, gender, BMI	Fitbit	RR, HR, HRV	2754 COVID-19 cases	0.77 +/- 0.03 AUC, sensitivity 47 %, specificity 65 %	Computed parameters: Shannon entropy of the noctural RR series, the mean noctural HR during deep sleep, pre-processing: transformation into z-score, algorithm: CNN	Some extra parameters were provided during training – among others: age, gender, BMI. HR together with RR is increasing during illness, HRV is decreasing.
[1]	Comparison of COVID-19 disease in the early outbreak, later outbreak and also with Influenza	Fitbit	self-report data, RHR, step counts, nightly sleep hours	41 COVID-19 cases, 4985 self-reported flu, 1265 pre-pandemic COVID-19	statistical differences in tests	Statistical evaluations	The authors demonstrate the higher intensity and variety in symptoms for COVID-19 cases than for normal flu.
[136]	Anomaly detection of COVID-19 disease in the early stage	Fitbit	HR, number of step taken	25 COVID-19 cases, 10 Influenza, 67 HC	0.946 precision, 0.234 recall, F-beta 0.918	PCorNet	F-beta is an unreliable metric.
[137]	Anomaly detection of COVID-19 disease	Fitbit	HR, number of step taken	32 COVID-19 cases, 74 HC	Anomalies were detected 23.5%-40 % earlier in comparison to [59] 21/29 found anomalies for RHR-OCSM 24/29 found anomalies for HROS-OCSVM	OC-SVM	There is no consideration of classification problem.
[139]	Checking the influence of temperature on the classification task	Oura Ring	temperature, HR, HRV, RR, MET	73 COVID-19 cases, approximately 63000 HC	AUC = 0.819 for all modalities, AUC = 0.770 for all modalities without temperature	Random Forest	The temperature was confirmed to be valuable for COVID-19 detection.
[144]	Determination of the need of hospitalization	VitalConnect VitalPatch, Proactive, Prox@ Finger Pulse Oximeter 20110	raw 125 Hz ECG, 50 Hz triaxial accelerometer, 0.25 Hz skin temperature, SpO <sub>2</sub>	22 positive cases (required hospitalization), 308 negative cases	AUC = 0.84	Gradient Boosting	The determination of the need for hospitalization was decided based on CDI.

## 3.2 Parkinson’s Disease Detection based on Symptoms

MRI, PET, and CT are considered as the most accurate methods for detecting PD. Nevertheless, those methods are expensive and for this reason, are not suitable to serve as a screening test. Moreover, the PSG is an accurate test for recognising sleep disorders in PD, however, it requires a hospital environment to be carried out and it is not appropriate for early diagnosis. Because of this reason, there is a need for the utility of more approachable techniques [77]. Methods based on video analysis [151], wearable sensor analysis [152,153], and audio analysis [154] appear to be more cost-effective and accessible for the detection of PD. The combination of the modalities found also a research gap by the scientists in [155–157]. The video modality is applied to human gait analysis [151] and hypomimia recognition in PD [7].

### 3.2.1 Hypomimia

There has been a limited amount of research published on PD detection based on hypomimia. The main issue is the missing access to the dataset containing hypomimia records. Some of the works were focused only on statistical analysis [110,158], whereas the studies in [7,159,160] considered emotion recognition in PD. The emotions which are taken into account are: sadness, surprise, happiness, anger, neutrality, disgust, and fear [161,162]. The meticulous review of hypomimia analysis in PD is presented in [161].

There were identified two groups of facial features extraction techniques for the automatic evaluation of hypomimia, namely: statistic-based and geometry-based. Statistics-based methods rely on measurements based on differences in illumination between pixels [163]. Whereas, the geometry-based methodologies utilise the facial landmarks and calculate the distances between those landmarks or compute the areas between some of the detected facial landmarks [164].

There are a few techniques that were applied for hypomimia analysis, i.e., the affectograms, *facial action coding system* (FACS), the automatic *maximally discriminative facial movement coding systems* (MAX), *facial electromyography* (fEMG), the *Action Units* (AUs), automatic *facial expression recognition* (FER), and techniques with the usage of AI for emotion recognition [63,161]. There could be distinguished two main groups of techniques utilizing video, or image, and ML. The methods belong to the first group are detecting pixels or facial landmarks on a face. The second group represents the solutions that utilize neural networks to extract features from images or videos [161].

The facial landmarks were detected and 12 features were obtained based on them in [165]. Areas and distances were extracted features. The performed exercise was a one-minute monologue of native speakers in the Czech language. In this exercise, 79 HC and 91 de-novo (in the early stage) and drug-naive (untreated) patients participated, excluding those suffering from depression. The classification task was performed thanks to the 5 features, the leave-one-subject-out cross-validation with binary Logistic Regression as a classifier. The metrics which were calculated were as follows: AUC = 0.87, accuracy = 78.3 %, sensitivity = 79.1 %, and specificity = 77.8 %. However, the computed features did not include the dynamic of facial muscle movements and motion of the anthropomorphic distances.

The work in [162] is dealing with the separation of PD from HC thanks to the analysis of emotions with the usage of the function. The experiment involved 8 HC and 7 PD patients. A function based on the duration, frequency, and intensity of FACS is used to compute facial expressivity to distinguish between two groups. Emotions were taken into consideration, particularly: fear, surprise, anger, amusement, disgust, and sadness. The task was to self-assessment of emotional state after watching the movie, nevertheless, this evaluation could be received as a drawback of the experiment because of the subjectiveness.

Another work was dealing with scrutinising the discrepancy in the ability to express the emotions between the PD and HC thanks to the analysis of video in [88]. The distinction was carried out thanks to the computed features vector of the Euclidean face between the neutral baseline and while expressing another emotion (happiness, anger, disgust, and sadness). The capability of expressing emotion for 17 PD patients and 17 HC was calculated this way. The statistical difference between the groups was proved with the conducted two-tailed t-test. Moreover, the most problematic emotions to express occur disgust and anger.

The combination of geometric and texture features was utilised in [6]. The differentiation between neutral and expressed emotion was taken into consideration. This was measured for geometric features by *facial expression factors* (FEFs) for activated states and *facial expression change factors* (FECFs) for detecting the moving trajectories of activated states. Whereas, the extended *histogram of oriented gradients* (HOG) was calculated for the texture features, including three dimensions, i.e., HOG-XY, HOG-YT, and HOG-XT. The 47 PD patients and 39 HC took part in the experiment. The *Principal Component Analysis* (PCA), 5-fold cross-validation was used as the ML methodology. The process flow of the designed approach is shown in Fig. 3.3. The best results achieved the combination of geometric and texture features by the Random Forest and SVM. The F1-score was equal to 0.9991 and 0.9997 for the Random Forest and SVM, respectively. However, in the opinion of the authors, the conducted PCA was vague and could suffer from overfitting.

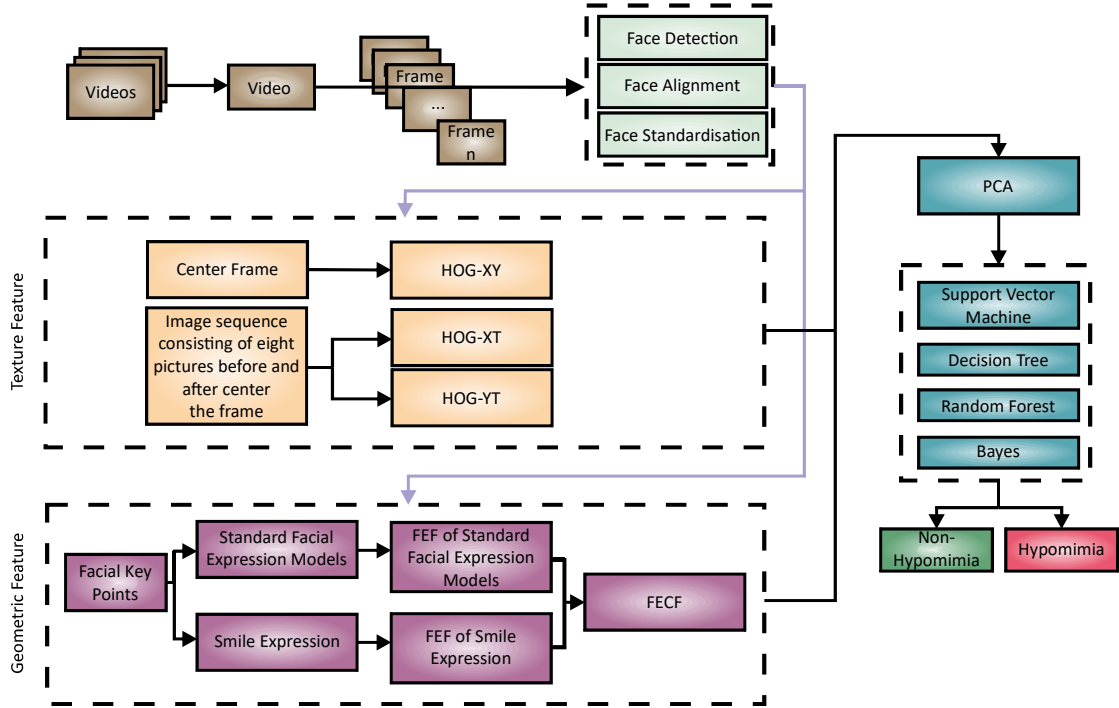


Fig. 3.3: The process flow of the approach in [6].

Another study [110] investigated the changes in entropy during the smiling. The video records were collected for 12 HC and 12 PD patients. 7 emotions were taken into account: surprise, anger, happiness, fear, disgust, and sadness. The hypomimia symptom was examined thanks to the shift in pixel intensity. The main point from the experiment is that reduced facial movement and bradykinesia were observed more often for PD patients than for HC for all feelings.

There were published also a few works with the usage of AUs. The set of AU and facial features for PD recognition was presented in [162]. The authors proposed the function of the frequency of AU. The statistical importance of the created function for separating PD from HC cases were proved. Moreover, the measurements of facial EMG were carried out to check the statistical significance between the two groups.

Furthermore, the AUs were computed in [160], and data were gathered from *three dimensional* (3D) sensors. The collected dataset contains records of 15 HC and 15 PD patients. The AU and linear regression were used to create a ML model. The obtained accuracy was in the range of 0.90 to 0.99. Nevertheless, the dataset is relatively small.

The evaluation of the impairment of facial movements was done by grading AU from 0 to 5 in [166]. The 1812 video records were collected by a webpage

tool <sup>1</sup> for 61 PD patients and 543 HC. The single video includes the recording of repeated 3 times emotions with a pause for the neutral expression. Three feelings were examined: surprise, smiling, and disgust. The tests lasted between 10 and 12 seconds. Examination of the disease was conducted using the FACS and ML. The SVM was used as a classifier with the leave-one-out cross-validation and was preceded by the *Synthetic Minority Oversampling Technique* (SMOTE) on the whole dataset. The results of the classification were equal to 95.6 % accuracy. Moreover, the Logistic Regression was applied to define the interpretability of made decision and identify the valuable AU. The most significant features were AU\_01 (inner brow raiser), AU\_06 (cheek raiser), AU\_12 (lip corner puller) for smiling, and AU\_04 (brow lowerer) for disgust. Nevertheless, because of the approach using SMOTE, the obtained outcome suffers from overfitting.

The substitute for facial analysis by video records is the usage of facial EMG. The fEMG was utilized in the research works [162], and [167]. The activity of the facial muscles during expressing emotions was examined. Nevertheless, this kind of test is disrupting and troublesome for patients. Moreover, this kind of evaluation of hypomimia suffers from subjectivity because the patients need to determine their emotional state, there is no, for instance, fixed speech exercise.

The extension of the work from [6] was published in [168]. The gathered dataset contains 39 HC and 47 PD patients. The authors introduced a *Semantic Feature based Hypomimia Recognition Network* (SFHR-NET). One of its components of it is *Semantic Feature Classifier* (SF-C). The role of SF-C is to fit the feature salient map. The semantic loss and classification loss were tuned thanks to the *Progressive Confidence Strategy* (PCS). The neural network consists of a spatial encoder and temporal encoder with *red, green, blue* (RGB) spatial representations and optical flow, respectively. Furthermore, the *Gradient-weighted Class Activation Mapping* (GRAD-CAM) was used to interpret the approximate activate area. This neural network is an end-to-end solution and includes *Visual Geometry Group* (VGG) as the backbone, segmenter, SF-C, PCS, and optical flow. The dataset was divided into a training set of 60 %, a validation set of 10 %, and a testing set of 30 %. The results of classification were equal to 99.39 % accuracy and 99.49 % F1-score. Unfortunately, the dataset is imbalanced in the number of cases from each group, and the cross-validation was not conducted.

Additionally, the detection of PD was performed using transfer learning in [169]. A total of 107 records of PD were collected. The dataset was splintered into the training and testing dataset. The testing set contains 27 HC and 27 PD patients. The labeling of the dataset was done by the two neurologists. The CNN was trained

---

<sup>1</sup><https://www.parktest.net>



on the database Youtube Faces Database. This dataset is a collection of 3245 videos from 1595 people. The used network was VGG. The result of the prediction was the density distribution of the hypomimia score. The obtained *area under the receiver operating characteristic* (AUROC) as a metric of the classification was equal to 0.75. The clinical influence of the medication was examined by the Tufts Clinical Data based on the mean of 3639 frames per video. The evaluation of medication was provided for 33 PD patients and among them 76 % were recognised in the off-medication state, and 67 % were in on medication state. This methodology might be used as an examination of the treatment's impact on PD patients.

Furthermore, it has to be emphasised that the process of defining the kind of emotion by the physicians is subjective. Moreover, the expressing emotions varies among cultures and different for personalities, thereby could be biased on some level [63].

For the unified emotion recognition by the automatic systems, a few solutions with the neural network were introduced and trained on the FER2013 dataset. This set consists of 35 685 images of 7 following feelings: happiness, surprise, disgust, neutral, fear, sadness, and anger [170]. The task is FER. The human recognition of accuracy based on the FER2013 dataset is  $65.5 \pm 5$  % of accuracy [170, 171]. One of the proposed solutions, trained on FER2013, was a *deep neural networks* (DNN) with two convolutional layers, max-pooling, and four Inception layers. The obtained accuracy was equal to 66.4 % [171]. Furthermore, another introduced neural network was simple CNN with the submission of activation function from softmax to linear SVM. The result of classification of this task was equal to 71.2 % accuracy [172]. Moreover, the potential augmentation of emotions was considered thanks to the usage of VGG16 with a soft label constructor. The accuracy of the obtained classification was 73.3 % [173].

Based on the results of the conducted classification on FER2013, the prediction of the emotions by proposed automatic methodology could exceed human abilities. Clinicians may be able to recognize PD more accurately thanks to the utilisation of ML-based solutions to recognize emotions.

Furthermore, there were published works which were dealing with the progress of the disease based on hypomimia symptoms. In one paper [174], the authors collected the records of 727 PD patients. In this study, the author did not consider the classification of PD disease, the HC class was missing. The aim of the exercise was to describe the positive and negative memory by the participants. The measurements of height and width of the eye, mouth, and eyebrow were computed during the features extraction step. The Random Forest Regressor with 9-fold cross-validation was applied for the regression task.

The methods for determining the progress of the PD were presented in [7]. In this

approach, the *Mel-frequency cepstral coefficients* (MFCC) and AU were extracted from audio and video modalities, respectively. The regression method classifies PD patients into four categories for the development of the disease. The collected 772 records were gathered from 117 PD patients. The aim of the exercise was to talk about their positive and negative experiences by them. A *Hierarchical Bayesian neural network* (HBNN-C) was used as a ML method. The scheme of the performed experiment is presented in Fig. 3.4. The multiclass classification was equal to 0.55 F1-score. Unfortunately, the experiment is not repeatable because the dataset is private.

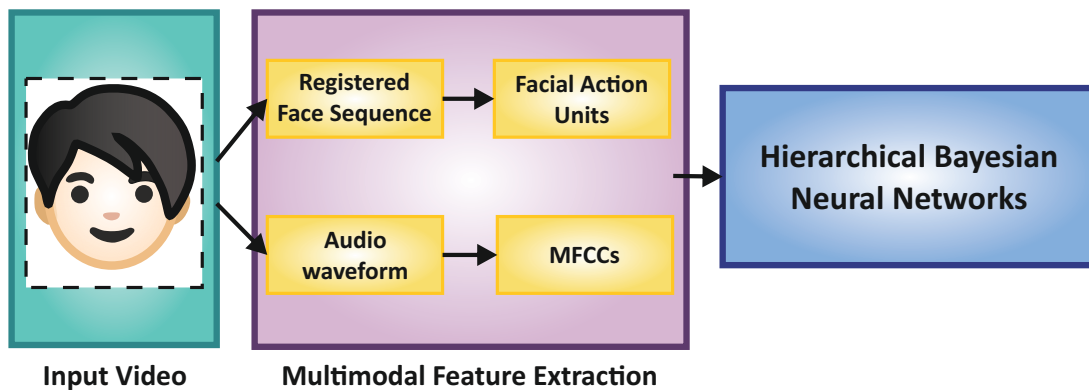


Fig. 3.4: The flow of the experiment from [7].

The detection based on the fusion of modalities for PD detection was presented in [156]. The authors used the audio and video modality. Two types of data were distinguished. One of them was the so-called training dataset which contained the data of 111 HC and 112 PD patients in the ‘on’ phase. The second dataset was a validation dataset and contained records of 74 HC and 74 PD patients in the ‘off’ phase. The device used for gathering the data was the smartphone. The task in the experiment was to read the text by the participants. As the feature extraction step, 20 features were computed. Among them were: gender, age, 6 key mouth- and eye-related features, pitch variance, average pitch, pause percentage, voice volume variance, reading time, and phonetic score. The used ML methods were 10-fold cross-validation and nine classifiers. The 0.85 AUROC was achieved thanks to the Logistic Regression for the training dataset. 0.90 AUROC was obtained with the usage of the AdaBoost classifier for the validation dataset.

Tab. 3.2: Overview of the works analysing hypomimia in patients suffering from PD.

Reference	No. of HC	No. of PD patients	Task	Access	Modality	Comment	Metrics
[162]	8	7	Expressing 6 emotions	Private	Video	Detection	Differences in group for the worked-out functions
[88]	17	17	Expressing emotions: anger, disgust, happiness, sadness	Private	Video	Detection	p-value <0.05
[160]	15	15	Watching funny movies and answering 5 questions	Private	Video	Detection	0.90-0.99 AUC
[110]	12	12	Posing with different emotional expressions	Private	Video	Detection	p-value <0.05
[109]	23	11	Selfie photos	Private	Image	Detection	p <0.05, 0.79 specificity/0.82 sensitivity, 0.58/0.54 sensitivity/specificity
[104]	50	50	Photo	Private	Image	Detection	67% accuracy
[158]	15	-	Watching cartoon	Private	Image	Detection	p-value <0.05
[162]	8	7	Watching movie clip	Private	EMG	Detection	p-value <0.05
[166]	543	61	Expressing 3 emotions	Private	Video	Detection	95.6%
[6]	39	47	Expressing neutral mimic and emotion	On request	Video	Detection	0.9991 F1-score, 0.9997 F1-score
[174]	0	727	Describing patients' negative or positive experience	Private	Video	Regression	0.560 Mean Absolute Error (MAE)
[7]	0	772	Describing patients' negative or positive experience	Private	Video, Audio	Regression, Multiclassification	0.48 MAE, 0.55 F1
[169]	27	27	Assessment of hypomimia and checking the influence of the medication on this symptom	Public/Private	Video	Detection	for class detection: 0.75 AUROC, for comparison of the differences in state between medications: 76% off, 67% on
[168]	39	47	Neutral mimic and smiling	On request	Video	Detection	99.39% accuracy, 99.49 F1-score
[165]	75	91	Assessment of hypomimia and indication on valuable features	On request	Video	Detection	0.87 AUC, 78.3% accuracy, sensitivity 79.1%, 77.8% specificity
[156]	112 'on', 74 'off'	111, 74	Assessment of video and audio records	Private	Video, Audio	Detection	0.85 AUROC for 'on' state, 0.90 AUROC for 'off' state
[157]	13	13	Assessment of eye fixation and gait patterns	Private	Video	Detection	Accuracy, sensitivity and specificity up to 100 %

Furthermore, the method with the usage of multimodality for PD detection which was analysing changes in facial expression was introduced in [157]. This time, eye fixation and gait were taken into consideration. The dataset contains the records of 13 HC and 13 PD patients. The possibility to retain gazing at the defined point was evaluated and defined as the ocular fixation. The fundamentals of using this modality are the deviating frequency of microsaccades eye movement interval for PD patients (5.7 Hz) from HC (1-2 Hz) [157, 175]. In this study, the two kinds of features were computed. They were kinematic features calculated from optical flow and deep features extracted from the convolutional neural network. As the next step, the spatial distribution of the features was used thanks to computing the covariance matrices. The final features were computed as the temporal mean of covariance matrices. The final classification was achieved thanks to the cross-validation leave one-patient-out and Random Forest. The obtained accuracy of PD detection was equal to up to 100 %.

The overview of the studies which considered automatic analysing of the hypomimia for PD detection and progress of the disease is presented in Table 3.2.

### 3.2.2 Dysarthria

Speech is another modality that is the subject of research about PD detection. The representatives of the Speech and Movement Disorders Study Group proposed recommendations for dysarthria acoustic analysis of movement disorders [176]. The published guideline is about extracting the following acoustic features which are expressing HD in the domain of respiration, prosody, phonation, and articulation, i.e.: *standard deviation* (std) of the fundamental frequency, harmonic-to-noise ratio, the std of intensity, mean intensity, vowel space area, shimmer, jitter, voice onset time, diadochokinetic regularity, diadochokinetic rate, etc.

In [97], the basic features recommendation was utilised for the detection of PD. Three groups were taken into account, namely: HC (30), PD patients (30), and 50 individuals suffering from *idiopathic rapid eye movement sleep behavior disorder* (iRBD). The third group was considered because it is the early marker of PD. The data were gathered by smartphone for speech exercises. They were the monologue, the diadochokinetic exercise – repeating of pa-ta-ka, and the sustained phonation of vowel [a]. PD vs. HC, and PD vs. iRBD were classified based on collected records. The outcome of the detection of PD was equal to 0.85 AUC, 75.0 % sensitivity, and 78.6 % specificity, for Logistic Regression. Moreover, the benefit was the usage of smartphone technology for the prodromal detection of the disease. Furthermore, the most profitable biomarkers occurred to be decreased rate of follow-up intervals, inappropriate silences, and the monopitch. Further, the classification of the second

scenario, PD vs. iRBD, had an AUC of 0.78, the sensitivity of 66.7 %, and specificity of 71.0 %.

The studies which were introduced are not dealing with the recommendation of basic acoustic features, however, they proposed new features and solutions for analysing HD. They are presented below.

The set of features based on the spectro-temporal sparsity was proposed in [177]. It was registered that PD has lower temporally sparse speech spectral coefficients than HC. This fact led to the creation of the features. The introduced parameters were: non-parametric sparsity measures (Shannon entropy, Gini-index, l1-norm) and parametric measures (the spare parameters of a Weibull distribution and the shape parameter of a Chi distribution). The dataset consists of records of 45 PD patients and 45 HC of Colombian Spanish native speakers [178]. The achieved accuracy was equal to 83.3 % with the usage of an SVM classifier with a radial basis kernel function. The parametric sparsity measures (shape parameters of the Chi and Weibull distribution) and the Gini index occurs to be the most valuable features for PD detection purpose.

Another solution for PD detection based on speech analysis was introduced in [179]. The authors proposed a forced Gaussian-based methodology that can distinguish PD from HC analysing independently various phonetic units. The dataset was diversified and consists of 50 HC and 50 PD Colombian patients, 32 HC and 47 PD Spanish patients, 20 de-novo PD Czech individuals and 14 HC. The achieved outcome of classification was 87 % of AUC with the usage of cross-corpora validation.

Furthermore, the novel feature for PD detection based on the biomechanical model was introduced in [180]. The feature was named the *absolute kinematic velocity* (AKV). The analysed model was about articulation and speech, whereas the feature was computed thanks to the existing dependency between the jaw-tongue reference position displacements and format oscillations. The considered dataset has 16 PD individuals and gender-, age-matched 16 HC. The analysed feature AKV characterised the increased distinguishing abilities of groups in comparison to format centralisation ratio or vowel space area.

The interpretable features were proposed in [181]. They were in particular articulatory kinetic biomarkers. The 50 HC and 50 PD patients were included in the research. The diadochokinetic speech exercise was taken under analysis. The authors made two observations. Firstly, the envelope of the speech was explored between classes because of its indirect connections to the distribution of forces controlling the articulators. Next, to evaluate the kinetics of speech PD individuals, the velocity of the mid-term air pressure was researched. The envelope of the speech was finally analysed in view of the dependency on the mid-term airflow pressure. Regarding the used ML approaches, the obtained kinematic biomarkers were used

as features, a future selection step was carried out thanks to the sequential floating selection, and the SVM with linear kernel was chosen as the classifier. The obtained accuracy was equal to 85 %.

There are nowadays a registered plethora of studies dealing with deep learning for support methodologies [182, 183]. A few of the works also dealt with utilising the DNN for PD detection based on speech modality [184–186].

The example of utilisation CNN was introduced in [184] for PD detection. The dataset which was used in the research consists of 88 PD and 88 HC German, 50 PD and 50 HC Columbian, and 20 PD and 15 HC Czech individuals. The CNN for the training obtained *short-time Fourier transform* (STFT) and wavelet transform of transitions between the onset and offset of phonation. The detection of the disease for this CNN was 89 % of accuracy. Moreover, this same team used transfer learning for this same dataset and increased the classification accuracy by up to 8% [8]. The analysed languages were Czech, German, and Spanish. The schematic presentation of the applied transfer learning is visible in Fig. 3.5. It could be implied that deep learning has the potential for PD diagnosis tasks.

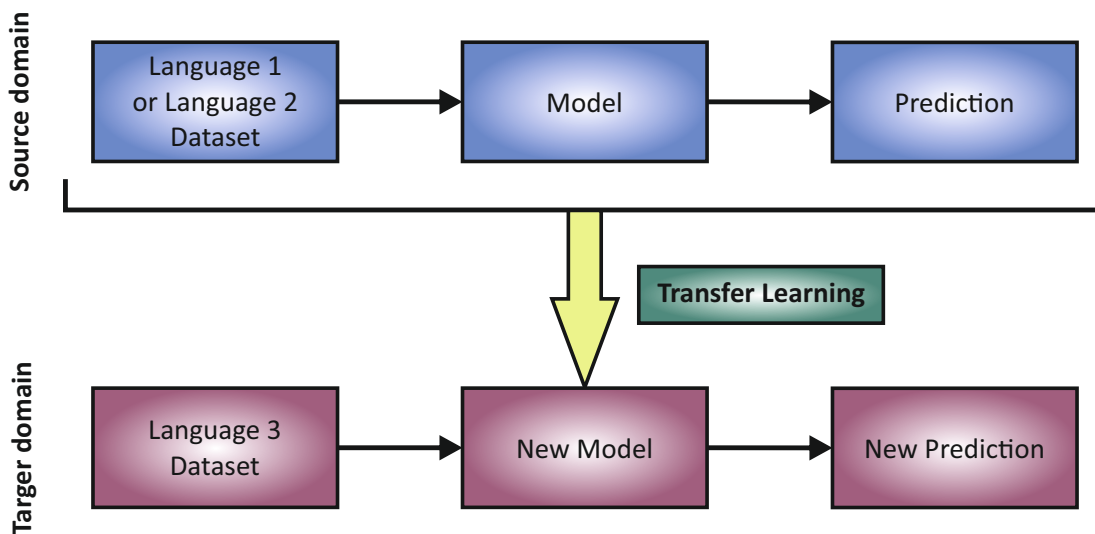


Fig. 3.5: The scheme of the used transfer learning in [8].

Recent studies considered the impact of gender on PD, particularly in speech. Generally, the low-frequency content of the speech was found for men suffering from PD, while high-frequency content was typical for women patients [187, 188]. In this research, into account were taken four datasets and confounding factors. Furthermore, the authors of [189] discovered that women suffering from PD characterise having more satisfactory vocal control. The analysed dataset contained the instance

of 60 male and 40 female HC and the same amount of PD patients.

Generally speaking, speech and voice analysis found a niche in the detection of PD disease. The described works are a good instance of this claim. More information about the used approaches so far could be found in [84,118,119].

### 3.2.3 Methods for Parkinson's Disease Detection based on Sleep Disorders and Actigraphy

This section illustrates the transferable algorithms and methodology dedicated to sleep disorders analysis in PD. Those methods were used earlier as different applications in various diseases and disorders such as ADHD, Alzheimer, and bradykinesia. The analysis of sleep/awake stages was also taken into consideration. The records which were analysed were gathered by wearable - actigraph. The suitable features extraction was discussed and used ML were depicted. They were: SVM, Naive Bayesian, k-NN, XGBoost, or Random Forest, Logistic Regression, GRU, LSTM, and CNN, neural ODE, 1-D CNN, Deep-ACTINet, AdaBoost, *Time-aware Toeplitz Inverse Covariance-based Clustering* (TICC) and CNN (TATC). Additionally, the examples of actigraphs are mentioned, which signals they are collecting, as well as the methods, how to measure the movement by this device, are described. Moreover, this section concentrates on the sleep-related parameters which vary among PD and HC. Furthermore, the processing and storing of data in actigraphs are explained.

#### Monitoring movement with sensors and actigraphs

Actigraphy is the name of the technique to monitor human activity or rest movements and cycles in an unintrusive way. Actigraphy could be used for tracking activity time, sleeping time, resting time, or the development of the illness. The sensors such as gyroscopes, accelerometers, and magnetometers are widely used in wearable actigraphs. The purpose is to measure the body position, movement, and rotation of the body. Moreover, they could monitor the skin temperature, HR, skin conductance, sound, and ambient light [190], [9]. The actigraph could be placed on the wrists, thighs, ankles, and hips. The most valuable position of the wearable on the body appears to be the wrist. This placement of the wearable on the body is especially desirable for PD monitoring [191].

The market offers a few kinds of actigraphs. The most frequently occurring are ActiGraph GT9X Link (made by ActiGraph), 9-axis IMU (mbientlab), wGT3X-BT, GENEActiv (Activinsights), Vivoactive3 (Garmin), Versa (Fitbit), Fit2 Pro (Samsung), Actiwatch Spectrum Pro (Philips Respironics), Charge 3, Vapor (Misfit), Ticwatch E (Mobvoi) [192].

## **Sleep disorders in Parkinson's disease**

The outcome of the disease prediction with the usage of actigraph and PSG could vary. For example, an accurate prediction could be obtained for brain injury and sleeping apnea. On the other hand, the achieved sensitivity for insomnia could be low [190,193]. The differences between the detection of the disease imply the need for a profound understanding of the nature of the sleep disorders in PD to be able to detect this illness on the appropriate level. The PD patients manifest a variance comparing the HCs in the following computed, sleep-related parameters: higher *fragmentation index* (FI), *sleep onset latency* (SOL), longer *wake after sleep onset* (WASO), shorter total sleep time, and lower *sleep efficiency* (SE) [194]. Moreover, the markers of sleep disorders in PD are also: periodic limb movement, RLS, and chronic insomnia [195]. Additionally, nocturnal hypokinesia is characteristic of people suffering from PD [196]. The factor which evokes this symptom is the lower secretion of dopamine. Turns in bed are less frequent in the PD patients group than in the HC group. The speed and acceleration of PD patient are lower and manifest smaller degrees. The typical is also akinesia and often occurring a supine position, especially during the second half of the night.

## **Processing and storing of data in actigraphs**

The methods of the IoMT are used for the transmission of the data from the actigraph [52]. This concept contains the following procedure: data acquisition, communication gateway, and server/cloud. The gateway is a physical device (for example a smartphone) or software [197]). It is considered that the gateway acts as a point of connection between the device and the server or cloud via the field connector - communication protocol. Cloud computing serves as the data holding, advances calculations, the constraint of the exploration of the wearable device (computations are instead performed in the cloud), likewise place for implementing support methodologies [198–200]. The advances in the development of the IoMT concept facilitate the progress of the eHealth solution, silver economy including applications for PD detection [60, 201].

The telemedicine solution could be utilised by doctors for tracking the existence and progress of marks among PD patients such as FOG, sleep, bradykinesia, hypomimia, tremor, and so on [63,202]. The solution with the usage of the smartphone was used in [203] to detect PD. The accelerometer was utilised to collect the data related to gait which were transmitted to the cloud. The proposed application obtained 81 % accuracy.

Additionally, the accuracy could depend on some factors. One of them is the recording modes. The signal could be transmitted without changes or after modi-



fication in mode. Three kinds of modes are distinguished: the *time above threshold* (TAT), *zero-crossing mode* (ZCM), and the *proportional-integral mode* (PIM). The advantage of the usage of recording mode is the reduction of the size of the sent data. Nevertheless, the drawback could be the decreased accuracy of the following designed model [190]. Moreover, the impact on the achieved accuracy of the support model has the position on the body actigraph, time window (so-called epoch), and sampling frequency. Sufficient sampling rate is considered as 20-25 Hz, the sampling rate above that value will not add more value [204].

30 s and 1 min of the epoch is the common value chosen by manufactures as the default value for ML model. The lower value of the epoch makes the quality of the model worse [204, 205].

### **An overview of transferable algorithms and methodologies for sleep disorders recognition with the usage of Actigraphy dedicated to Parkinson's disease detection**

This section is overviewing ML algorithms being frequently utilised for the detection of different diseases than PD with the assistance of wearables. Moreover, the solutions for the recognition of PD based on various modalities besides sleep were considered. The Table 3.3 summarises the potential techniques which could be used for the recognition of PD based on sleep disturbances. The table contains information about the main aim of the research, utilised features and architectures, achieved results, what kind of disease or case is considered, and added comments about the study.

The computed features could be in the statistics domain, time domain, and morphology-based. Moreover, the extracted features could be in the frequency domain, obtained thanks to the Wavelet Transform, or Fourier Transform. Furthermore, there are many techniques for computing the features from the raw signal or based on activity counts. The features could be derived manually or automatically. The most frequently extracted features include: minimum, maximum, entropy, the energy of the signal, kurtosis, skewness [206], 10th, 20th, 50th, 75th and 90th percentiles, std, mean, peak-to-peak amplitude, peak intensity, *interquartile range* (IQR) [207], the sum of values, zero crossings, coefficient of variation, TAT, signal power, *root mean square* (RMS) value of signal components, the difference between maximum and minimum signal peak values, maximum frequencies, the mean normalized frequency of the signal's spectrum, the median normalized frequency of the signal's spectrum [208], band energy, crest factor, and spectral flux [209].

Algorithms can be divided into two groups based on their complexity: Shallow Learning and Deep Learning. First of them are algorithms such as SVM, Naive

Bayesian, k-NN, XGBoost, or Random Forest, Logistic Regression [204]. Because the researchers are suffering from a lack of data, shallow learning obtains comparative outcomes.

The potential application of ML solutions, including deep learning for PD based on sleep records, were taken under analysis. Those solutions were applied for the recognition of sleep/awake status, sleep-correlated illnesses, and daily activity. The architectures which were utilised for those aims are: time-aware architectures such as GRU, LSTM, and CNN, likewise the combinations of algorithms [47, 206, 210]. Promising usage seems to be the application of ODEs [211] or temporal CNN [212]. Moreover, the use of the spectrogram seems to be justified in view of the resemblance of the oscillating speech signal to the accelerometry signal, whereas the spectral representation of this signal was used together with CNN for speech recognition [206].

The common issue which is regarded in the sleep records is the sleep/awake status. The method which is often used for sleep/awake recognition is the events counting activity. This method is named activity counting. Nevertheless, activity counting has not an application in the detection of PD based on sleep records because of its simplicity [213].

The individual-related factors like biological factors, age, and lifestyle-related parameters were analysed in [207] to develop the personalised method for recognition of sleep/awake stages with the usage of actigraph. The dataset contained 54 individuals. The normalisation was performed, and 18 features were computed. Moreover, 21 actigraphy parameters were utilised. The used algorithms were Random Forest, XGBoost, AdaBoost, *Regularized logistic regression* (RLR) with an L2 regulariser and *Stochastic gradient descent* (SGD), and Naïve Bayes. The achieved accuracy was equal to 86 % for XGBoost. For this same classifier, the specificity was equal to 95 %, and the sensitivity was equal to 45 %.

Another solution for the distinction between sleep and awake stages was presented in [9]. The authors used the bidirectional version of LSTM. The scheme of the this architecture is shown in Fig. 3.6. The visible input  $x_t$  on the picture is the feature matrix at time t, and the output  $y_t \in [0, 1]$  is the predicted sleep probability at time t. The data were collected for 186 individuals. They considered various combinations of signals, however, the best occurred to be a 3-axis acceleration signal together with skin temperature. The obtained accuracy was equal to 96.5 % accuracy. Additionally, the solution of LSTM dedicated to real-time applications was tested, and it obtained lower accuracy by 0.2-1 %. By the same token, it applies to real-time solutions. The architecture for real-time application with the usage of LSTM is presented in Fig. 3.7.

The methodology of sleep/awake recognition was performed in [209]. The combination of 1-D CNN and LSTM was proposed as Deep-ACTINet for this purpose.

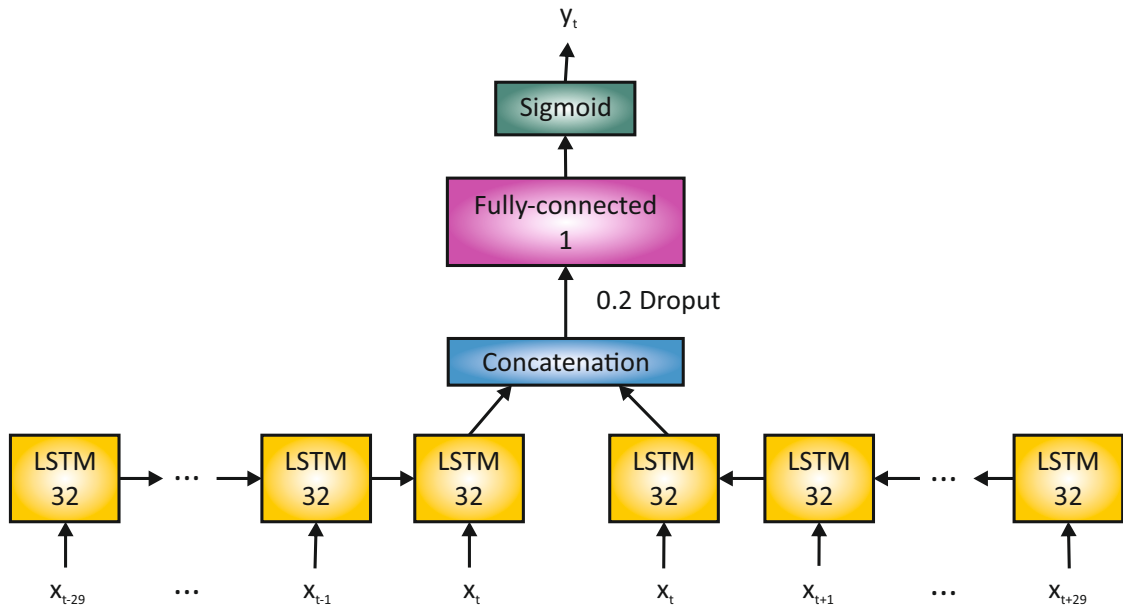


Fig. 3.6: The bidirectional LSTM [9].

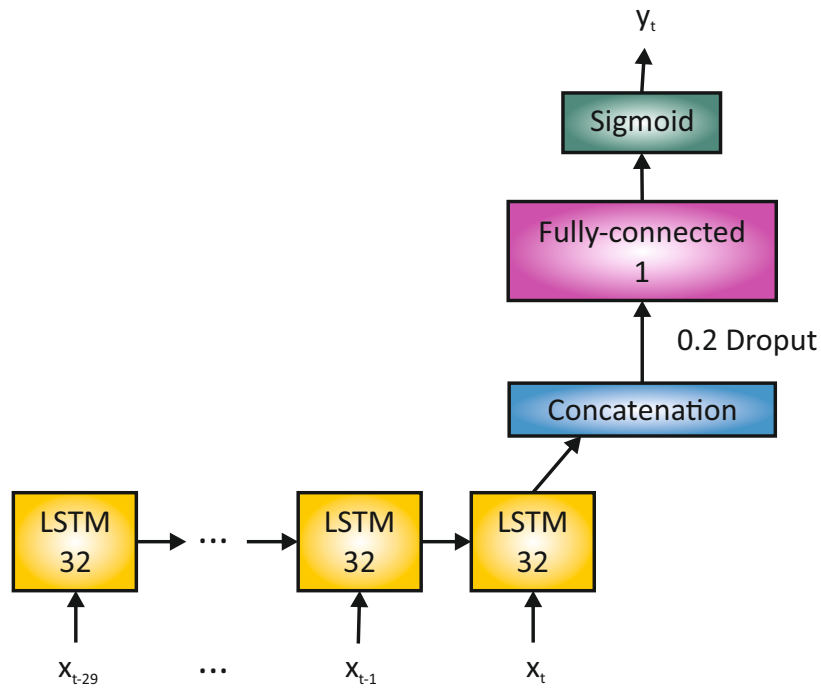


Fig. 3.7: The architecture for real-time detection of sleep with the usage of LSTM [9].

The outcome of the recognition sleep/awake stage was provided also for two common algorithms: Cole-Kripke, and Sadeh, likewise for Naive Bayes, Random Forest,

*Linear Discriminant Analysis* (LDA), and feature-based CNN. The result for the Deep-ACTINet was registered as the highest and the accuracy was equal to 89.65 %.

The common problem which is occurring is the limitation in the size of the dataset. Moreover, the model which is trained on a small dataset could suffer from lower-quality prediction. The answer to this issue could be the usage of the augmentation technique. The promising method of augmentation was proposed in [214] – 3D augmentation. The magnitude scaling, the timewise scaling, and the random rotation belong to them *inter alia*. The dataset was obtained from DREAM PD Digital Biomarker Challenge [214]. Moreover, the techniques such as normalisation and balancing the imbalanced dataset are common procedures for developing solid machine-learning models. The gait of the PD patients was registered which contains the periods of an irregular cyclic pattern, variations in walking speed, and tremors during the quiet standing interval. The obtained accuracy with the usage of normalisation, augmentation, and CNN was equal to 87 %. This solution outperformed the state-of-the-art methods.

The study in [192] considered the data steamed from PSG (Newcastle polysomnography) with the records gathered from the wrist wearable accelerometer sensor. The results of classification sleep/awake stage were equal with the usage of XGBoost to 80 %, and with the usage of XGBoost, and the SMOTE to 84 %. Furthermore, cross-validation was performed.

Another study [206] utilised three datasets including the Daphnet FOG dataset of PD patients. The main idea of the research was to transform the triaxial accelerometer signal into the spectrogram. The proposed technique contained two parts: the unsupervised pre-training part, and the supervised discriminative part. The DBNs were used as architecture. The detection of the FOG was correct at 91.5 %. The key to the successful classification was the pretrained part with the feature extractions.

The utilisation of the spectrogram was also presented in [10]. The authors detect ADHD thanks to feeding CNN with the spectrogram. The used architecture is depicted in Fig. 3.8. The lower activity was thresholded from the signal, not to noise the signal with the informative part of manifesting ADHD. Moreover, the aggregation of the signals was performed as the magnitude of the three-axis and normalisation. The length of the window had also an influence on the final result of classification, and the most desirable length was 300s - the medium size. Nevertheless, the most suitable length of the window for PD detection could vary. The achieved accuracy was equal to 98.6 %.

Another work [215] also dealt with the recognition of ADHD. The signals from the accelerometer and gyroscopes were taken into account. The feature selection was

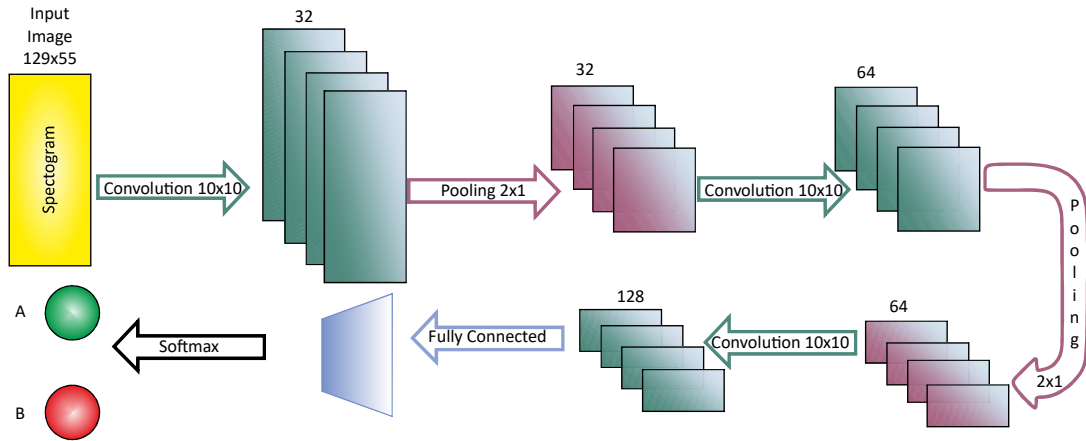


Fig. 3.8: The scheme of the CNN used in [10].

performed and 15 features were chosen. It was estimated that half of the features are 'high-resolution histograms'. Those features are considered as appropriate attributes of early ADHD. The authors used SVM as the classifier, and the obtained accuracy was equal to 95 %.

The analysis of the bradykinesia among PD patients was provided in [47]. The gathered signal came from the 3-axial accelerometer, Shimmer3. The participants of the experiment performed given motor exercises. Thanks to the usage of *deep learning* (DL), the study outperformed the state-of-the-art techniques by 4.6 % accuracy. The AdaBoost, SVM, and k-NN gave worse accuracy than the DL methods. The applied CNN achieved an accuracy of 90.9 %.

The time-aware solution which could consider the circadian rhythm for the detection of Alzheimer's disease was presented in [210]. The data were collected for this aim by the actigraph. The architecture is named TICC and CNN (TATC). Those two parts - first, TICC is an unsupervised technique destined to recognizing stages such as exercising and sleeping, and second - CNN learns temporal features from time series. The uniqueness depends on the ability to analyse and recognise the data collected during various periods of the day (night or day). The data gathered during the day could be emphasized in comparison to those collected during the night, while the GRU and LSTM are destined to focus on long-temporal dependencies. The obtained accuracy was equal to 86.2 %, and, a similar value was registered for sensitivity and specificity.

## Conclusion

It has been confirmed by many discussed research studies that the wearable devices have large potential in early detection of PD. The need for early detection is especially important for this disease, because then the treatment could be faster applied, and could limit the development of the disease. Sleep appears to be a promising modality for PD detection. The usage of wearables such as actigraph could be explored for the eHealth sector.

This subsection is answer to **RO6**, which identified the possible solutions for PD detection based on sleep disorders, including information about used architectures and achieved accuracies. The advantage of the summary is the identification of the potentially transferable applications detecting various diseases that could be used for PD recognition based on sleep disorders. They are ADHD, Alzheimer's, and recognition of the sleep/awake stages. Additionally, the methodology detecting PD based on different modality (bradykinesia) could be transferred. These two examples are the reply to question **RQ6.2**.

The literature shows the variance in the sleep parameters between PD patients and HC. FI, SOL, WASO, shorter total sleep time, and lower SE [194], periodic limb movement, RLS, and chronic insomnia [195], incidents of turns in bed during the night, speed and acceleration of person are the varying sleep parameters and are answers to question **RQ6.3**.

Furthermore, the factors which could impact the final classification of PD patients were identified. They are the length of the analysed signal - time window, the modality, and kind of the gathered signal, the augmentation of the data, the size of the dataset, the feature extraction step, the subset of features used for classification, balancing the dataset and applied architecture. The XGBoost appears to be the most powerful among the classic classifier. The algorithms of deep learning which are recommended are for instance GRU, LSTM, 1-D CNN, CNN, and usage of the spectrogram. Moreover, TICC because of its ability to analyse circadian rhythm seems to be worth considering. The aforementioned algorithms are the answer to question **RQ6.1**. It should be emphasised that the increase in prediction could be potentially feasible by combining the signals from various sensors.

Tab. 3.3: Potential methodologies for detecting PD based on sleep disorders [13].

Main Aim	Features/Architectures	Results	Disease/Case	Comments
Evaluation of temporal and spatial augmentation [214]	Deep CNN	0.87 AUC	The gait of PD patients	Data augmentation improved the prediction
Testing XGBoost and SMOTE for sleep/awake stages detection [192]	XGBoost	80 % accuracy, 84 % with SMOTE	Sleep/awake stages	Using SMOTE improved accuracy; XGBoost was a good choice of algorithm
Evaluation of the performance of personalised factors, Testing various algorithms [207]	<b>Features:</b> 10th, 20th, 50th, 75th, and 90th percentiles, mean, sum of values, coefficient of variation, peak-to-peak amplitude, skewness, kurtosis, signal power, peak intensity, zero crossings, std time above threshold, and maximum value, IQR <b>Algorithms:</b> XGBoost, Naive Bayes, RLR with L2 regularizer, SGD, AdaBoost	86 % accuracy, 95 % specificity, 45 % sensitivity	Sleep/awake stages	Personalised sleep/awake prediction is better, the best observable results are for XGBoost
Detection of ADHD [215]	Features extracted from accelerometer and gyroscope records, SVM	95 % accuracy	ADHD in children	It is possible to recognise ADHD in children. It is needed extension of the database
Usage of CNN and spectrograms for detection of ADHD based on 1-day record [10]	CNN + spectrograms	97.62% sensitivity, 99.52% specificity, AUC values over 99%	ADHD	Deep learning together with actimetry records allows for the detection of the ADHD
Converting triaxial accelerometer signals into spectrograms [206]	Unsupervised pre-training step, supervised discriminative step, DBNs, hybrid approach of deep learning and Hidden Markov Models	91.5 % accuracy	Freezing of Parkinson's gait	Pre-training with feature extractions allows achieving better prediction
Detection of Alzheimer disease in the early stage with time-aware NN [210]	TICC	86.2 AUC	Alzheimer disease	The great potential of using TICC for predicting Alzheimer disease
Comparison of the proposed method with different solutions [209]	Clustering TICC and CNN (TATC)	89.65 % accuracy	Sleep/awake	Achieved detection of sleep/awake stages with the end-to-end deep learning model
Testing bidirectional version of LSTM for sleep/awake stages [9]	Deep-ACTINet (1-D CNN and LSTM)	96.5 % accuracy	Sleep/awake	The solution adequate for real-time application obtained good results
Comparing algorithms according to the detection of bradykinesia [47]	Bidirectional LSTM	90.9 % accuracy	Bradykinesia	CNN outperformed other algorithms about 4.6 %
Checking if it possible to automatically detect periodic limb movements and actigraphy analysis could give results as PSG [208]	CNN	74.2 % accuracy	Periodic limb movements	Actigraphy records with support system methodology could serve as a method for recognising periodic limb movements
Personalised detection of sleep/awake stages [207]	Time-based, frequency-based and signal morphology-based features, LDA 10th, 20th, 50th, 75th, and 90th percentiles, mean, sum of values, std, coefficient of variation, peak-to-peak amplitude, IQR, skewness, kurtosis, signal power, peak intensity, zero crossings, time above threshold, and max value with normalized actigraphy records, Naive Bayes, RLR, SGD, Random Forest, AdaBoost, XGBoost	Up to 91% accuracy	Sleep/awake	The differences in sleep patterns were confirmed between the groups, the highest results were registered for AdaBoost and XGBoost

## 3.3 Deep learning and Time Series Analysis for EEG Analysis

The primary applications of EEG are presented in this section. The seizure detection, and sleep-stages recognition likewise BCI. The ML is used for those purposes likewise the wireless solutions are mentioned. Moreover, the Granger Causality and its advancement Sparse Granger Causality are described together with their use cases. Furthermore, the novelty neural ODE is illustrated. The principle of operation with its advantages are depicted. Additionally, the application of *Ordinal Partition Transition Networks* (OPTNs) is presented.

### 3.3.1 Literature Overview

The most common issue which tackles the usage of EEG is seizure detection, sleep stage recognition likewise BCI [216].

The detection of the seizure brings information about the potential occurrence of epilepsy. The suitable approach for dealing with this problem is the usage of EEG. The likely solutions utilised for this aim are the energy analysis likewise wavelet transform. Moreover, the authors in [134] computed spectrograms to analyse the data in the time-frequency domain. The considered data obtained from the Children's Hospital in Boston. Furthermore, the architectures which were fed with spectrograms were *stacked sparse denoising autoencoder* (SSDA) and ConvA (denoising and convolutional AE). The accuracy reached 94.37 %, and F1-score was 85.34 %.

Additionally, the detection of seizure based on raw signal EEG was proposed in [217]. The data came from the Department of Epileptology at the University of Bonn. The authors applied LSTM and cross-validation for this aim. The accuracy was equal to 95.54 % and AUC to 0.9582.

The gold standard for the measurement of sleep-related parameters is the PSG. A few tests which are typically performed during this medical examination, include: EEG, EMG, ECG, and EOG belong to them. The unified description of sleep is important because this human activity has impacts on memorising, dealing with emotions, and the learning process [216].

The application of 1 channel wireless EEG (Fpz-Cz at 100 Hz) was described in [11]. The solution utilised *Bluetooth Low Energy* (BLE). This technique is dedicated to sleep stage detection because it is more comfortable for the patients than the standard wired 22-channels EEG. The data originated from the Sleep-EDF from Physionet-bank. The accuracy reached 85.3 %. The outcome of the classification was ready after 30 s with the usage of a 1-D CNN. The used architecture is presented



in Fig. 3.9. The part of it - Base-CNN contains three repeated bunch of two 1-D convolutional (Conv1D) layers, 1-D max-pooling, and spatial dropout layers. Next, Base-CNN consists of two Conv1D, 1-D global max-pooling, dropout, and dense layers.



Fig. 3.9: The scheme of the architecture used in [11].

Another solution for sleep stage recognition – scoring was presented in [218]. The data originated from the Sleep-EDF database, and the public dataset *Montreal Archive of Sleep Studies* (MASS). The raw signal gathered by single-channel EEG was taken under the analysis. The problem for the feature extraction appears to be the transition between the sleep stages in the EEG signal. The combination of the mixed CNN (to generate the time invariants features) with BLSTM (to register the metamorphosis between the sleep states) was used as DeepSleepNet architecture. The outcome of classification was equal to 78.9% accuracy, and 73.7 % macro F1-score for Sleep-EDF, likewise 86.2 % accuracy, and 81.7 % macro F1-score for MASS.

The two main components are necessary to control the BCI. There are the assistive signal from the records of brain waves likewise the ML model. The application of noise-proof EEGNet was utilised in [219] to generate, at the same moment, a myriad of features from the signal. The used architecture is a CNN network. The best outcome was obtained for the four datasets, i.e., Feedback Error-Related Negativity, Sensory Motor Rhythm, P300 Event-Related Potential, Movement-Related Cortical Potential. The AUC for prediction based on the P300 dataset was equal to 0.9054.

A detailed review of the usage of deep learning for analysing and creating support methodologies could be found in [135].

### 3.3.2 Casual Inference from Neural Time Series Data

The method which found application in the EEG analysis is the Granger Causality. The advancement of this algorithm – Sparse Granger Causality was proposed in [220], so-called SC-SGA. This architecture characterises the lower complexity of the algorithm and the potential for achieving higher accuracy of prediction. This technique was applied to generate features. Ridge regression, SVM, and Logistic Regression were used as classifiers together with 5-cross validation. Two types of

datasets were investigated for this research. The first of them was *Shanghai Jiao Tong University* (SJTU) Emotion EEG Dataset SEED dataset. The records of three emotions were gathered, i.e., negative, neutral, and positive. The DEAP dataset was the second used dataset, which includes records of 32-channel EEG signals, and additional peripheral physiological signals.

The proposed SC-SGA technique analyses the dependency between the EEG sensors as prior information and chooses the valuable features. This Sparse Granger Causality uses the least absolute shrinkage,  $L_{1/2}$  norm, and the selection operator (LASSO). Moreover, it takes the  $L_2$  norm for Logistic Regression and the Pearson similarity coefficient, whereas the standard Granger Causality utilises the  $L_2$  norm. The SC-SGA technique obtained better outcomes in contrast with the LASSO-GA,  $L_2$  Granger Analysis, and LAPPS. This method (SC-SGA) achieved better accuracy of 2.46 % to 21.81 % on two datasets. Moreover, this algorithm reduces the noise such as the impact of blinking, and head motion.

### 3.3.3 Possibility of Reducing the Number of Training Samples and Increasing Accuracy thanks to the ODE

The paradigm shift in the area of the neural network was introduced at the end of 2018. It is named a Neural ODEs [211]. It is not a representation of the standard sequence of the hidden layers, whereas this network characterises the continuous-profound model. The black-box differential equation solver computes the outcome which assesses the hidden unit dynamics  $f$  if needed to determine the result with the expected accuracy (Eq. 3.2). The advantage of this model is the parameter efficiency. This solution is dedicated to the diversly sampled time series. Moreover, the ODE network is malleable with the constant memory cost at the function of depth. The parameters which can be under control are the profoundness of the model likewise it can be optimised the harmony between the cost of the performed model versus achieved accuracy.

The illustration of the ODE network could be presented as the continuous version of the residual network. The discrete version of the ODE solver could be visualised as the residual blocks in a neural network. This novelty introduced a vector field of continuous smooth transformations [211]. Whereas residual network computes a discrete and approximate sequence of finite transformations. The obtained calculation by ODE is more precise.

The equations of the computations between blocks in the residual network can be expressed mathematically as:

$$h_{t+1} = h_t + f(h_t, \theta_t), \quad (3.1)$$

where  $h_t$  is the hidden information at time step  $t$  ( $t \in \{0, \dots, T\}$ ) in the current block,  $h_{t+1}$  is the hidden information in the following block,  $\theta_t$  is the bunch of the parameters of the model (the weights and the biases),  $f(h_t, \theta_t)$  is the malleable function of the actual hidden information.

The following equation (ODE) will be achieved, when the time step goes to zero:

$$\frac{dh(t)}{dt} = f(h(t), t, \theta), \quad (3.2)$$

where  $h(t)$  is the transition of the hidden information in the infinitely small-time,  $f(h_t, t)$  is the malleable function of the actual hidden information.

The utilisation of ODE was used recently for developing OPTNs in [221]. This method was implemented to detect the casual coupling structures underlying epileptic form activity from rodent brain slices. The microelectrode array was utilised for this purpose. The signal was detected at the onset of the ictal beginning with the usage of bipartite OPTNs.

To summarise, the ODE network could be especially used for the time series analysis because its continuous learning of the data depended on time. The complexity of the algorithm is lower and could be for instance applied to a wireless solution of sleep analysis based on EEG.

### 3.4 Conclusion

The meticulous review of the methodology of COVID-19 recognition and pandemic development with ML and wearables is depicted in this chapter. Moreover, the approaches of PD detection are illustrated. The symptoms which were profoundly discussed are hypomimia, HD, and sleep disorders. They are used for PD recognition. The novel solutions of mHealth and eHealth are taken into consideration. Furthermore, the benefit of this chapter is the presentation of the captivating transferable methodologies which would be appropriate for PD classification based on sleep disorders. The spectrum of difficulties manifesting in sleep disorders among PD patients is depicted to introduce the reader to the topic. Additionally, the techniques of EEG analysis based on deep learning methods are discussed, such as SSDA, 1-D CNN, ConvA, LSTM, CNN, BLSTM, DeepSleepNet. This section addresses the research objective **RO7.** and answers the research question **RQ7.2.** The special attention has neural ODE which is a possible solution to **RQ7.3.**



## 4 Wearables for COVID-19 detection - Practical Solutions

The purpose of this chapter is to investigate the use of wearable devices for the detection of COVID-19. Furthermore, the ML methodology was utilized for this aim. Chapter 2 and Chapter 3 present the state-of-the-art methods and related works, whereas this Chapter 4 presents my research, experiments, and obtained results. The introduced in the thesis ML approaches are built upon two papers [59] and [1] that sought to identify COVID-19 among analysed cohorts. In the first them, 4642 volunteers were involved in the experiment by Stanford University, whereas 114 of them were diagnosed with COVID-19 disease. Additionally, the cohorts had HC group and Influenza. The dataset had records of the heart rate and the number of steps. The sampling rate was 1 per minute. The personal activity was expressed as the heart rate value divided by the number of steps. The research idea in this thesis is the extension of this paper [59]. Nevertheless, the novelty in this thesis is focusing on the data classification problem, not just anomaly detection like in the original paper. The two scenarios were considered. The first of them distinguished COVID-19 cases from HC, while the second scenario focused on the classification of ill cases from HC. The physiological base of the assumption of the thesis was taken *inter alia* from [1]. The resting time of the heart rate is higher for an ill person than in the case of HC group. Moreover, there are differences in heart rate rhythm between COVID-19 and Influenza cases, they last longer and begin earlier. Furthermore, the highest contagiousness period is regarded as -2 to 1 day after the onset of COVID-19, which determines the necessity of early COVID-19 detection. This issue was likewise considered in this thesis. The second dataset from [1] contains the heart rate record and the number of steps but at a different sampling rate. Three groups can be distinguished: COVID-19, Influenza prior to the main pandemic, and Influenza during the main pandemic. The thesis also focuses thanks to the combination of the two mentioned reused datasets on more diverse datasets in terms of demographic. The experiments provided the statistical analysis of the datasets likewise creation of support methodologies suitable for COVID-19 diagnosis in the early stage.

## 4.1 COVID-19 Diagnosis at Early Stage Based on Smartwatches and Machine Learning Techniques

The primary objective of this study was to develop a support system methodology for the early detection of COVID-19 [2, 59]. Furthermore, the criteria were to focus on wearable measurements. This was accomplished by reusing the publicly available dataset prepared by Stanford University with cases of COVID-19, Influenza, and HC. The heart rate record and the number of steps were taken into account. Additionally, this work focused on developing a ML model suitable for use as a screening test. Taking into consideration the contagiousness and incubation periods of the analysing sample, the model should be able to identify which of the analysing sample is ill or healthy during the prodromal stage. The two scenarios were investigated, i.e., COVID-19 detection and Illness recognition when COVID-19 and Influenza were treated as one group and the second was regarded as HC. Developing the appropriate set of features, which captures the time and frequency dependency in the data at various stages of the disease process, was a crucial step. Mann-Whitney U test was used to evaluate the statistical significance of the features. The experiment scheme is shown in Fig. 4.1. Experiments were conducted in the following manner. First, the ratio of heart rate to steps was calculated. Next, the time windows - selected interval taken under analysis, for two scenarios were defined, i.e., COVID-19, Influenza, and HC likewise for COVID-19 and HC. There are three types of windows: five-day, seven-day, and ten-day. For each window, a set of features was computed. The difference in the windows between the later and earlier set of features was then calculated. *Maximum Relevance Minimum Redundancy* (mRMR) was used for feature pre-selection with 50 features. Lastly, stratified cross-validation was performed. The following classifiers were used: Random Forest, Decision Tree, Logistic Regression, SVM, k-NN, XGBoost, and *Generalised Learning Vector Quantisation* (GLVQ). Thus, the results of classification between illness and HC were evaluated in terms of accuracy, sensitivity, specificity, and *Matthews correlation coefficient* (MCC) [2].

### 4.1.1 Data Characterization

Data for this study were obtained from [59]. For the research, the data were collected by Stanford University using wearable devices and the MyPHD app. Data from smartwatches - heart rate, number of steps, and stages of sleep and their duration were analysed. In view of their limited numbers, the sleep data are omitted, whereas the selected dataset contains the records of steps per minute and heart rate per second. This study enrolled 5262 participants. The data were gathered

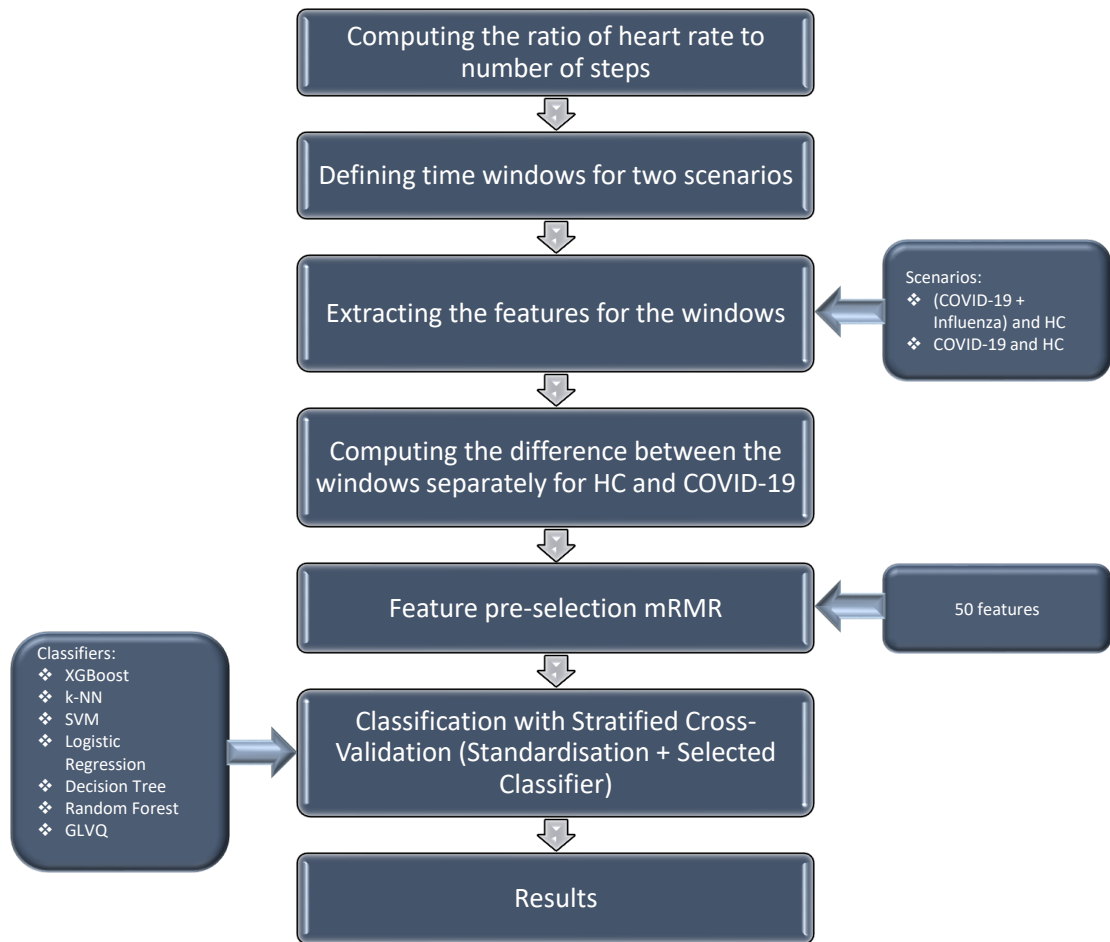


Fig. 4.1: Experiment scheme.

from February 2020 until June 2020. COVID-19 disease was diagnosed in 114 of the participants. The full records of 34 HC, 27 COVID-19 patients were taken under analysis, they were collected by Fitbit device. HC and COVID-19 cases were balanced in this study, along with 7 Influenza cases. The

#### 4.1.2 Feature Extraction and Machine Learning

Physiological data collected from wearable devices are typically continuous time-series records. Data amounts are often quite limited in the majority of cases. Due to this, time-series signals require manual extraction of features [222]. The inspiration for analysing physiological signal features was taken from the following articles: [222–225]. These are the most frequently used features for physiological signals. Unfortunately, there is still a limited number of samples available for measurement. The decision was to use hand-crafted features in order to extract features from these

samples since they span a relatively long period of time. 1-D CNN and LSTM were also evaluated but without any significant results. Three types of features were extracted, i.e., temporal, statistical, and spectral.

A Python package called tsfel [226] was used to extract the features. A diagram illustrating how features are extracted is shown in Fig. 4.2.

Two windows of features were computed for the HC cohort in order to extract the HC samples ( $p_{HC1}$  - earlier window and  $p_{HC2}$  - later window). Windows were fixed in size and separated by a specified spacing. The difference between a set of features for each window was calculated. Where:

- Based on the earlier healthy state, the vector of features is expressed as follows:  $\vec{f}_{HC1}$
- In the later healthy state, the vector of features is expressed as follows:  $\vec{f}_{HC2}$
- For HC, the final vector is as follows:  $\vec{f} = \vec{f}_{HC2} - \vec{f}_{HC1}$
- There is an end point to the earlier healthy state window described as follows:  $t_{HC1}$
- In order to indicate when the later healthy state window begins, it was used the following variable:  $t_{HC2}$

Fig. 4.2 shows the scheme for HC feature extraction.

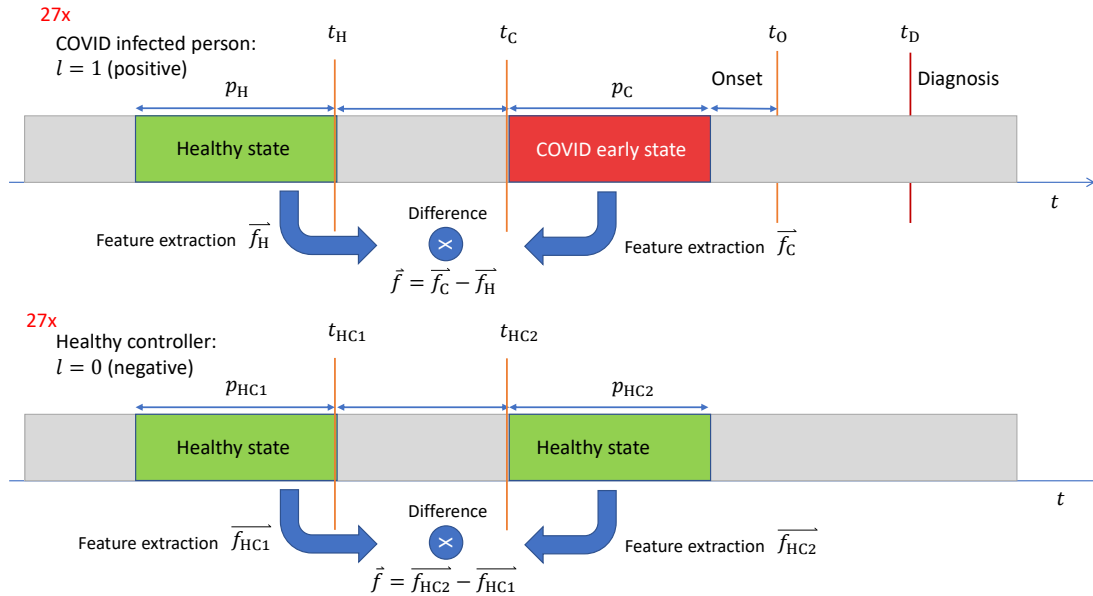


Fig. 4.2: An outline of the feature extraction process for cases of HC and COVID-19.

Similarly, COVID-19 cases were extracted using the same procedure. To detect disease in the prodromal stage, shifts in the computation of windows were defined. Due to the contagious nature of this disease, the highest contagious peak occurs two



days before the disease begins. Following the same steps as when extracting HC samples, the next step was the extraction of COVID-19. As shown in Fig. 4.2, the steps are described in a systematic manner.

Where:

- An analysis of the healthy state yields the following vector of features:  $\vec{f}_H$
- COVID-19 early state features are expressed as a vector as follows:  $\vec{f}_C$
- COVID-19's final vector can be expressed as follows:  $\vec{f} = \vec{f}_C - \vec{f}_H$
- A healthy state window ends as follows:  $t_H$
- COVID-19 begins at the following time:  $t_C$
- Symptoms begin to appear at the following time:  $t_0$
- COVID-19 is diagnosed as follows:  $t_D$
- In terms of the Onset, it is:  $Onset = t_C + p_C$
- The shift between the diagnosis time and the Onset of the disease is expressed as:  $SHIFT = t_D - t_0$

In order to evaluate the algorithms' predictions, metrics such as accuracy, sensitivity, and specificity were computed, as well as the MCC. In order to perform the statistical analysis, the Mann-Whitney U test was used. SVM, Logistic Regression, k-NN, Decision Trees, Random Forests, XGBoosts, and GLVQ were used in this study.

### 4.1.3 Results

It was intended to assess the support system methodology for two cases: in the first scenario, for cohorts containing COVID-19 cases and HC, and, in the second scenario, for cohorts containing COVID-19 cases with Influenza and HC. A 7 day interval between windows is used and a 2 day SHIFT is used (please, check the designation in Fig. 4.2). There is a detailed description of the fixed parameters for experiments in Table 4.1. The extracted features were statistically evaluated using the Mann-Whitney U test with *false discovery rate* (FDR) correction. A table showing the results for scenarios with 5-day windows can be found in Table 4.2 and Table 4.3. A total of 381 features were extracted, and their descriptions can be found at [227].

For the cohort of individuals suffering from COVID-19 disease and HC, Mann-Whitney's U-test revealed the following important features: the sets of MFCC, *Fast Fourier Transform* (FFT) mean coefficient, and *linear prediction cepstral coefficients* (LPCC), spectral slope, maximum frequency, spectral roll-off, spectral kurtosis, fundamental frequency, spectral skewness, zero-crossing rate, slope, min, spectral centroid, median frequency, *empirical Cumulative Distribution Function* (ECDF) percentile and signal distance. For the assumed confidence level  $\alpha = 0.05$ , Table

Tab. 4.1: The scenario in which the experiment was conducted.

Cases	Len_window	SHIFT	Spacing
COVID+HC+Influenza	5	2	7
	7	2	7
	10	2	7
COVID+HC	5	2	7
	7	2	7
	10	2	7

Tab. 4.2: Statistical analysis of COVID-19 cases vs. HC based on Mann-Whitney U test with FDR correction.

Features	pval	pval_FDR	Features	pval	pval_FDR
MFCC_11	0.0045	0.3546	Zero crossing rate	0.0283	0.3546
FFT mean coefficient_117	0.0070	0.3546	LPCC_3	0.0297	0.3546
FFT mean coefficient_189	0.0070	0.3546	LPCC_9	0.0297	0.3546
Spectral slope	0.0102	0.3546	Slope	0.0297	0.3546
FFT mean coefficient_43	0.0134	0.3546	Min	0.0304	0.3546
FFT mean coefficient_254	0.0134	0.3546	FFT mean coefficient_21	0.0309	0.3546
FFT mean coefficient_233	0.0140	0.3546	FFT mean coefficient_243	0.0321	0.3546
Maximum frequency	0.0160	0.3546	FFT mean coefficient_175	0.0333	0.3546
Spectral roll-off	0.0160	0.3546	FFT mean coefficient_144	0.0346	0.3546
FFT mean coefficient_130	0.0174	0.3546	MFCC_7	0.0346	0.3546
MFCC_0	0.0174	0.3546	FFT mean coefficient_163	0.0374	0.3546
FFT mean coefficient_49	0.0189	0.3546	Spectral centroid	0.0374	0.3546
FFT mean coefficient_149	0.0189	0.3546	LPCC_0	0.0388	0.3546
FFT mean coefficient_0	0.0206	0.3546	FFT mean coefficient_249	0.0403	0.3546
FFT mean coefficient_202	0.0215	0.3546	Median frequency	0.0403	0.3546
MFCC_2	0.0224	0.3546	ECDF Percentile_0	0.0409	0.3546
FFT mean coefficient_37	0.0243	0.3546	FFT mean coefficient_39	0.0418	0.3546
FFT mean coefficient_247	0.0243	0.3546	FFT mean coefficient_185	0.0418	0.3546
MFCC_9	0.0243	0.3546	FFT mean coefficient_242	0.0418	0.3546
FFT mean coefficient_167	0.0263	0.3546	FFT mean coefficient_6	0.0434	0.3546
FFT mean coefficient_188	0.0263	0.3546	FFT mean coefficient_57	0.0450	0.3546
Spectral kurtosis	0.0263	0.3546	Signal distance	0.0450	0.3546
Fundamental frequency	0.0274	0.3546	FFT mean coefficient_235	0.0467	0.3546
Spectral skewness	0.0274	0.3546	FFT mean coefficient_154	0.0484	0.3546
Histogram_5	0.0283	0.3546	FFT mean coefficient_194	0.0484	0.3546

4.2 presents the features that passed the test. All of the checked features failed the test after applying the FDR correction. However, it remains a strong criterion. In this case, 0.3546 was the minimum value obtained. A comparison between the two scenarios is possible by comparing the p-values with FDR corrections. For the scenario involving COVID-19, Influenza, and HC cases, the minimum p-value was lower than for the scenario involving COVID-19 and HC cases only, i. e. 0.2389 (see Table 4.3).

Tab. 4.3: Statistical analysis of COVID-19 cases, Influenza vs. HC based on Mann-Whitney U test with FDR correction.

Features	pval	pval_FDR	Features	pval	pval_FDR
FFT mean coefficient_163	0.0024	0.2389	FFT mean coefficient_56	0.0212	0.2421
FFT mean coefficient_243	0.0031	0.2389	Min	0.0216	0.2421
FFT mean coefficient_189	0.0032	0.2389	FFT mean coefficient_236	0.0225	0.2421
FFT mean coefficient_202	0.0040	0.2389	FFT mean coefficient_53	0.0238	0.2421
FFT mean coefficient_149	0.0059	0.2389	FFT mean coefficient_29	0.0245	0.2421
FFT mean coefficient_242	0.0065	0.2389	FFT mean coefficient_15	0.0252	0.2421
Spectral kurtosis	0.0065	0.2389	FFT mean coefficient_165	0.0259	0.2421
FFT mean coefficient_182	0.0070	0.2389	FFT mean coefficient_247	0.0259	0.2421
FFT mean coefficient_167	0.0075	0.2389	FFT mean coefficient_152	0.0267	0.2421
Maximum frequency	0.0094	0.2389	FFT mean coefficient_185	0.0267	0.2421
Spectral roll-off	0.0094	0.2389	Spectral centroid	0.0267	0.2421
FFT mean coefficient_254	0.0101	0.2389	FFT mean coefficient_134	0.0299	0.2586
FFT mean coefficient_117	0.0118	0.2389	Slope	0.0299	0.2586
Histogram_5	0.0122	0.2389	Median frequency	0.0316	0.2615
FFT mean coefficient_25	0.0126	0.2389	Spectral spread	0.0316	0.2615
Zero crossing rate	0.0130	0.2389	FFT mean coefficient_125	0.0333	0.2647
FFT mean coefficient_233	0.0138	0.2389	FFT mean coefficient_155	0.0333	0.2647
FFT mean coefficient_194	0.0147	0.2389	FFT mean coefficient_250	0.0352	0.2738
MFCC_11	0.0152	0.2389	FFT mean coefficient_50	0.0372	0.2832
FFT mean coefficient_43	0.0157	0.2389	FFT mean coefficient_160	0.0382	0.2851
FFT mean coefficient_39	0.0162	0.2389	FFT mean coefficient_230	0.0392	0.2872
FFT mean coefficient_150	0.0162	0.2389	FFT mean coefficient_175	0.0413	0.2874
Spectral skewness	0.0162	0.2389	FFT mean coefficient_255	0.0413	0.2874
FFT mean coefficient_143	0.0167	0.2389	FFT mean coefficient_222	0.0424	0.2874
FFT mean coefficient_0	0.0172	0.2389	FFT mean coefficient_229	0.0447	0.2874
FFT mean coefficient_130	0.0172	0.2389	FFT mean coefficient_231	0.0458	0.2874
FFT mean coefficient_251	0.0172	0.2389	FFT mean coefficient_30	0.0470	0.2874
Spectral slope	0.0177	0.2389	Signal distance	0.0470	0.2874
FFT mean coefficient_235	0.0183	0.2389	FFT mean coefficient_180	0.0483	0.2874
FFT mean coefficient_207	0.0188	0.2389	FFT mean coefficient_196	0.0483	0.2874
FFT mean coefficient_48	0.0212	0.2421	FFT mean coefficient_187	0.0495	0.2874

It is then necessary to explain how the parameters are selected. Due to the registration of the highest contagiousness peak exactly 2 days before the patient's clear onset [71], the shift was set as 2 days. Regarding the incubation period, the window interval was fixed as seven days. 2 days to up to 11 days are considered to be the incubation period. Considering this parameter, 7 days were chosen, so the sum of the later windows (from which the features were calculated) and spacing is greater than the maximum period registered for incubation. A variable in this study was the length of the window, i.e., 5-, 7-, and 10-day windows were tested. Results were obtained with XGBoost, k-NN, SVM, Logistic Regression, Decision Tree, Random Forest, and GLVQ classifiers.

There is a presentation of the best results of the classifications for the cohort

that contains COVID-19 cases and HC in Table 4.4. The most accurate (0.78) with specificity 0.77, sensitivity 0.80, and MCC 0.60, were registered for k-NN during the 5-days window (see Table 4.4). The best sensitivity was registered for GLVQ: 0.81. Stratified cross-validation was conducted on 27 HC and 27 COVID-19 cases. A ML models were also optimized. It was determined that the following parameters were most optimal for the best k-NN: 11 nearest neighbors, Manhattan distance as the distance metric. The weight function was also used (see Table 4.4).

Tab. 4.4: COVID-19 disease detection results for 5-day windows (cohorts: 27 HC, 27 COV).

Classifier	Accuracy	Sensitivity	Specificity	MCC
XGBoost	0.71	0.72	0.71	0.46
k-NN	<b>0.78</b>	0.77	<b>0.80</b>	<b>0.60</b>
SVM	0.65	0.66	0.65	0.33
Logistic Regression	0.69	0.69	0.69	0.41
Decision Tree	0.50	0.52	0.49	0.01
Random Forest	0.62	0.59	0.66	0.27
GLVQ	0.76	<b>0.81</b>	0.71	0.55

There is created second sub-dataset (Subsection 4.1.1) that includes COVID-19 disease cases, Influenza patients, and HC. The COVID-19 disease and Influenza were grouped together, while HC was grouped separately. There was also a balance in the data this time. 34 cases of HC, 27 cases of COVID-19, and 7 cases of Influenza were included in the cohort (see Table 4.5). As a result of the case for a 5-day window, the best accuracy (0.73) and the best specificity (0.76) were recorded for k-NN, as well as the best MCC (0.49). With Logistic Regression, the sensitivity was the highest (0.76). GLVQ obtained also the accuracy equal to 0.73. k-NN was optimized by selecting three nearest neighbors, Euclidean distance as the best distance, and weighing every point equally in the neighborhood. The GLVQ algorithm used distance function squared Euclidean, activation function 'swish', with parameter  $\beta = 3$ , solver type steepest gradient descent with parameters maximum runs = 56, and step size = 3.5. According to the results of the Logistic Regression, the L2 penalty and the 'saga' optimization algorithm are the most appropriate parameters, and the inverse of regularization strength is  $C=464$  (see Table 4.5).

Additionally, there was provided the test for 7- and 10-day windows which are presented for the scenario with COVID-19 and HC in Tables 4.6 and 4.7 likewise for the scenario with COVID-19 and Influenza treated as one class and HC as second class in Tables. 4.8 and 4.9.

The outcomes of the classification for the 7-day windows are shown in Table 4.6. As compared to Table 4.4, the results obtained for classification of 7-day windows

Tab. 4.5: Detection of COVID-19 disease and the presence of Influenza cases within a 5-day window (cohorts: 34 HC, 27 COV, and Influenza 7).

Classifier	Accuracy	Sensitivity	Specificity	MCC
<b>XGBoost</b>	0.66	0.68	0.65	0.35
<b>k-NN</b>	<b>0.73</b>	0.71	<b>0.76</b>	<b>0.49</b>
<b>SVM</b>	0.71	0.75	0.68	0.45
<b>Logistic Regression</b>	0.69	<b>0.76</b>	0.62	0.40
<b>Decision Tree</b>	0.52	0.50	0.55	0.05
<b>Random Forest</b>	0.56	0.56	0.56	0.13
<b>GLVQ</b>	<b>0.73</b>	0.73	0.72	0.47

were lower. In terms of accuracy (0.68), k-NN and GLVQ had the highest results, while XGBoost had the highest specificity (0.66). Additionally, the best sensitivity: 0.84 and MCC:0.38 were observed also for GLVQ.

Table 4.7 presents the results with a 10-day window. It has been found that k-NN produced the best results in accuracy (0.71), sensitivity (0.84), and MCC (0.46). Among the results obtained for Logistic Regression, a score of 0.68 was obtained for specificity.

Tab. 4.6: COVID-19 detection results for 7-day windows (cohort: 26 HC, 26 COV).

Classifier	Accuracy	Sensitivity	Specificity	MCC
<b>XGBoost</b>	0.67	0.68	<b>0.66</b>	0.35
<b>k-NN</b>	<b>0.68</b>	0.73	0.63	0.37
<b>SVM</b>	0.66	0.70	0.63	0.35
<b>Logistic Regression</b>	0.63	0.60	0.64	0.26
<b>Decision Tree</b>	0.54	0.50	0.57	0.08
<b>Random Forest</b>	0.59	0.56	0.61	0.18
<b>GLVQ</b>	<b>0.68</b>	<b>0.84</b>	0.51	<b>0.38</b>

A Logistic Regression model (0.71) produced the best results for scenarios including influenza cases within the 7-day window (Table 4.8). There was also a high level of specificity (0.68) and MCC (0.45) for this classifier. k-NN had the highest sensitivity (0.89). Lastly, k-NN provided the best performance in terms of accuracy (0.73), sensitivity (0.82), and MCC (0.50) for the 10-day window length (see Table 4.9). Logistic regression provided the highest specificity (0.66). Considering the length of the window, the most beneficial analysed interval occurs to be a 5-day time window for both scenarios, i.e. classification of COVID-19 cases likewise detection of ill cases.

Tab. 4.7: The results of the COVID-19 disease detection for 10-day windows (cohorts: 24 HC, 24 COV).

Classifier	Accuracy	Sensitivity	Specificity	MCC
XGBoost	0.65	0.65	0.67	0.34
k-NN	<b>0.71</b>	<b>0.84</b>	0.60	<b>0.46</b>
SVM	0.67	0.67	0.67	0.36
Logistic Regression	0.70	0.72	<b>0.68</b>	0.42
Decision Tree	0.53	0.61	0.46	0.07
Random Forest	0.58	0.53	0.63	0.18
GLVQ	0.66	0.67	0.64	0.34

Tab. 4.8: Detection of COVID-19 disease as well as Influenza cases within a 7-day window (cohort: 33 HC, 26 COV, Influenza 7).

Classifier	Accuracy	Sensitivity	Specificity	MCC
XGBoost	0.63	0.64	0.62	0.28
k-NN	0.70	<b>0.89</b>	0.51	0.44
SVM	0.68	0.80	0.57	0.39
Logistic Regression	<b>0.71</b>	0.74	<b>0.68</b>	<b>0.45</b>
Decision Tree	0.54	0.50	0.57	0.08
Random Forest	0.60	0.55	0.65	0.21
GLVQ	0.66	0.73	0.60	0.36

Tab. 4.9: Detection results for COVID-19 disease and Influenza cases for 10-day windows (cohort: 31 HC, 24 COV, Influenza 7).

Classifier	Accuracy	Sensitivity	Specificity	MCC
XGBoost	0.63	0.63	0.62	0.27
k-NN	<b>0.73</b>	<b>0.82</b>	0.64	<b>0.50</b>
SVM	0.68	0.72	0.65	0.39
Logistic Regression	0.67	0.67	<b>0.66</b>	0.35
Decision Tree	0.52	0.54	0.50	0.04
Random Forest	0.59	0.58	0.59	0.18
GLVQ	0.66	0.73	0.59	0.36

#### 4.1.4 Discussion and Summarisation

An early detection methodology for COVID-19 was presented in this section. In this study, they were considering the data collected from wearables, i.e., heart rate and the number of steps. Based on data gathered by Stanford University [59], the experiment was carried out. Among them, it was selected 27 COVID-19 cases, 7 Influenza cases, and 72 HC cases. It is a limited dataset in the number of cases.

Nevertheless, the sampling rate is satisfactory and it makes this dataset valuable. Additionally, the advantage is that this dataset is publicly accessible. With consideration of the incubation period and the highest contagiousness period, a few scenarios were designed with changeable window sizes, fixed spacing between the windows, and appropriate shifting of the windows. The parameters were chosen specifically to avoid analysing people who were already on quarantine. The difference between the features used in later and earlier stages of the disease was used to determine the nature of the disease. The features were computed in the spectral, frequency, and statistical domains. They were suitable for time signals. It was possible to identify the valuable features through statistical evaluation. It was found that MFCC, FFT, histogram, spectral-based, and LPCC could be used to distinguish between HC and COVID-19. The MFCC are used not only for speech analysis [7], but also were used for recognising abnormal heart rhythm based on ECG [228]. The most important features for the scenarios COVID-19 and Influenza vs. HC were spectral-based, FFT, and MFCC. Nevertheless, after applying stronger criteria, i.e., FDR correction, none of the features fulfilled the requirements with significance level  $\alpha = 0.05$ . However, for the cohort with Influenza, the p-value after FDR correction was lower. Among the classifiers, k-NN appeared to be the most accurate. Furthermore, Logistic Regression and XGBoost performed well in the scenario with COVID-19 and HC. The highest results for k-NN may indicate that there are clear boundaries between clusters in terms of dimension. It was determined that a 5-day time window was the most suitable. Using k-NN, the classifier achieved an accuracy of 0.78, a sensitivity of 0.77, a specificity of 0.80, and a MCC of 0.60. These results allow for the design of a shorter in-time detection methodology for COVID-19 than those presented in the original paper [59]. For cases with Influenza and COVID-19 treated as one group, the accuracy was greater than 0.70 for each type of time window. In this case, the best results were obtained for 5-day, i.e., 0.73 accuracy for k-NN and GLVQ. It was also found that k-NN had the highest specificity (0.76), and MCC (0.49). Based on the simple Logistic regression, it is possible to distinguish between ill cases and HC, and HC is detected with 0.76 specificity. COVID-19 heart-related symptoms tend to begin earlier and last longer than Influenza. Nevertheless, the difficulties in distinguishing ill cases from HC occur [1]. Due to their ease of use and accessibility, likewise relative affordability, wearables are ideal for serving as screening tests. Only pedometers can provide a more accurate measurement of the number of steps taken by a person than smartwatches, but they are not as widely distributed and are less user-friendly.

Data were collected using a variety of devices. During the experiment, people wore a variety of smartwatches, but only Fitbit data were analysed. The data have been unified in relation to the user device. Additionally, it is possible to improve the

results by increasing the number of modalities. There is potential to increase the percentage of detection of COVID-19 in the cohort by incorporating parameters such as skin conductance, skin temperature, acceleration, *Blood Volume Pulse* (BVP), and HRV [229,230]. Nevertheless, the device's quality is reflected in its price. In those modalities is equipped Empatica. It is a medical device that can be used to collect data with greater precision, but it is intended for use with cohorts under observation rather than in real-life standard conditions. Therefore, Empatica cannot be used for screening tests [231]. Additionally, in contrast to the original paper on which the research was based, the support methodology was introduced with classification instead of simply anomaly detection, as in the original paper. While the original paper [59] examined 32 cases of COVID-19, the carried-out research presented in this thesis examined 27 cases of COVID-19. A total of 25 cases were found in original paper to deviate from the norm, with 22 of them occurring at an early stage. The metrics: accuracy, sensitivity, and specificity were calculated in this study. Spectral, frequency, and statistical features were used to classify the state problem. The authors developed both online and offline algorithms in their original work. However, these methods do not have the capability of clearly distinguishing between people with COVID-19 and those with HC. Online algorithm - CuSum - considered 28 days of a person's health. Using this algorithm, 62.5 % of COVID-19 cases were detected, however, the algorithm did not provide specificity.

The research presented in this thesis aims to provide support system methodology, not only anomaly detection. The promising classification was possible through the use of k-NN, which had an accuracy of 78 % and a sensitivity of 77 % for the 5-day window. One of the main advantages of this work is that it does not require a long period to detect the disease. Furthermore, some obstacles to the research should be identified. The demographic distribution of the reused cohort from [59] was not specified. The race, ethnicity, and gender of the applicant were not provided. A number of instances in the dataset were limited. As a result of these two factors, ML models may be biased and overfit. The dataset should be extended in order to improve the classification outcome. The process of collecting the data is another obstacle to obtaining higher accuracy. The research aims to analyse data from a specific device, i.e., Fitbit. Each smartwatch has its own preprocessing steps. Data collected by this particular device were used to train the ML model. Therefore, it is not learned how to differentiate correctly between ill cases gathered by Fitbit and those collected by other devices. Furthermore, wearing the smartwatch during the experiment is a personal responsibility, for this reason, some data may be missing. Furthermore, data collected by smartwatches may be noisy and another sensor may interfere with the collection of data. In addition, Fitbit is not a medical device and does not include a wide range of functionality.



To summarize, the methodology was developed to aid the detection of COVID-19 disease at the prodromal stage. Moreover, the character of the disease was taken into consideration, i.e., incubation and contagiousness period. A model based on five-day windows enables an accuracy of 78 percent in prediction. Several parameters were tested. Realistically, the solution based on 5-day windows is likely to be the most useful and practical. Based on [59], this study was conducted and differs from the original. This research focuses on creating a more powerful classification algorithm than the originally proposed model to detect anomalies. Based on the technology used in this experiment, it may be possible to perform a screening test utilising smartwatches. The study revealed the statistical importance of the majority of features from the statistical and spectral domain based on the Mann-Whitney U test. Moreover, the advantage of this research is that the model results were learned for two different cohorts, i.e., on COVID-19 disease and HC, as well as on COVID-19 disease, Influenza, and HC. Both cases had similar results, with Influenza's extended cohort showing slightly worse results. Considering algorithms, k-NN and Logistic Regression produced the best results, showing that the datasets are not complexly dependent. XGBoost and GLVQ were also successful in some cases. In the future, it may be possible to extend the database and utilise neural networks to handle larger databases. It is also expected that the use of a medical device - Empatica - and the collection of a greater variety of sensor data will enhance the results.

## **4.2 The Distinction between COVID-19 Cases and Two Types of Influenza with Wearable Devices and Machine Learning**

The major objective of this study was to distinguish COVID-19 cases from Influenza cases using ML and wearable technology. The two types of Influenza cases from various periods (before and during the pandemic) were examined. The data were retrieved from [1]. There are records of heart rate and the number of steps. The presented in the thesis support methodologies were developed to confirm the conclusions and assumptions from the original paper regarding the differences in heart rate between the types of viruses tested. Moreover, the incubation and contagiousness periods were taken into consideration to create a solution suitable for early COVID-19 detection. Fitbit was the device used to gather data for this study. The flow of the applied algorithm is visible in Fig. 4.3. As a first step, the time window was selected to extract features concerning the contagious period and the incubation period. The features were also extracted from a 5-day window covering 7 to 2 days prior to the visibility of the onset. A feature pre-selection method was

then applied to select the most valuable features. Next, a 10-fold cross-validation method was used. The applied classifiers were Logistic Regressions, Random Forests, k-NNs, XGBoosts, SVMs, Decision Trees, and GLVQs. Using cross-validation, the classification results were determined.

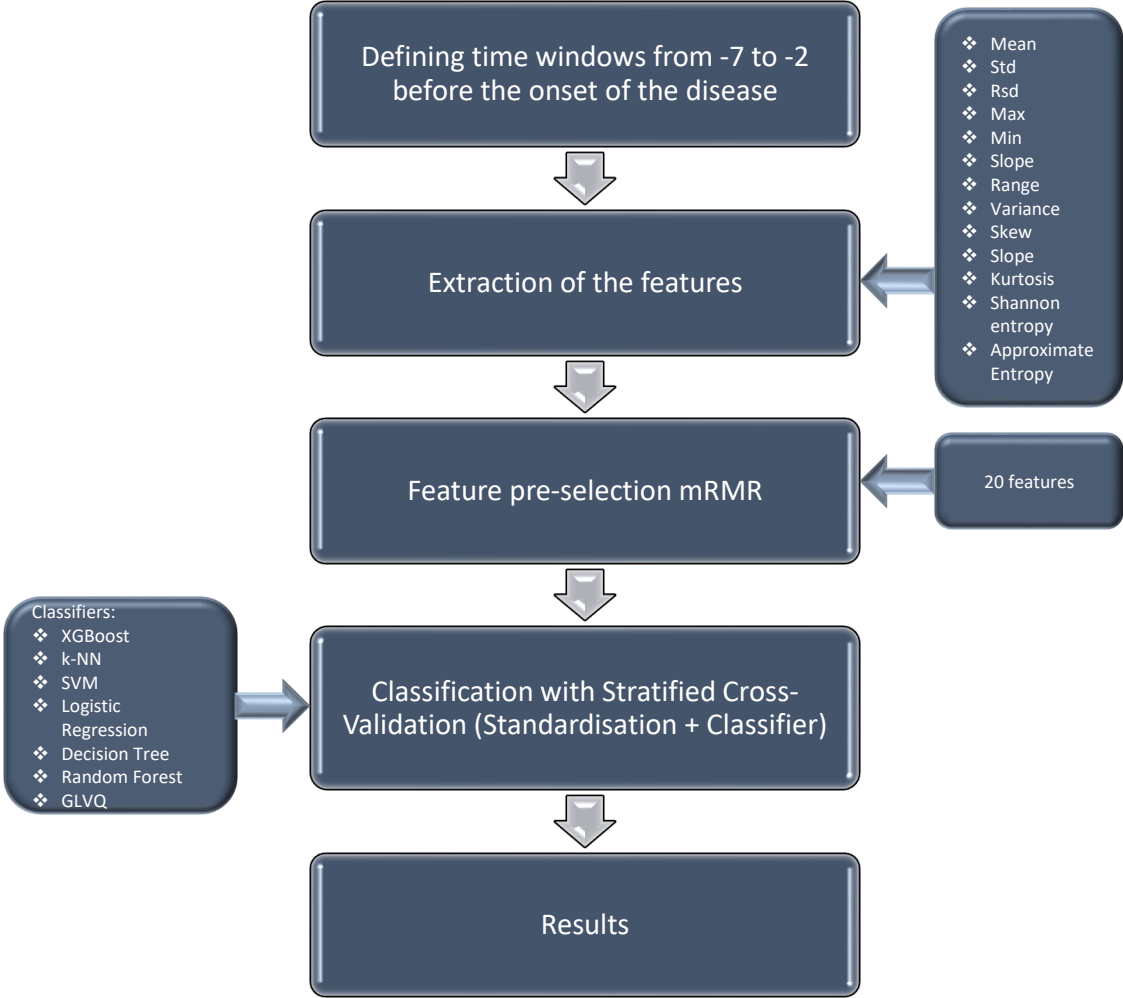


Fig. 4.3: The flow of the algorithm

### 4.2.1 Data Characterization

It is difficult to collect the data by yourself, it would have been needed cooperation with hospitals, local authorities, and ethical approval to organise such an initiative. Therefore, the data were retrieved from [1]. There are two types of Influenza and COVID-19 cases in the dataset. 41 patients with COVID-19, 85 non-COVID-19 flu patients (data collected in the middle of pandemic for Influenza cases which were gathered 03.2020-05.2020), 1126 pre-COVID-19 flu patients (data collected before

the main pandemic for Influenza cases and data were gathered (11.2019-03.2020) were among them. Unfortunately, the HC group was not involved in this study. The dataset has records of the heart rate and the number of steps likewise the lacking records of sleep stages. It was necessary to recompute the data and resample them at a frequency of one sample per day. Due to the lack of data, the dataset for analysis was limited. 37 cases of non-COVID-19 flu, 37 cases of pre-COVID-19 flu, and 21 cases of COVID-19 were included in the recalculated and filled dataset. An overview of the demographic characteristics of the analysed cohort can be found in the original paper [1].

## 4.2.2 Feature Extraction and Machine Learning

A few gradual computations were conducted in order to obtain the features ready for teaching by ML models. In the beginning, the time window was calculated with respect to the period of incubation and contagiousness. The length of the window was 5 days. Generally, the window of time covers the period spanning from seven days before the diagnosis of disease onset to two days before the diagnosis. Several features were calculated for the isolated time window, including *std*, *skew*, *variance* (*var*), *range*, *minimum* (*min*), *maximum* (*max*), *mean*, *kurtosis*, *slope*, and *approximate entropy*. *Accuracy*, *sensitivity*, *specificity*, *MCC*, and *F1-score* were computed to assess the ML algorithms.

## 4.2.3 Results

The purpose of this subsection is to present the results of the experiment. Below is a description of each table in this section:

- Table 4.10: Differentiating Influenza cases during pandemics from cases caused by COVID-19; It was checked if there are differences between diseases caused by two different viruses during the same analysed period.
- Table 4.11: Identifying Influenza cases during the pandemic and before the pandemic; The existing possible differences between Influenza, which cases were registered in various periods, were taken under analysis.
- Table 4.12: Distinguishing the Influenza cases before the main pandemic and COVID-19 cases; It was evaluated if there are discrepancies between viruses collected in various intervals. One of the virus was COVID-19.
- Table 4.13: Multiclass classification of Influenza before the main pandemic, Influenza during the main pandemic, and individuals with COVID-19; The possibility to distinguish parallelly three types of diseases caused by distinct viruses was checked.

Tab. 4.10: Identifying Influenza cases during the middle of the pandemic and cases of COVID-19.

Classifier	ACC Bal	Accuracy	Sensitivity	Specificity	MCC
XGBoost	0.67 ± 0.20	0.71 ± 0.19	0.56 ± 0.37	0.81 ± 0.19	0.38 ± 0.42
k-NN	<b>0.73 ± 0.19</b>	<b>0.77 ± 0.16</b>	0.58 ± 0.33	<b>0.87 ± 0.16</b>	<b>0.49 ± 0.39</b>
SVM	0.64 ± 0.20	0.66 ± 0.19	0.56 ± 0.34	0.72 ± 0.23	0.29 ± 0.42
Logistic Regression	0.68 ± 0.20	0.70 ± 0.18	<b>0.61 ± 0.35</b>	0.75 ± 0.22	0.38 ± 0.42
Decision Tree	0.58 ± 0.20	0.62 ± 0.19	0.44 ± 0.35	0.72 ± 0.25	0.17 ± 0.44
Random Forest	0.58 ± 0.20	0.61 ± 0.19	0.50 ± 0.36	0.67 ± 0.25	0.18 ± 0.42
GLVQ	0.70 ± 0.19	0.74 ± 0.17	0.57 ± 0.34	0.83 ± 0.21	0.43 ± 0.40

The outcome of the classification of COVID-19 cases and Influenza cases during the pandemic are shown in Table 4.10. The most successful algorithm in balanced accuracy, specificity, and MCC was k-NN. The balanced accuracy was equal to 0.73, specificity achieved 0.87 and MCC was 0.49. The sensitivity was the highest for Logistic Regression and was equal to 0.61.

Tab. 4.11: Identifying Influenza cases before and during the pandemic.

Classifier	ACC Bal	Accuracy	Sensitivity	Specificity	MCC
XGBoost	0.80 ± 0.15	0.80 ± 0.15	0.86 ± 0.18	<b>0.74 ± 0.24</b>	0.63 ± 0.29
k-NN	0.79 ± 0.14	0.79 ± 0.14	0.89 ± 0.15	0.69 ± 0.24	0.61 ± 0.27
SVM	0.79 ± 0.15	0.79 ± 0.15	0.91 ± 0.14	0.67 ± 0.25	0.61 ± 0.29
Logistic Regression	0.76 ± 0.16	0.76 ± 0.16	0.81 ± 0.20	0.71 ± 0.24	0.55 ± 0.32
Decision Tree	0.80 ± 0.14	0.80 ± 0.14	0.95 ± 0.13	0.66 ± 0.24	0.64 ± 0.27
Random Forest	0.78 ± 0.15	0.78 ± 0.15	0.87 ± 0.18	0.69 ± 0.24	0.59 ± 0.29
GLVQ	<b>0.82 ± 0.13</b>	<b>0.82 ± 0.13</b>	<b>0.96 ± 0.12</b>	0.68 ± 0.23	<b>0.68 ± 0.24</b>

In the second experiment, Influenza cases were classified before and during a pandemic. The outcome is presented in Table 4.11. **GLVQ! (GLVQ!)** achieved the best results. This classifier obtained achieved the highest balanced accuracy, i.e., 0.82. sensitivity (0.96), MCC (0.68), while XGBoost achieved the highest specificity (0.74).

The results of the distinction between COVID-19 cases and the Influenza cases registered before the main pandemic are shown in Table 4.12. The most successful in balanced accuracy among all tested classifiers was the GLVQ (0.84). This classifier also achieved the best MCC (0.71). Moreover, the best specificity was equal to 0.77 for k-NN and the best specificity was registered for SVM.

The multiclass classification results indicate that the k-NN is capable of distinguishing the cases at a level of 0.64 F1-score (Table 4.13). The balanced accuracy for this same classifier was equal to 0.69 also for k-NN and MCC was equal to 0.54 for k-NN.

Tab. 4.12: Identifying COVID-19 cases and Influenza cases prior to the main pandemic.

Classifier	ACC Bal	Accuracy	Sensitivity	Specificity	MCC
XGBoost	0.80 ± 0.19	0.84 ± 0.16	0.93 ± 0.14	0.68 ± 0.35	0.64 ± 0.37
k-NN	0.83 ± 0.15	0.85 ± 0.13	0.89 ± 0.15	<b>0.77 ± 0.27</b>	0.69 ± 0.28
SVM	0.82 ± 0.17	<b>0.86 ± 0.14</b>	<b>0.96 ± 0.10</b>	0.68 ± 0.33	0.68 ± 0.33
Logistic Regression	0.78 ± 0.17	0.79 ± 0.16	0.82 ± 0.19	0.74 ± 0.31	0.57 ± 0.34
Decision Tree	0.75 ± 0.19	0.79 ± 0.17	0.89 ± 0.21	0.61 ± 0.34	0.54 ± 0.38
Random Forest	0.74 ± 0.18	0.73 ± 0.18	0.70 ± 0.25	0.79 ± 0.31	0.50 ± 0.36
GLVQ	<b>0.84 ± 0.17</b>	<b>0.86 ± 0.14</b>	0.92 ± 0.13	0.76 ± 0.31	<b>0.71 ± 0.32</b>

Tab. 4.13: Result of multiclass classification for COVID-19, Influenza cases prior to the main pandemic, and Influenza during the main pandemic.

Classifier	F1-score	ACC Bal	Accuracy	MCC
XGBoost	0.61 ± 0.17	0.63 ± 0.15	<b>0.67 ± 0.14</b>	0.50 ± 0.22
k-NN	<b>0.64 ± 0.17</b>	<b>0.69 ± 0.13</b>	0.66 ± 0.15	<b>0.54 ± 0.22</b>
SVM	0.62 ± 0.16	0.64 ± 0.15	<b>0.67 ± 0.14</b>	0.51 ± 0.21
Logistic Regression	0.56 ± 0.16	0.58 ± 0.16	0.59 ± 0.15	0.39 ± 0.24
Decision Tree	0.46 ± 0.09	0.53 ± 0.09	0.62 ± 0.10	0.42 ± 0.18
Random Forest	0.48 ± 0.15	0.52 ± 0.14	0.57 ± 0.14	0.35 ± 0.22
GLVQ	0.61 ± 0.17	0.63 ± 0.15	0.66 ± 0.14	0.49 ± 0.23

#### 4.2.4 Discussion and Summarisation

The purpose of this study is to determine if it is possible to distinguish between COVID-19 cases, Influenza cases before the main pandemic, and Influenza cases during the pandemic. ML and wearable devices were used for this purpose. The answer was presented through the summary of results in tables after analysing a few combinations of the problem. Because of the fact that the data were imbalanced, the balanced accuracy was calculated. The classification of COVID-19 cases and Influenza cases before the main pandemic achieved the best distinction in terms of balanced accuracy for the performed scenarios. During a pandemic, people's lifestyles changed and also their activities changed. Quarantine has a significant impact on the lives of people. In addition, differences in heart rates between Influenza and COVID-19 cases were reported in [1]. It is possible that these observations could explain the good results obtained for the classification process. It could be interpreted that the SVM classifier was able to detect boundaries between the datasets. Furthermore, the algorithm was able to identify Influenza cases more easily than COVID-19 cases. There is also a good level of distinction between the two types of Influenza: the one prior to the main pandemic and the one in the middle of the

pandemic. It could be explained by the change in lifestyles that occurred during the pandemic, as well as the fact that there were two types of Influenza viruses and flu during this period. From a medical perspective, the most significant classification was the distinction between COVID-19 and Influenza cases during the pandemic. Based on this analysis, it is possible to distinguish the cases at a level of 0.73 balanced accuracy. This finding confirms the hypothesis from the original work [1] that there is a difference in physiological signals: heart rate and personal activity before the onset of the disease likewise the intensity of the changes in symptoms. Logistic regression was the most successful in identifying COVID-19 cases (0.61 sensitivity), whereas k-NN was more successful in identifying Influenza cases (0.87 sensitivity). It was easier for the classifier to identify Influenza cases than COVID-19 cases based on the binary classification for COVID-19 and Influenza. A multiclass classification, however, indicates that there is a low probability of distinguishing between cases (0.64 F1-score). Research limitations include the inclusion of people who are perhaps self-quarantined. Nevertheless, the research is limited due to the absence of the HC group. Consequently, the generated models could not be used as a screening tool. Regarding the statement in the original paper, there were statistically significant differences between races and ages. Additionally, the cohort was gathered in the *United States of America* (USA), which may have influenced the results. There was also probably a higher rate of hospitalization in the class during the pandemic than before it. According to the original paper [1], the RHR collected by wearables was observed to be higher for COVID-19 cases than for Influenza cases, and the changes lasted longer. Additionally, it should be noted that social contact restrictions in the USA may differ from those in other countries.

To summarize, the research aimed to distinguish between each type of case, i.e.: COVID-19 cases, Influenza before the main pandemic, and during the pandemic. To accomplish this purpose, ML methodologies and wearable devices were used. The analyses were performed based on the data from [1] along with heart rate records and a number of steps. The most important of four performed classifications show that COVID-19 cases and Influenza in the middle of the pandemic can be distinguished with a 0.73 balanced accuracy via k-NN. Moreover, the contribution of this study is the introduction of models differentiating two types of influenza likewise COVID-19 cases vs. Influenza cases before the main pandemic. The achieved balanced accuracies for GLVQ were equal to 0.82 and 0.84, respectively. Several factors could be responsible for the differences between the analyses, including the different types of influenza, differences in symptoms associated with heart rate and self-quarantine, as well as changes in people's lifestyles. The sampling rate and the size of the dataset are the limitations of this study. In addition, the lack of a HC group makes it impossible to create screening tests based on those models.

### 4.3 Wearable Analytics and Early Diagnostic of COVID-19 Based on Two Cohorts

The purpose of this study was to combine two datasets previously explored in chapters 4.1 and 4.2. Both datasets considered COVID-19 disease. First of them with a higher sampling rate contains COVID-19 cases, Influenza, and HC (referred to as dataset A). The second dataset does not have HC cases, however, it has the representatives of two types of Influenza – before the main pandemic and during the pandemic, and COVID-19 cases (referred to as dataset B). A combination of the datasets should result in obtaining a more diverse and flexible ML model than was created in section 4.2 for the dataset with a lower sampling rate. Both datasets (A and B) contain heart rate records and steps taken. They were collected by the wearable – Fitbit device. A major difference between the datasets was the homogeneity of dataset B, which primarily represented overweight individuals from the USA. The scheme of the carried-out experiment is presented in Fig. 4.4.

Tab. 4.14: The combinations of datasets and classes for each experiment.

Types of data	Experiment 1a	Experiment 1b	Experiment 2	Experiment 3	Experiment 4	Experiment 5
COVID-19 A	1	1	1	1	1	2
COVID-19 B	1	1		1	1	2
Influenza A			1	0	1	1
Non-Covid-19 Flu				0	1	1
Pre-Covid-19 Flu				0	1	1
HC	0	0	0		0	0

In the beginning, dataset A was undersampled in order to unify both datasets. Two modalities were sampled at a rate of one sample per day. The next step was to merge the datasets. The 5-day time window for each time series was then provided. According to the contagiousness of the disease and the incubation period, the time window was extracted from -7 to -2 days before the onset of the disease. Next, the features were calculated. They are mentioned in Fig. 4.4. Then, the ratios of the heart rate-related features to the number of step-related were computed. mRMR was used to select the features. Twenty features were chosen from a total of 36. Additionally, 10-fold stratified cross-validation was conducted. Classifiers used in this study included k-NN, SVM, Random Forest, Decision Tree, Logistic Regression, XGBoost, and GLVQ. Experiments were conducted according to Table 4.14. There were the binary classifications as well as a multiclass classification. The experiments vary in analysed classes, from which datasets data were taken, likewise the number of chosen features during the features selection step. For example, Experiment 1a treated COVID-19 cases from A and B datasets as one group and HC as a second class. The whole set of features was used in the classification step.

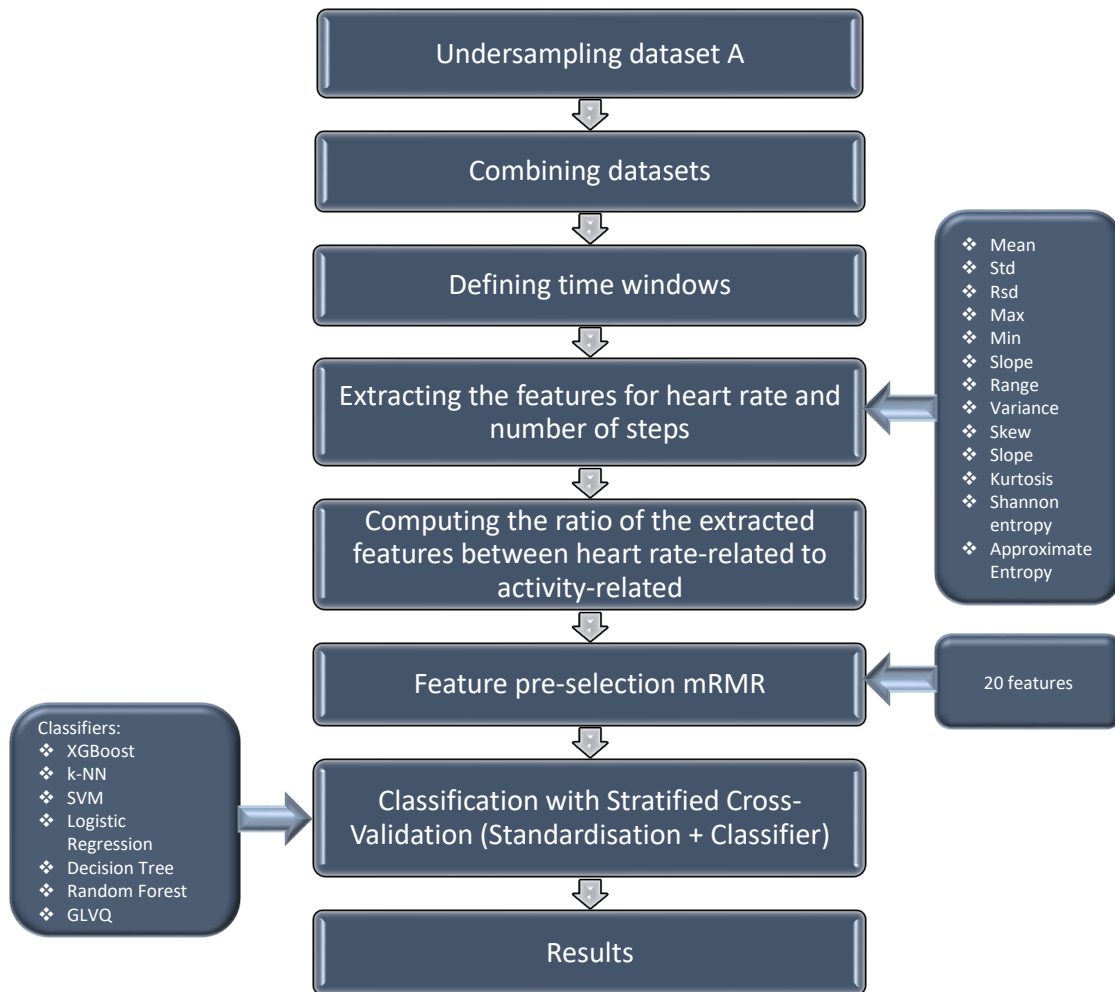


Fig. 4.4: Flow of the algorithm.

Whereas, Experiment 2 considered COVID-19 and Influenza cases from dataset A as one class, while HC was the second group. The number of used features in Experiment 2 was 20. To analyse the data, the Mann-Whitney U test was used with FDR correction. The confidence level  $\alpha = 0.05$ .

### 4.3.1 Dataset Characterization

The dataset was merged from two analysing in the previous subsection datasets 4.1.1 and 4.2.1. The two datasets vary in sampling rate. The data from 4.1.1, here named dataset A, already mentioned before, have COVID-19, Influenza, and HC cases. The second dataset B contains COVID-19 cases (data were collected from 03.2020 to 05.2020), and two types of Influenza: gathered before the main pandemic (11.2019-03.2020) and during the COVID-19 pandemic (03.2020-05.2020). Both



Tab. 4.15: Dataset mixtures and sample numbers for each experiment.

Types of data	Experiment 1a	Experiment 1b	Experiment 2	Experiment 3	Experiment 4	Experiment 5
COVID-19 A	27	27	27	27	27	27
COVID-19 B	21	21	0	21	21	21
Influenza A	0	0	7	7	7	7
Non-Covid-19 Flu	0	0	0	19	10	20
Pre-Covid-19 Flu	0	0	0	20	9	21
HC	48	48	34	0	74	48

types of data contain the record of heart rate and the number of steps taken during the day. For dataset A, the sampling rate for heart rate was 1 per minute and for the number of steps taken during the 1 hour. Dataset B characterises collected samples per day for both parameters. From the first dataset were chosen 27 COVID-19 cases, 15 Influenza, and 73 HC. 21 COVID-19 cases, 37 Non-COVID-19 Flu, and 675 Pre-COVID-19 Flu were in dataset B. A few scenarios of experiments were performed - classification among cases, which may be found in Table 4.15. Data balance was taken into account.

### 4.3.2 Feature Extraction and Machine Learning

The features were extracted based on the [2, 59, 61]. Based on the contagiousness period and the incubation period, the features were obtained. The highest contagiousness period is regarded as -2 to 1 days after the beginning of the onset of the disease. Five days were set as the length of the windows. A time window of -7 to -2 days prior to the onset of the disease was selected in order to detect the disease at its prodromal stage. To unify two datasets, dataset A was undersampled. The features were extracted according to the receipt from [61]. Fig. 4.5 illustrates the scheme for calculating the time window with respect to the mentioned onset disease.

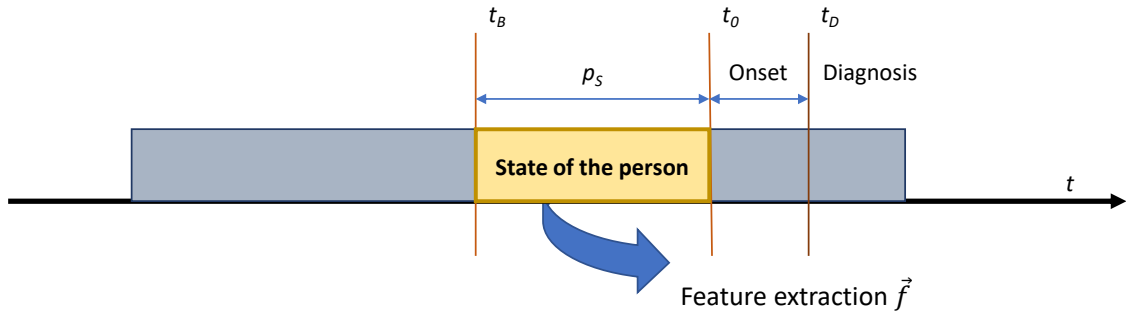


Fig. 4.5: Feature extraction.

The following time points are computed:

- $t_D$  is the detection of the disease  $t_D = t_0 + 2$
- $t_0$  is the visible Onset of the illness  $t_0 = p_S + t_B$
- $p_S$  is the duration of the time window
- $t_B$  is the beginning of the disease

Subsequently, a few parameters were calculated for the time windows of heart rate and the number of steps for the datasets A and B. These were: max, min, mean, std, *relative standard deviation* (rsd), range, Shannon entropy, approximate entropy, skewness, kurtosis, variance, and slope. Furthermore, the ratios between the parameters of heart rate and the number of steps were calculated. For example, the maximum heart rate to the maximum number of steps taken to express changes in personal activity was computed.

### 4.3.3 Results

The statistical analysis, as well as the binary classification and multiclass classification, were conducted according to Tables 4.14 and 4.15.

The outcome of the Mann-Whitney U test and its variant with FDR correction is presented in:

- Tab 4.16 for COVID-19 cases vs. HC
- Tab 4.17 for COVID-19 cases, Influenza vs. HC.

The procedure of Benjamini-Hochberg was carried out to minimise the impact of the number of type I errors. For the first case, having analysed COVID-19 cases vs. HC, the 13 features were statistically significant after FDR correction. For the scenario with COVID-19 cases, Influenza vs. HC, 17 features passed the test with FDR correction for the confidence level  $\alpha = 0.05$ .

The distinction of COVID-19 (cases from datasets A and B) from HC gave 0.73 accuracy for XGBoost, and this same classifier achieved 0.75 sensitivity and 0.48 MCC. The results are presented in Table 4.18. Whereas, k-NN had the best specificity of 0.77. The balance of the dataset was taken into consideration for all provided classifications. The details of used samples for each classification and regarded classes could be found in Tables 4.14 and 4.15.

A second analysis was conducted using the identical sub-dataset but after feature selection, using 20 features out of 36 (Table 4.19). XGBoost obtained 0.73 accuracy, 0.71 sensitivity, and 0.48 MCC. While GLVQ achieved a specificity of 0.91.

The outcome of distinction COVID-19, Influenza vs. HC (all from dataset A) are visible in Table 4.20. The highest accuracy was observable for XGBoost, and it was equal to 0.63, whereas MCC was equal to 0.28 and was the best among all tested classifiers. SVM achieved the best sensitivity: 0.90, while Decision Tree has the highest specificity of 0.67.

Tab. 4.16: Analysis of COVID-19 cases, Tab. 4.17: Analysis of ill and HC cases and HC cases using the Mann-Whitney U test and FDR correction. Using the Mann-Whitney U test and FDR correction.

Feature	pval	pval_FDR
activ_std	0.00001	0.00014
activ_variance	0.00001	0.00014
activ_range	0.00001	0.00015
activ_rsd	0.00003	0.00027
activ_entropy_shannon	0.00092	0.00644
heart_std	0.00184	0.00921
heart_variance	0.00184	0.00921
heart_rsd	0.00291	0.01249
heart_range	0.00321	0.01249
steps_entropy_shannon	0.00750	0.02624
steps_range	0.01214	0.03540
activ_max	0.01138	0.03540
steps_slope	0.01593	0.04288
steps_std	0.02331	0.05439
steps_variance	0.02331	0.05439
steps_max	0.03409	0.07457
heart_max	0.03736	0.07692
heart_entropy_shannon	0.04967	0.09658
steps_skew	0.05699	0.10498
activ_mean	0.06626	0.11596
steps_kurtosis	0.09272	0.15453
activ_min	0.09860	0.15687
steps_mean	0.10960	0.16224
steps_min	0.11125	0.16224
heart_mean	0.12333	0.16602
activ_slope	0.12333	0.16602
steps_rsd	0.14433	0.18709
heart_min	0.24971	0.31214
activ_kurtosis	0.27791	0.33541
heart_slope	0.38690	0.45139
heart_skew	0.42406	0.47878
heart_kurtosis	0.64317	0.70347
activ_skew	0.92736	0.98357
steps_approx_entropy	1.00000	1.00000
heart_approx_entropy	1.00000	1.00000

Feature	pval	pval_FDR
heart_std	0.00000	0.00002
heart_range	0.00000	0.00002
heart_variance	0.00000	0.00002
activ_std	0.00000	0.00002
activ_range	0.00000	0.00002
activ_entropy_shannon	0.00000	0.00002
activ_variance	0.00000	0.00002
heart_rsd	0.00002	0.00009
activ_rsd	0.00002	0.00009
heart_max	0.00005	0.00019
heart_entropy_shannon	0.00010	0.00031
heart_mean	0.00040	0.00118
steps_entropy_shannon	0.00303	0.00815
heart_min	0.00395	0.00987
activ_max	0.00952	0.02222
activ_min	0.01776	0.03884
activ_mean	0.01951	0.04017
steps_min	0.05476	0.10648
steps_max	0.10284	0.17997
steps_skew	0.09805	0.17997
steps_mean	0.11923	0.19872
steps_slope	0.13656	0.21726
steps_range	0.21119	0.32138
steps_std	0.26776	0.36045
steps_variance	0.26776	0.36045
steps_kurtosis	0.26611	0.36045
activ_kurtosis	0.44079	0.57139
heart_skew	0.57789	0.72237
steps_rsd	0.65872	0.75941
activ_slope	0.67262	0.75941
activ_skew	0.63672	0.75941
heart_slope	0.77282	0.84528
heart_kurtosis	0.88901	0.94289
steps_approx_entropy	1.00000	1.00000
heart_approx_entropy	1.00000	1.00000

Tab. 4.18: The outcome of distinction COVID-19 (A and B dataset) from HC (for all 36 features).

Classifier	Accuracy	Sensitivity	Specificity	MCC
<b>XGBoost</b>	<b>0.73 ± 0.14</b>	<b>0.75 ± 0.22</b>	0.75 ± 0.20	<b>0.48 ± 0.30</b>
<b>k-NN</b>	0.68 ± 0.13	0.60 ± 0.21	<b>0.77 ± 0.18</b>	0.39 ± 0.27
<b>SVM</b>	0.68 ± 0.15	0.61 ± 0.22	0.75 ± 0.22	0.38 ± 0.20
<b>Logistic Regression</b>	0.66 ± 0.16	0.62 ± 0.23	0.70 ± 0.21	0.33 ± 0.32
<b>Decision Tree</b>	0.66 ± 0.16	0.65 ± 0.22	0.66 ± 0.22	0.32 ± 0.33
<b>Random Forest</b>	0.70 ± 0.16	0.64 ± 0.24	0.76 ± 0.20	0.42 ± 0.33
<b>GLVQ</b>	0.64 ± 0.14	0.55 ± 0.23	0.74 ± 0.20	0.30 ± 0.29

Tab. 4.19: The outcome of distinction COVID-19 (A and B dataset) from HC (for selected 20 features).

Classifier	Accuracy	Sensitivity	Specificity	MCC
<b>XGBoost</b>	<b>0.73 ± 0.14</b>	<b>0.71 ± 0.22</b>	0.75 ± 0.19	<b>0.48 ± 0.29</b>
<b>k-NN</b>	0.72 ± 0.15	0.58 ± 0.24	0.86 ± 0.16	0.47 ± 0.30
<b>SVM</b>	0.67 ± 0.14	0.60 ± 0.22	0.73 ± 0.19	0.35 ± 0.29
<b>Logistic Regression</b>	0.65 ± 0.15	0.59 ± 0.23	0.70 ± 0.20	0.31 ± 0.31
<b>Decision Tree</b>	0.68 ± 0.15	0.64 ± 0.23	0.71 ± 0.18	0.37 ± 0.30
<b>Random Forest</b>	0.70 ± 0.16	0.64 ± 0.24	0.76 ± 0.20	0.42 ± 0.33
<b>GLVQ</b>	0.66 ± 0.12	0.40 ± 0.21	<b>0.91 ± 0.12</b>	0.36 ± 0.26

Tab. 4.20: The outcome of distinction between COVID-19, Influenza vs. HC based on dataset A.

Classifier	Accuracy	Sensitivity	Specificity	MCC
<b>XGBoost</b>	<b>0.63 ± 0.18</b>	0.65 ± 0.26	0.62 ± 0.28	<b>0.28 ± 0.38</b>
<b>k-NN</b>	0.56 ± 0.18	0.56 ± 0.27	0.58 ± 0.26	0.14 ± 0.39
<b>SVM</b>	0.57 ± 0.14	<b>0.90 ± 0.20</b>	0.24 ± 0.22	0.18 ± 0.31
<b>Logistic Regression</b>	0.49 ± 0.18	0.49 ± 0.28	0.48 ± 0.26	-0.03 ± 0.40
<b>Decision Tree</b>	0.57 ± 0.17	0.48 ± 0.25	<b>0.67 ± 0.30</b>	0.16 ± 0.37
<b>Random Forest</b>	0.54 ± 0.17	0.51 ± 0.33	0.57 ± 0.34	0.08 ± 0.36
<b>GLVQ</b>	0.54 ± 0.15	0.71 ± 0.27	0.37 ± 0.32	0.09 ± 0.35

The outcome of the distinction of COVID-19 (A and B dataset) vs. Influenza (for all analysed cases, from A and B dataset) is presented in Table 4.21. The best accuracy (0.67) and MCC (0.36) were observable for XGBoost. The highest sensitivity: 0.66 was obtained for Random Forest and GLVQ, whereas the best specificity: 0.73 was achieved for SVM.

Tab. 4.21: The outcome of distinction COVID-19 (A and B dataset) vs. Influenza (all analysed cases, from A and B dataset)

Classifier	Accuracy	Sensitivity	Specificity	MCC
XGBoost	<b>0.67 ± 0.13</b>	0.63 ± 0.21	0.71 ± 0.19	<b>0.36 ± 0.27</b>
k-NN	0.62 ± 0.14	0.54 ± 0.22	0.70 ± 0.22	0.25 ± 0.30
SVM	0.62 ± 0.14	0.52 ± 0.22	<b>0.73 ± 0.21</b>	0.26 ± 0.29
Logistic Regression	0.61 ± 0.15	0.58 ± 0.22	0.64 ± 0.21	0.24 ± 0.31
Decision Tree	0.61 ± 0.15	0.62 ± 0.25	0.61 ± 0.21	0.24 ± 0.31
Random Forest	0.63 ± 0.14	<b>0.66 ± 0.22</b>	0.61 ± 0.21	0.28 ± 0.30
GLVQ	0.60 ± 0.14	<b>0.66 ± 0.23</b>	0.54 ± 0.27	0.22 ± 0.31

Tab. 4.22: The outcome of distinction of COVID-19 (A and B dataset), Influenza (for all cases, from A and B dataset) vs. HC.

Classifier	Accuracy	Sensitivity	Specificity	MCC
XGBoost	0.68 ± 0.12	0.62 ± 0.18	0.74 ± 0.16	0.37 ± 0.25
k-NN	0.70 ± 0.12	0.65 ± 0.17	0.75 ± 0.16	0.41 ± 0.24
SVM	<b>0.72 ± 0.12</b>	0.61 ± 0.19	0.82 ± 0.13	0.45 ± 0.25
Logistic Regression	0.67 ± 0.12	<b>0.66 ± 0.17</b>	0.68 ± 0.17	0.35 ± 0.25
Decision Tree	0.71 ± 0.11	0.53 ± 0.18	<b>0.90 ± 0.11</b>	<b>0.47 ± 0.21</b>
Random Forest	0.67 ± 0.11	0.57 ± 0.18	0.78 ± 0.18	0.37 ± 0.24
GLVQ	<b>0.72 ± 0.12</b>	0.64 ± 0.18	0.80 ± 0.16	0.45 ± 0.24

The results of the classification of COVID-19 cases (A and B dataset), Influenza (for all three cases, from A and B dataset) vs. HC are presented in Table 4.22. SVM and GLVQ obtained the highest accuracy of 0.72. Logistic Regression had a sensitivity equal to 0.66, and it was the highest, whereas the best specificity (0.90) and MCC (0.47) were observed for the Decision Tree.

Moreover, the multiclass classification was performed. The following three classes were considered combined for A and B dataset: COVID-19 cases and Influenza, and HC. XGBoost gave the highest F1-score and it was equal to 0.57. Accuracy was also the best for XGBoost: 0.58. MCC was registered the highest for XGBoost and GLVQ.

#### 4.3.4 Discussion and Summarisation

The purpose of this experiment was to develop support methodologies for identifying COVID-19, Influenza, and HC cases in different scenarios. Three types of Influenza were considered: Influenza from dataset A and two cases of Influenza from

Tab. 4.23: The outcome of multiclass classification of COVID-19, Influenza and HC for A and B datasets.

Classifier	F1-score	Accuracy	MCC
<b>XGBoost</b>	<b>0.57 ± 0.13</b>	<b>0.58 ± 0.13</b>	<b>0.38 ± 0.20</b>
k-NN	0.56 ± 0.14	0.57 ± 0.13	0.37 ± 0.20
SVM	0.53 ± 0.13	0.55 ± 0.12	0.34 ± 0.19
Logistic Regression	0.52 ± 0.13	0.54 ± 0.13	0.32 ± 0.19
Decision Tree	0.40 ± 0.07	0.51 ± 0.09	0.31 ± 0.15
Random Forest	0.53 ± 0.12	0.55 ± 0.11	0.34 ± 0.17
GLVQ	0.54 ± 0.12	0.57 ± 0.11	<b>0.38 ± 0.18</b>

dataset B - Influenza before and during the main pandemic. The sampling rate of the datasets was not high, but it was necessary to undersample dataset A to conduct the experiment with a mixture of datasets. The important factors in the experiment were the contagiousness and incubation period. The nature of heart rate during the illness was one of the basements of this research. Regarding the feature creation, the parameters used in [61] were applied. Additionally, the changes in personal activity were considered thanks to the ratio between the appropriate features, like the range of heart rate to the range of steps taken. To reduce the type I error, the Mann-Whitney U test was applied with FDR correction. It was checked if there was a statistically significant difference between COVID-19 cases from HC and ill cases (COVID-19 cases and Influenza) from HC. For the first analysis, the most valuable features occurred to be those indicated on personal activity. Furthermore, heart rate-related parameters have statistical significance. The most important metrics were range, std, rsd, and variance together with Shannon entropy. As a result, the deviations from the norm demonstrate the differences between the cases. The statistical analysis of ill cases vs. HC revealed the importance of more than half of the prepared features. Changes in the heart confirmed the differences between cases, as well as variations in personal activity. The Shannon entropy of the steps taken is also important. Six versions of classification, including multiclass classification, were performed in order to test the quality of distinction between the classes. Among the tested classifiers, XGBoost proved to be the most successful. As a result, XGBoost succeeded in the distinction due to the complexity of its algorithm since the data dependencies were most likely more complex. Nevertheless, overfitting could have a small impact on the results. XGBoost achieved the highest accuracy and also for MCC - the metric which is suitable for imbalanced dataset. For two classifications, the accuracy was equal to 0.73 (Tables 4.18 and 4.19), and this metric was equal to 0.72 for one scenario (Table 4.22) thanks to the usage of XGBoost, SVM and

GLVQ. They were differentiations of COVID-19 cases from HC (Tables 4.18 and 4.19) and distinctions of COVID-19 cases, Influenza vs. HC (Table 4.22), respectively. Those models could simply be assisting in screening testing. Furthermore, the classification with the participation of the data from dataset A, i.e., COVID-19 cases, Influenza vs. HC obtained only 0.63 accuracy. Nevertheless, the dataset is highly undersampled in the comparison to the solution from [2] where the support methodology for such combinations of the data and connected to its higher number of features achieved 0.73 balanced accuracy. Moreover, the separation of the COVID-19 sample vs. Influenza was obtained at a lower level of 0.67 considering data from the A and B datasets, than another achieved in [61] based just on dataset B. It is a small probability of correctly identifying each of the three classes for multi-class classification (0.57 F1-score). The case of distinction COVID-19 cases from HC does not outperform the solution from [2]. Here, the accuracy of the classification between COVID-19 vs. HC was equal to 0.73, whereas in [2] 0.78. Nevertheless, the spectrum of features and sampling rate of the time series were much lower. Additionally, dataset B is biased by people with overweight and the population comes from just on country - from the USA. The people could have specific clinical and demographic characteristics likewise epidemiologic features [232]. The comparison of the results from [61] (See Section 4.1) (0.73) with this solution (0.67) showed that the proposed approach in this Section with the usage of two datasets achieved lower accuracy for COVID-19 cases vs. Influenza. It would be beneficial to extend the gathered signals from various measurements in order to enhance the accuracy of the prediction. There could be temperature body signals, barometers, magnetometry, or gyroscope.

To summarize, the support methodology of COVID-19 detection based on two cohorts was proposed. The combined dataset is one of the largest presented in the literature (see Subsection 3.1). The valuable features from the point of view of distinguishing COVID-19 cases from HC cases were identified by statistical analysis. They were derived from the heart rate and the number of steps taken records. The proposed models are one of a kind. Among the six performed classifications, XGBoost was found to be the most powerful algorithm. The distinction of COVID-19 cases from HC from both datasets was possible in 0.73 balanced accuracy, whereas differentiation of ill cases achieved 0.72 balanced accuracy for k-NN and GLVQ.

## 4.4 Conclusion

A few methods using ML and wearables exist for detecting COVID-19 disease. By the same token, there is a research gap to introduce screening tests based on those technologies. The wearables could be especially utilised because they are relatively

inexpensive and approachable for a wider part of society. In this chapter, there were presented several approaches to COVID-19 detection and they are related to stated research objectives **RO1.**, **RO2.**, and **RO3.** Two datasets (A and B) were taken into account from [59] and [1]. The benefit of the conducted research is that all of the presented methods respect the contagiousness and incubation period. Additionally, all of them were destined for early disease recognition. This approach is a big advantage because it could potentially effectively detect the disease at the early stage, and thereby limit the number of infected people. Most of the papers presented in the literature do not consider early detection. Furthermore, the utilised datasets contain the records of COVID-19, two types of Influenza (before and in the middle of the pandemic) from dataset B, and Influenza from dataset A, likewise HC. A few types of Influenza were analysed and this diversity of the dataset makes the methodology unique. The gathered signals were records of HR and the number of steps taken. The foundations of the conducted and presented studies in this chapter, were observed differences in HR between the types of Influenza and COVID-19 likewise the longer duration of the disease. Nevertheless, the authors in [1] made statistical analysis, whereas the solution in this thesis introduced the classification between the healthy and ill groups. The differentiation between the COVID-19 disease and Influenza was not carried out earlier in the research. Additionally, the combination of the datasets allows for obtaining a more demographically diverse and numerous cohort. The main differences between the proposed approaches in each section are the number of the datasets, representation of classes, and sampling frequency of the used datasets.

The most advanced methodology was proposed in section 4.1. The contribution is the identification of the most statistically significant features, they were frequency- and spectral-related. They were: MFCC, FFT, histogram-related, spectral-based, and LPCC. The identified valuable features are the answers to the question **RQ1.2.**

The advantage of this research is the proposal of the classification methodology. The best results based on dataset A were achieved for a 5-day window and the classification of COVID-19 cases and HC was equal to 0.78 accuracy for k-NN. The distinction between the ill cases (COVID-19 cases or Influenza) from HC reached 0.73 accuracy for the k-NN and GLVQ classifiers and 5-day window. While just the anomaly detection was presented in the original paper [59], reaching 63 % of correctly detected COVID-19 cases likewise the specificity was not computed. It is a reply to question **RQ1.1.**



Tab. 4.24: The overview of the achieved outcomes in Sections 4.1, 4.2, and 4.3.

Approach	Dataset	Algorithm	Accuracy Balanced	Sensitivity	Specificity	MCC	F1-score
COVID-19 detection (5-day window)	A	k-NN	0.78	0.77	0.80	0.60	-
Illness detection (5-day window)	A	k-NN	0.73	0.71	0.76	0.49	-
		GLVQ		0.73	0.72	0.47	-
Distinction between Influenza cases during the middle of the pandemic and cases of COVID-19	B	k-NN	0.73	0.58	0.87	0.49	-
	B	GLVQ	0.82	0.96	0.68	0.68	-
Distinction between COVID-19 cases and Influenza cases before to the pandemic	B	GLVQ	0.84	0.92	0.76	0.71	-
	B	k-NN	0.69	-	-	0.54	0.64
COVID-19 detection	A and B	XGBoost	0.73	0.71	0.75	0.48	-
	A	XGBoost	0.63	0.65	0.62	0.28	-
Illness detection	A and B	XGBoost	0.67	0.63	0.71	0.36	-
	A and B	SVM GLVQ	0.72	0.61 0.64	0.82 0.80	0.45	-
Multiclass classification for COVID-19, Influenza and HC	A and B	XGBoost	0.58	-	-	0.38	0.57

Furthermore, four unique classifications were carried out for dataset B, considering also two types of Influenza. It is one-of-a-kind research, not explored earlier in the literature. This study also confirmed the existing differences between COVID-19, and two Influenza cases occurring in various periods. The differentiation between the COVID-19 cases from Influenza during the pandemic reached 0.73 balanced accuracy for k-NN, whereas the distinction between the COVID-19 cases from Influenza before the main pandemic was equal to 0.84 for GLVQ. Those achieved accuracies are the answers to question **RQ2.1**. The classification of two kinds of Influenza gave the distinction on the level of 0.82 balanced accuracy for GLVQ. This accuracy is the response to the question **RQ2.2**. The presented methodology is simpler with lower obtained results, however, it requires less complex data in comparison to the presented solution based on dataset A.

Additionally, the five unique scenarios were conducted based on the mixture of datasets A and B. Those classifications were performed to evaluate if they could surpass the results achieved based on separate dataset A or B. The distinction between COVID-19 cases and HC was possible on the level of 0.73 accuracy for XGBoost. This accuracy is partial answer to the question **RQ3.2**. The classification of ill cases vs. HC reached 0.72 accuracy for SVM and GLVQ. This accuracy is the fragmentary reply to the question **RQ3.2**. It is a similar outcome achieved on dataset A (0.73). Moreover, the contribution is the identification of the statistically important features extracted from a mixture of A and B datasets. They were the parameters of the ratio between the measurements of heart rate to the measurements of the number of steps taken. Those features are the reply to the question **RQ3.3**.

The summary of the obtained results in this chapter is presented in Table 4.24.

There are significant differences in how the methodologies were designed for the detection of COVID-19 in each scenario, based on datasets A, B, and a combination of both datasets. The developed support methodology for the detection of COVID-19 cases hinged on dataset A needed two-time windows. The spacing between the windows was chosen concerning the incubation period. The shift was equal to 7 days. The placement of the interval considered the highest contagiousness of the diseases. Subsequently, the set of features from spectral, frequency, and statistical domains was extracted for each window. The further bunch of features was subtracted from the earlier set of features. Next, the features selection was performed thanks to the mRMR. The 10-fold cross-validation was done with the classification step. The evaluated classifiers were XGBoost, SVM, k-NN, Random Forest, Decision Tree, Logistic Regression, and GLVQ. The aforementioned methodology is the reply to the question **RQ1.3**. The approach to detecting COVID-19 based on dataset B was various from the already mentioned methodology based on dataset A. The main limitation was the sampling rate of the dataset B. Because of that, one 5-day time

window was chosen instead of two. Nonetheless, the character of the disease in the calculation was maintained - the incubation period and contagiousness of the disease. The feature extraction was performed differently. The computed features were std, skew, var, range, min, max, mean, kurtosis, slope, and approximate entropy of the heart rate signals and number of steps taken. The next steps were computed similarly to the previous support methodology, i.e., the features pre-selection was done thanks to the mRMR, and the 10-fold cross-validation was performed with classification steps. Those same classifiers were used as in the previous example. This approach is a response to the question **RQ2.3**. The procedure which was used to create support methodologies based on datasets A and B was analogous to this used in the case of dataset B. However, dataset A had to be undersampled to combine both datasets. It is the answer to question **RQ3.1**. Additionally, the analysed datasets were gathered by the Fitbit devices. Thereby, created solutions are targeted at Fitbit. The device which could be used is Empatica because it collects a wider spectrum of signals. However, it is more expensive and not so widely distributed in society. Moreover, the COVID-19 restrictions could vary between the countries and thereby have an influence on the collected cohort and obtained results. To sum up, the increase in accuracy could be achieved thanks to the extension of the dataset with a higher sampling rate likewise incorporating other modalities such as skin temperature, BVP, RR, HRV, skin conductance, and oxygen saturation. However, the balance would be difficult to achieve between the probably elevated accuracy and the price of the device embedded in extra sensors.



# 5 mHealth Dedicated Solutions for Parkinson's Detection

This chapter is targeting the detection of PD and is dedicated to mHealth and eHealth solutions for AAL. The undoubted advantage of mobile phones is their broad ownership in society. Moreover, the monitoring and detection of Parkinson's motor and non-motor symptoms are becoming approachable for the elders and their families. The additional clear advantage of this is the reduction of the cost of the healthcare system. This chapter presents the automatic analysis of changes in emotions to recognise PD likewise the multimodal detection of PD based on hypomimia and HD symptoms with the usage of audio and video records. Furthermore, the chapter discusses the used material and methods, the collected dataset, and the feature extraction methodology. Moreover, the used ML approaches with the solutions for the interpretability of the model, the used metrics, and the discussion of the key findings are provided. Those kinds of techniques allowed the creation of support system methodologies for the detection of PD together with their interpretability.

## 5.1 Parkinson's Disease Detection based on Changes of Emotions during Speech

This research aims to develop a methodology which is detecting PD. Symptoms of hypomimia manifested in the difficulty of expressing emotions were the basis of the study. Using a numerical analysis of the changes in emotions over time, the set of features was determined. *Face expression recognition* (FEC) based on neural networks was used to detect differences frame by frame. For the evaluation of the disease, a tongue twister and reading aloud long texts were tested. Fig. 5.1 presents a schematic representation of the experiment. The first step was to calculate the intensity of the emotion in each frame using a neural network. Next, scalars for each emotion were calculated based on the time series. Using the mRMR, the feature pre-selection process was then performed. A stratified cross-validation with standardization and classification was conducted. k-NN, SVM, Random Forest, Decision Tree, Logistic Regression, and XGBoost were used as classifiers.

A Mann-Whitney U test was used with FDR correction to eliminate the influence of type I errors in the statistical evaluation. SHAP values were used to calculate the interpretability of the classifier.

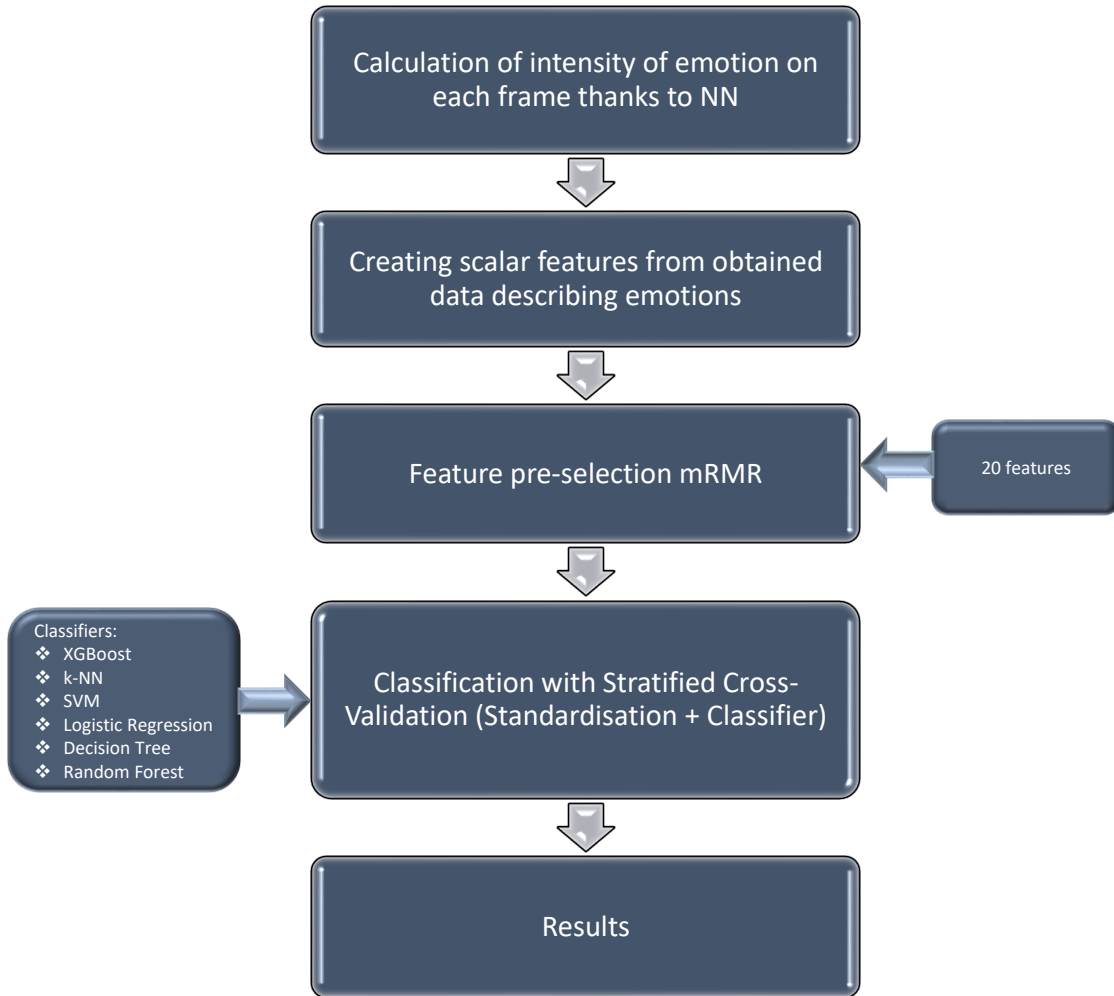


Fig. 5.1: The scheme of the experiment.

### 5.1.1 Dataset Characterization

For the purpose of this research, 45 HC (21 females (mean age  $62 \pm 9.22$ ), 24 males (mean age  $66 \pm 9.17$ )) and 70 PD patients (27 females (mean age  $68 \pm 8.04$ ), 43 males (mean age  $66 \pm 7.83$ )) were involved.

It was found that the *Unified Parkinson's Disease Rating Scale* (UPDRS) III mean for female Parkinsonians was  $21.6 \pm 13.50$  and the mean duration of PD in years was  $7.2 \pm 4.82$ , but the mean UPDRS III for male Parkinsonians was  $26.3 \pm 11.31$  and the mean duration of PD in years was  $7.9 \pm 4.64$ . HC and individuals with Parkinson's disease were recorded while being asked to pronounce and read long texts.

During this experiment, a Czech sentence (Celý večer se učí sčítat)<sup>1</sup> was pro-

<sup>1</sup>Link for the pronunciation of the sentence: <https://bit.ly/2DVPJ5M>

nounced by the participants. The sentence means "He's been learning to count all night", however, the difficulty in pronouncing sentence matters in the case of this experiment.

Ethical approval was granted by the Masaryk University Ethical Committee. The data were gathered by neurologists.

The total number of video records for speech exercise is 70 PD plus 45 HC, which equals 115. Records vary in length. During the acquisition of the video, 25 *frame per second* (FPS) sampling was used.

### 5.1.2 Feature extraction and Machine Learning

The detection of PD was conducted in a few computational phases. The primary objective of this study was to assign the numerical values to emotions and to classify illness based on the differences in intensity of emotional changes during the pronunciation of the task. A FEC method was applied [233] to achieve this aim. FER 2013 dataset [234] was used to train this publicly available neural network. Initially, the face was detected by using *Multitask Cascaded Convolutional Networks* (MTCNN) [235]. The neural network architecture was used in the second step to determine the intensity of seven analysed emotions. Aside from surprise and neutral emotions, disgust, sadness, happiness, fear, and anger are considered. Regression intensities were calculated for each frame, and these were then aggregated for the seven time series for each participant. Fig. 5.2 illustrates the instances of recognition emotion per frame. The time series for each participant is shown in Fig. 5.3. As a next step, the scalar values were computed from seven time series for each participant. The following features were determined: approximate entropy, Shannon entropy, skewness, kurtosis, std, rsd, range, max, and min.

Four metrics were used to evaluate the models: sensitivity, specificity, balanced accuracy and MCC. The used classifiers include XGBoost, k-NN, SVM, Decision Tree, Random Forest, and Logistic Regression. XGBoost emerged as the most promising. This classifier is among them the most powerful algorithms for structured data due to its ability to optimize specific loss functions and regularization techniques. Due to the use of parallelization techniques, the algorithm is also executed more rapidly.

### 5.1.3 Results

As can be seen in Table 5.1, a statistical analysis of the features has been conducted. This table presents the top 10 features based on their *p-values* (pVals). Furthermore, there is also provided the pVal after FDR correction, as well as the median and IQR of PD and HC. P-values and P-values with FDR correction are all

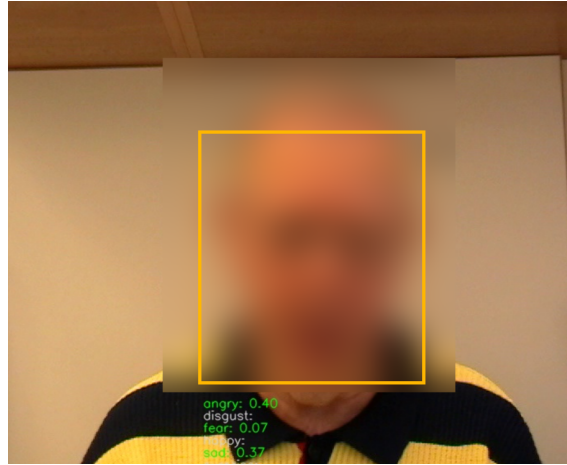


Fig. 5.2: Emotional recognition based on the face image. (The image has been blurred for privacy reasons).

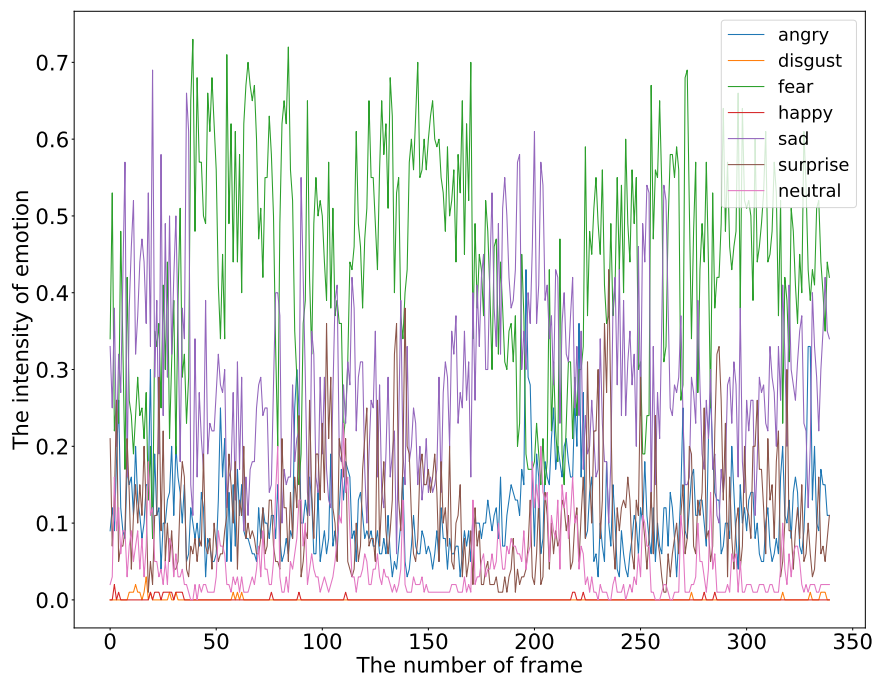


Fig. 5.3: An analysis of the changes in emotions during a speech exercise.

below the set significance level  $\alpha = 0.05$ . The following features are notably important: fear\_std, fear\_variance, angry\_mean, angry\_std, angry\_variance, fear\_max, fear\_mean, angry\_min, fear\_range, sad\_approx\_entropy.

The outcomes of the tongue twister classification are presented in Table 5.2.



Tab. 5.1: Statistical analysis of the created features among PD and HC.

Features	pVal	pVal(FDR)	Median (PD)	Median (HC)	IQR (PD)	IQR (HC)
fear_std	0.0034	0.0408	0.0564	0.0350	0.0649	0.0390
fear_variance	0.0034	0.0408	0.0032	0.0012	0.0078	0.0032
angry_mean	0.0038	0.0408	0.1762	0.1276	0.1953	0.1639
angry_std	0.0039	0.0408	0.0819	0.0562	0.0605	0.0499
angry_variance	0.0039	0.0408	0.0067	0.0032	0.0097	0.0060
fear_max	0.0040	0.0408	0.2950	0.2100	0.3050	0.1800
fear_mean	0.0045	0.0408	0.0988	0.0725	0.1711	0.0719
angry_min	0.0050	0.0408	0.0600	0.0400	0.0700	0.0600
fear_range	0.0050	0.0408	0.2500	0.1900	0.2700	0.1500
sad_approx_entropy	0.0053	0.0408	0.3624	0.4990	0.3005	0.2781

Tab. 5.2: The prediction for tongue twister based on the various classifiers.

Classifier	ACC Bal	Sensitivity	Specificity	MCC
k-NN	0.63 ± 0.13	0.69 ± 0.17	0.57 ± 0.23	0.27 ± 0.27
SVM	0.59 ± 0.13	0.70 ± 0.16	0.49 ± 0.25	0.19 ± 0.27
Decision Tree	0.57 ± 0.13	0.41 ± 0.17	<b>0.73 ± 0.24</b>	0.15 ± 0.27
Random Forest	0.62 ± 0.14	0.68 ± 0.18	0.56 ± 0.23	0.25 ± 0.29
Logistic Regression	0.59 ± 0.13	<b>0.73 ± 0.16</b>	0.45 ± 0.24	0.18 ± 0.28
XGBoost	<b>0.69 ± 0.14</b>	0.71 ± 0.17	0.67 ± 0.22	<b>0.39 ± 0.29</b>

Based on balanced accuracy, XGBoost was rated as the best classifier with a score of 0.69. MCC for this classifier was 0.39. Logistic Regression had the highest sensitivity of 0.73, while the best specificity was observed for Decision Tree.

The SHAP values provided for the XGBoost classifier the interpretability of the model (see Fig. 5.4). The positive correlation of the features to PD was recorded for std, variance, range, maximum, and mean for fear (fear\_std, fear\_variance, fear\_range, fear\_max, fear\_mean) and maximum and variance for anger (angry\_max, angry\_variance), as well as the mean for surprise (surprise\_mean). PD was negatively correlated with approximate entropy for surprise and sadness (surprise\_approx\_entropy, sad\_approx\_entropy).

In addition, the results of the classification of PD for two different speech exercises for XGBoost are presented in Table 5.3. Not only the tongue twister was evaluated but also the participants were asked to read long text. Prediction results for the second speech exercise are 0.60 balanced accuracy, 0.66 sensitivity, 0.54 specificity, and 0.20 MCC.

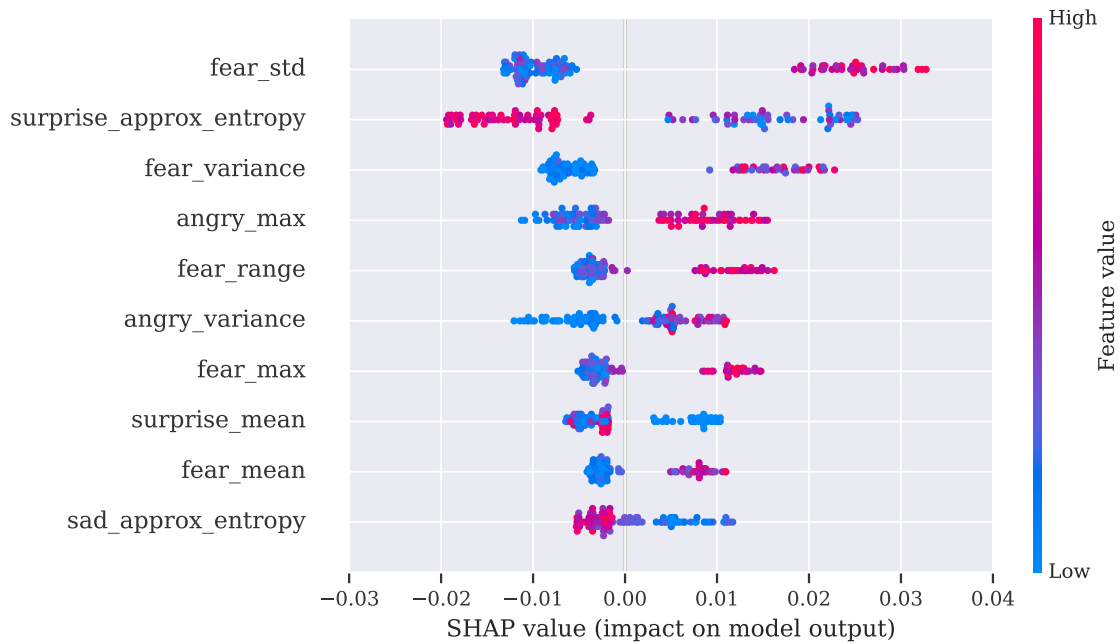


Fig. 5.4: The values of SHAP were derived from the XGBoost tongue twister model.

Tab. 5.3: XGBoost predictions for tongue twister and reading text exercise.

Speech exercise	ACC Bal	Sensitivity	Specificity	MCC
<b>Tongue twister</b>	$0.69 \pm 0.14$	$0.71 \pm 0.17$	$0.67 \pm 0.22$	$0.39 \pm 0.29$
<b>Long text</b>	$0.60 \pm 0.16$	$0.66 \pm 0.21$	$0.54 \pm 0.27$	$0.20 \pm 0.34$

### 5.1.4 Discussion and Summarisation

A methodology for the automatic detection of PD based on changes in emotions is presented. In this study, hard-to-pronounce Czech tongue twister was examined as well as the reading of a long text by participants. Based on the detection of seven emotions in each video frame, it occurs that fear is the most meaningful emotion. The results indicated that the most informative variables were mean, std, variance, maximum, mean and range of fear, mean, std, variance, min of anger, and approximate entropy of sadness. SHAP values confirmed the importance of fear emotion (std, var, range, max, and mean) for the classification. This may be explained by the fear of difficulty in correctly pronouncing the speech exercise by PD patients and laboratory conditions. The changes in entropy for surprise and sadness

were negatively correlated with PD, which could be explained by impairment to express emotions generally by PD patients. A tongue twister speech exercise proved to be more predictive and robust for detecting PD than reading a lengthy text. For this task, the XGBoost achieved a balanced accuracy of 0.69.

As a conclusion, this study was designed to provide support methods for the detection of PD in order to assist clinicians in their diagnosis of this disease. This research is exploring the potential of rarely analysed - hypomimia symptom for PD detection. Moreover, the generation of scalars from the speech exercise record appears to be appropriate for the state problem. The contribution of this study is the identification of fear as the most statistically significant emotion based on SHAP values for the XGBoost model. XGBoost delivered the 0.69 balanced accuracy for a tongue twister, indicating that it is the most important speech task that has been evaluated. Thereby, the tongue twister appeared to have clinical value. The future direction of this research is to extend the database and test other tongue twisters for the detection of PD.

## 5.2 Multimodal Detection of Parkinson Disease

The PD methodology detection based on a multimodality approach (combination of video and audio) was set up as the goal. The 43 characteristic speech exercises were used to evaluate the PD thanks to the video and audio analysis. The feature extraction was designed to capture the hypomimia symptoms and changes in HD dimension between HC and illness cases. Especially, the facial features were created based on facial landmarks, which are valuable anthropometrically. The prediction of the PD was conducted according to the scheme presented in Fig. 5.5. In the first step, a previously introduced dataset in Subsection 5.1.1 was prepared for extracting features for each speech exercise. The exact feature extraction for video and audio modality was then provided separately. The regression out of the confounding factors was performed. Statistical analysis was then carried out using the Mann-Whitney U test and FDR correction. Next, the feature preselection method mRMR was also applied. The Stratified 10-fold Cross-Validation with XGBoost was used. The SHAP values were applied for the clinical interpretation of the model.

### 5.2.1 Dataset Characterization

A total of 73 people were included in the analysis, including 43 males (education length  $14.76 \pm 2.97$ , mean age  $66 \pm 7.83$ ) and 30 females (education length  $13.04 \pm 2.70$ , mean age  $68 \pm 8.20$ ) and 46 people with HC (24 males (mean age  $66 \pm 9.17$ ), 22 females (mean age  $62 \pm 9.02$ )). The dataset was collected under the

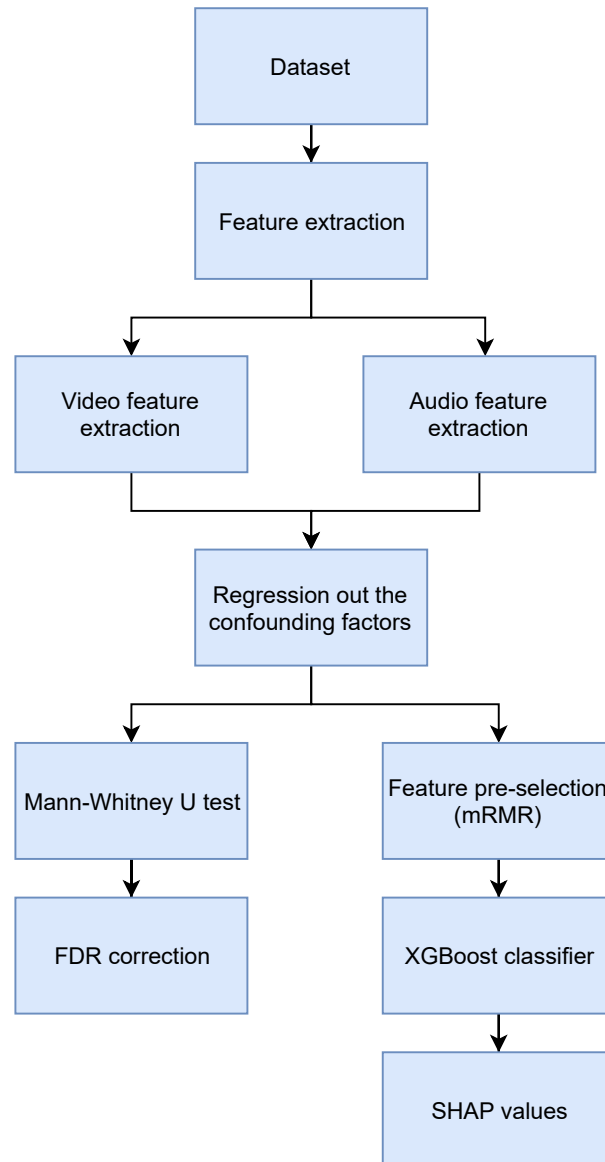


Fig. 5.5: Flow of the algorithm.

same conditions as those described in subsection 5.1.1. The clinical condition and demographic information can be found in Table 5.4. Additionally, the subsection provides the UPDRS III (Fig. 5.6) as well as the kernel density estimation of the duration of PD (Fig. 5.7).

Individuals participated in a variety of speech exercises during the experiment. The Czech language was considered, but pronunciation difficulties rather than the meaning of the expression played a role. A variety of exercises were examined, including tongue twisters, poems, free speech, diadochokinesis tasks, reading texts, sentences, words, vowels, and others. The used devices were the camera and micro-

Tab. 5.4: Clinical and demographical information of the dataset.

	Mean	Std	Min	Q1	Median	Q3	Max
<b>Age</b>	66.90	7.95	49	62.00	67.0	72.00	82
<b>Duration of PD</b>	7.80	4.39	1	4.00	7.0	11.00	22
<b>UPDRS III</b>	24.91	11.91	3	14.75	25.5	33.00	55
<b>UPDRS IV</b>	3.16	2.73	0	0.00	3.0	5.00	10
<b>FOG</b>	7.16	5.79	0	2.00	7.0	11.00	20
<b>NMSS</b>	38.37	23.06	2	19.00	34.5	54.00	112
<b>RBDSQ</b>	3.79	3.21	0	1.00	3.0	6.00	13
<b>LED [mg]</b>	1006.04	542.94	0	621.25	879.5	1325.50	2275
<b>ACE-R</b>	87.15	8.01	60	82.75	87.5	93.00	100
<b>MMSE</b>	28.04	2.38	16	28.00	29.0	29.00	30
<b>BDI</b>	10.41	6.06	0	6.00	9.0	13.50	27
<b>DX</b>	74.32	8.90	30	71.00	76.0	79.00	88

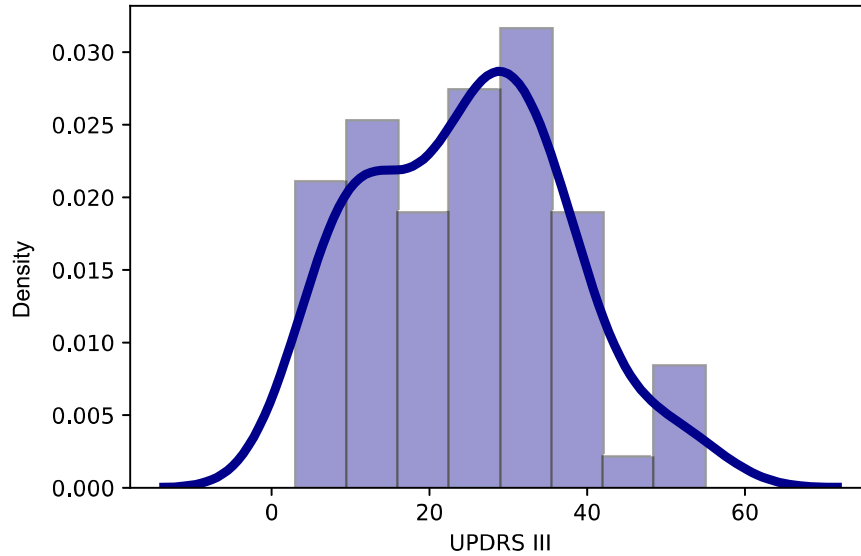


Fig. 5.6: Kernel Density Estimation of level of UPDRS III.

phone. For this study, a PANASONIC SDR-H20 camera was used with a sampling frequency of 25 FPS. At a distance of 20 cm from the mouth, the cardioid microphone M-AUDIO Nova was mounted on the arm, with a sampling frequency of 48 kHz, and a 16-bit resolution. The UPDRS - diagnostic assessment was also used to assess the severity of PD illness. The data were collected at the hospital in Czechia. The Masaryk University Ethical Committee approved the study.

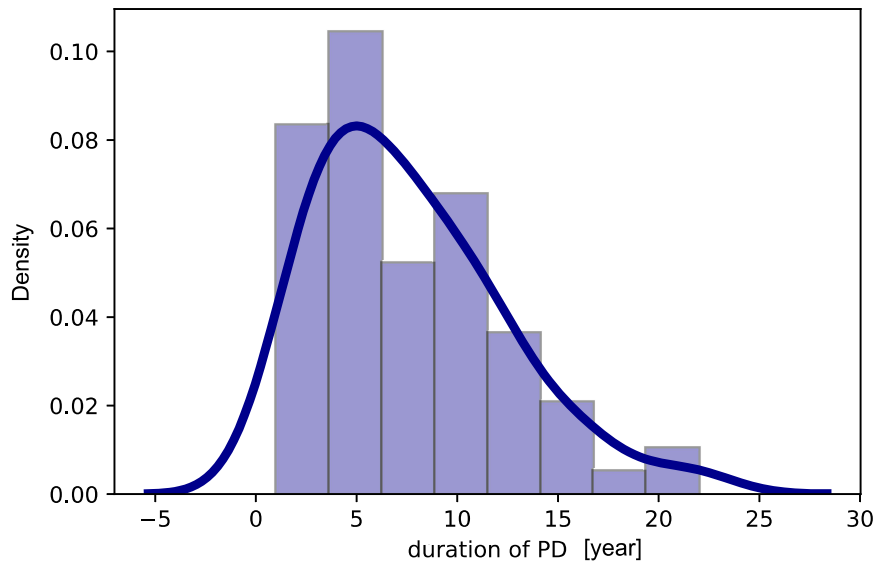


Fig. 5.7: Kernel Density Estimation of duration of PD.

## 5.2.2 Feature Extraction

There are two steps involved in the feature extraction process. A description of each of the video and audio features extraction is located in this section.

### Facial features

A feature extraction algorithm was developed. This proposed methodology contains the extraction of facial landmarks and the computations of the differences between them within a specified period. The varieties between distances and angles in time were calculated from the facial measurements. 68 facial landmarks were detected using an open source framework <sup>2</sup>. Fig. 5.9 illustrates the schematic illustration. The algorithm involves two steps: first, it detects faces using the HOG and Haar feature-based cascade classifiers. Detection of facial landmarks was performed using a neural network in the second step. The founded positions of x and y were used to generate a time series. Lastly, the scalars were calculated to determine how the characteristic points on the face differentiated over time. The measurements were the *Shannon entropy* (se), the *approximate entropy* (ae), the max, the min, the std, the rsd, the var, the range, the slope, and the mean. The explanation of the created scalars is presented in Table 5.5. Fig. 5.8 illustrates the flow of the algorithm.

<sup>2</sup><https://pypi.org/project/face-recognition/>

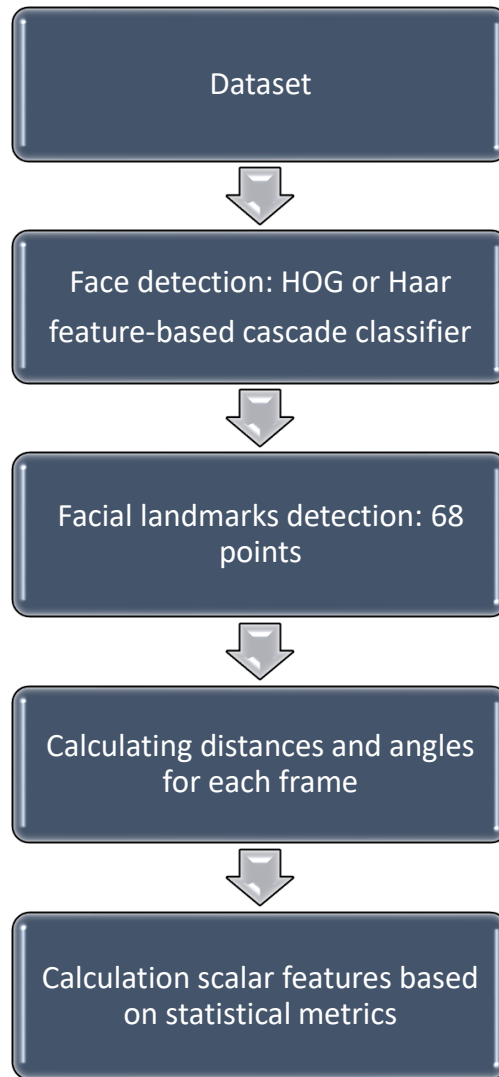


Fig. 5.8: Flow of the facial features extraction.

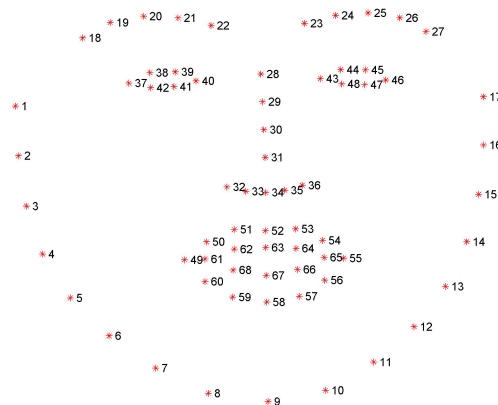


Fig. 5.9: Illustration of facial features [12].

Tab. 5.5: Features extraction explanation.

Name feature	Points, angle
D1	37, 49
D2	46, 55
D3	22, 23
D4	52, 58
D5	20, 38
D6	25, 45
D7	39, 41
D8	45, 47
D9	31, 9
D10	1, 17
D11	18, 22
D12	23, 27
D13	34, 52
EYEBROW1	Angle: (22,19) vs. (40, 43)
EYEBROW2	Angle: (22, 19) vs. (23, 26)
EYEBROW3	Angle: (22, 19) vs. (23, 26)
EYEBROW4	Vertical: 19, 37
EYEBROW5	Vertical: 26, 46
EYE1	37, 38
EYE2	37, 39
EYE3	46, 45
EYE4	46, 44
EYE5	40, 39
EYE6	40, 38
EYE7	43, 44
EYE8	43, 45
EYE9	37, 42
EYE10	37, 41
EYE11	43, 48
EYE12	43, 47
EYE13	40, 41
EYE14	40, 42
EYE15	46, 48
EYE16	46, 47
EYE17	38, 42
EYE18	45, 47
EYE19	39, 41
EYE20	44, 48
EYE21	37, 40
EYE22	43, 46
M1	49, 52
M2	49, 58
M3	55, 52
M4	55, 58
M5	49, 55
M6	52, 58
M7	60, 54
M8	50, 56
RATIO_MOUTH	M5/M6
MOUTH_AREA	The area of the inside of the mouth
LEYE_AREA	The area of the left eye
REYE_AREA	The area of the right eye
RATIO_FACE	D1/D2
RATIO_MOUTH_SKEW_UP	M3/M1
RATIO_MOUTH_SKEW_DOWN	M4/M2



Tab. 5.6: Description of the acoustic features. The details of the features implementation are provided in [14].

Code of Acoustic feature	Description of the features	HD dimension	Specific disorder
<b>DDK rate</b>	DDK rate	articulation	slow alternating motion rate
<b>DDK reg</b>	std of DDK cycle periods	articulation	irregular alternating motion rate
<b>DUV</b>	fraction of locally unvoiced frames	phonation	aperiodicity
<b>MPT</b>	total speech time	phonation	airflow insufficiency
<b>NSR</b>	net speech rate	prosody	unnatural speech rate
<b>SPIR</b>	speech index of rhythmicity	prosody	inappropriate silences
<b>jitter</b>	period perturbation quotient	phonation	microperturbations in frequency
<b>mean HNR</b>	mean of harmonic-to-noise ratio	phonation	increased noise
<b>relF0SD</b>	relative std of fundamental frequency	prosody	monopitch
<b>relF1SD</b>	relative std of 1st formant	articulation	rigidity of tongue and jaw
<b>relF2SD</b>	relative std of 2nd formant	articulation	rigidity of tongue and jaw
<b>relSEOSD</b>	relative std of short-time energy	prosody	monoloudness
<b>shimmer</b>	amplitude perturbation quotient	phonation	microperturbations in amplitude

Tab. 5.7: Meaning of the part of the exercises in Czech and English.

Code	In Czech	English translation
<b>TSK19</b>	Chcete vidět velký lov? Budu lovit v džungli slov. Osedlám si Pegasa Chytím báseň do lasa.	Would you like to see a big hunt? I will be hunting in a jungle of words. I will saddle the Pegasus, I will catch a poem into a lasso.
<b>TSK20</b>	Prostřete k obědu?	Will you lay the table?
<b>TSK21</b>	Prostřete k obědu!	Lay the table!
<b>TSK22</b>	Prostřete k obědu.	Lay the table.
<b>TSK23</b>	Teď musíš být chvíli trpělivý, než to dokončíme.	Now you have to be patient for a while until we finish.
<b>TSK24</b>	Tak dáš mi už konečně pokoj!	I urge you to leave me alone.
<b>TSK25</b>	Už mě to nebaví, dej mi už konečně pokoj!	I am fed up, I urge you to leave me alone.
<b>TSK26</b>	Tak co, jak to dopadlo?	So, what happened?
<b>TSK27</b>	rychlonožka	lightfoot
<b>TSK28</b>	marnotratný	wasteful
<b>TSK29</b>	horolezectví	mountaineering
<b>TSK30</b>	stříbrotepec	silversmith
<b>TSK31</b>	železobetonový	iron-concrete
<b>TSK32</b>	zákonodárce	legislator
<b>TSK33</b>	horkovzdušný	convection
<b>TSK34</b>	strastiplná	tortuous
<b>TSK35</b>	záviděníhodný	enviable
<b>TSK36</b>	československý	Czechoslovak
<b>TSK37</b>	Do čtvrt hodiny tam byla smršť.	In a quarter of an hour there was a whirlwind.
<b>TSK38</b>	Prohovořte to s ním dopodrobna.	Discuss it with him in detail.
<b>TSK39</b>	Při ústupu pluku duní bubny.	Drums are pounding during the retreat of regiment.
<b>TSK40</b>	Kuchařští učni nejsou jak zlatníci.	Apprentices of cookery school are not as those from goldsmith one.
<b>TSK41</b>	Celý večer se učí seřítat.	He is learning to add the whole evening.

Tab. 5.8: Definition of the acoustic features in detailed.

HD dimension and specific disorder	Vocal tasks	Acoustic feature		Feature definition
<b>Phonation</b>				
Airflow insufficiency	Expiration with closed (TSK2) or opened (TSK3) lips	MPT	Maximum phonation time.	
Irregular pitch fluctuations	Sustained phonation (TSK3 - TSK17)	relFISD	The standardized efficiency of the vocal tract measured as the maximum duration of the sustained vowel/consonant.	The standard deviation of fundamental frequency relative to its mean, variation in frequency of vocal fold vibration.
Microperturbations in frequency	Sustained phonation (TSK3 - TSK17)	jitter	Frequency perturbation, the extent of variation of the voice range.	Jitter is defined as the variability of the F0 of speech from one cycle to the next. In this case it is implemented as the five-point period perturbation quotient.
Microperturbations in amplitude	Sustained phonation (TSK3 - TSK17)	shimmer	Amplitude perturbation, representing rough speech.	Shimmer is defined as the sequence of maximum extent of the signal amplitude within each vocal cycle. In this case implemented as the five-point amplitude perturbation quotient.
Tremor of jaw	Sustained phonation (TSK3-TSK17)	relFISD, relP2SD	The standard deviation of the first (F1) and second (F2) formant relative to their mean.	Formants are related to the resonances of the oro-naso-pharyngeal tract and are modified by position of tongue and jaw.
Increased noise	Sustained phonation (TSK3 - TSK17)	mean HNR	Harmonics-to-noise ratio, the amount of noise in the speech signal, mainly due to incomplete vocal fold closure.	HNR is defined as the amplitude of noise relative to tonal components in speech.
Aperiodicity	Sustained phonation (TSK3-TSK17)	DUV	Degree of inverted segments, the fraction of pitch frames marked as unvoiced.	
<b>Articulation</b>				
Rigidity of tongue and jaw	Rhythmic units (TSK10), Basic intonation template (TSK20-TSK22), Reading with different emotions (TSK23-TSK26), Repeated word complicated for articulation (TSK27-TSK36), Repeated sentence complicated for articulation (TSK37-TSK41), Reading paragraph (TSK42), Monologue (TSK43)	relFISD, relP2SD	The standard deviation of the first (F1) and second (F2) formant relative to their mean.	Formants are related to the resonances of the oro-naso-pharyngeal tract and are modified by position of tongue and jaw.
Slow alternating motion rate	Diadochokinetic task (TSK48)	DDK rate	Diadochokinetic rate, representing the number of syllable vocalizations per second.	
Irregular alternating motion rate	Diadochokinetic task (TSK48)	DDK reg	Diadochokinetic regularity, defined as the standard deviation of distances between following syllables nuclei.	
<b>Prosody</b>				
Mondolousness	Rhythmic units (TSK10), Basic intonation template (TSK20-TSK22), Reading with different emotions (TSK23-TSK26), Repeated word complicated for articulation (TSK27-TSK36), Repeated sentence complicated for articulation (TSK37-TSK41), Reading paragraph (TSK42), Monologue (TSK43)	relSEOSD	Speech loudness variation, defined as the standard deviation of intensity contour relative to its mean.	
Monopitch	Rhythmic units (TSK10), Basic intonation template (TSK20-TSK22), Reading with different emotions (TSK23-TSK26), Repeated word complicated for articulation (TSK27-TSK36), Repeated sentence complicated for articulation (TSK37-TSK41), Reading paragraph (TSK42), Monologue (TSK43)	relFISD	F0 pitch variation, defined as the standard deviation of F0 contour relative to its mean.	
Inappropriate silences	Basic intonation template (TSK20-TSK22), Reading with different emotions (TSK23-TSK26), Repeated sentence complicated for articulation (TSK37-TSK41), Reading paragraph (TSK42)	SPR	Number of speech inter-pauses per minute.	
Unnatural speech rate	Basic intonation template (TSK20-TSK22), Reading with different emotions (TSK23-TSK26), Repeated word complicated for articulation (TSK27-TSK36), Repeated sentence complicated for articulation (TSK37-TSK41)	NSR	If we consider net speech time (NST) as a duration of speech without pauses, then the net speech rate (NSR) is defined as the number of phones per NST.	

## Voice features

The features were extracted with respect to [84]. Parameters related to personal impairments, phonation, articulation, and prosody were calculated. Details can be found in Tables 5.6 and 5.8.

Tab. 5.9: Carried-out vocal tasks.

Code	Vocal task	Description
TSK1	expiration	maximum phonation of [m] in one breath
TSK2	expiration	maximum phonation of [i] in one breath
TSK3	phonation	vowel [a] (sustained and normal intensity)
TSK4	phonation	vowel [e] (sustained and normal intensity)
TSK5	phonation	vowel [i] (sustained and normal intensity)
TSK6	phonation	vowel [o] (sustained and normal intensity)
TSK7	phonation	vowel [u] (sustained and normal intensity)
TSK8	phonation	vowel [a] (sustained and maximum intensity)
TSK9	phonation	vowel [e] (sustained and maximum intensity)
TSK10	phonation	vowel [i] (sustained and maximum intensity)
TSK11	phonation	vowel [o] (sustained and maximum intensity)
TSK12	phonation	vowel [u] (sustained and maximum intensity)
TSK13	phonation	vowel [a] (sustained and minimum intensity, but not whispering)
TSK14	phonation	vowel [e] (sustained and minimum intensity, but not whispering)
TSK15	phonation	vowel [i] (sustained and minimum intensity, but not whispering)
TSK16	phonation	vowel [o] (sustained and minimum intensity, but not whispering)
TSK17	phonation	vowel [u] (sustained and minimum intensity, but not whispering)
TSK18	diadochokinesis (DDK)	DDK pa-ta-ka
TSK19	rhythmical units	read poem
TSK20	main intonation pattern	same sentence read as interrogative
TSK21	main intonation pattern	same sentence read as imperative
TSK22	main intonation pattern	same sentence read as declarative
TSK23	intonation variability	monitoring prosody (declarative read sentence)
TSK24	intonation variability	monitoring prosody (imperative read sentence)
TSK25	intonation variability	monitoring prosody (imperative read sentence)
TSK26	intonation variability	monitoring prosody (interrogative read sentence)
TSK27	intelligibility of repeated words	repeated word complicated for the articulation
TSK28	intelligibility of repeated words	repeated word complicated for the articulation
TSK29	intelligibility of repeated words	repeated word complicated for the articulation
TSK30	intelligibility of repeated words	repeated word complicated for the articulation
TSK31	intelligibility of repeated words	repeated word complicated for the articulation
TSK32	intelligibility of repeated words	repeated word complicated for the articulation
TSK33	intelligibility of repeated words	repeated word complicated for the articulation
TSK34	intelligibility of repeated words	repeated word complicated for the articulation
TSK35	intelligibility of repeated words	repeated word complicated for the articulation
TSK36	intelligibility of repeated words	repeated word complicated for the articulation
TSK37	intelligibility of repeated sentences	repeated sentence complicated for articulation
TSK38	intelligibility of repeated sentences	repeated sentence complicated for articulation
TSK39	intelligibility of repeated sentences	repeated sentence complicated for articulation
TSK40	intelligibility of repeated sentences	repeated sentence complicated for articulation
TSK41	intelligibility of repeated sentences	repeated sentence complicated for articulation
TSK42	monitoring intelligibility and articulation	long read paragraph
TSK43	interview at the beginning - monitoring prosody, hesitations, time needed for response, etc.	free speech, usually the answer to "What are your hobbies?", "Where do you come from?", etc.

### 5.2.3 Statistical Evaluation with Machine Learning

To eliminate the influence of age and gender on the data, the regression out step was first performed. Videos and audio recordings were exempt from the confounding effect. The impact of confounding variables (gender and age) on independent variables (extracted features) and dependent variables (the existence of PD) was eliminated thanks to deleting the confounding effect. The confound was considered as a predictor, whereas the linear regression model was fitted on all features. The descriptions of the approach are illustrated in [236, 237]. Next, the statistical analysis was conducted using the Mann-Whitney U test [238] to detect the dependency of a separate feature on PD. The FDR correction was applied to limit type I errors. Subsequently, The set of features was chosen with the usage of mRMR to avoid the curse of dimensionality. Next, the Stratified 10-fold cross-validation was performed. The stratified cross-validation contains two steps: the stratified sampling when to equal distribution of the samples of each group is guaranteed in the training and test dataset, and the standardization of the data [239]. XGBoost was used as the classifier. This classifier has a few benefits, for instance, the robustness for the imbalanced dataset, a solid and effective approach for tabularised data, and the ability to find the non-linear correlation in the data. Moreover, this algorithm used a type of end-to-end tree ensembling model [240]. The SHAP values were used to understand the decision standing beyond the models about the PD detection. They can indicate the correlation of features to the decision model [241, 242]. Model quality was evaluated using the following metrics: accuracy, sensitivity, specificity, balanced accuracy, and MCC.

### 5.2.4 Results

Results are provided separately for statistical analysis and results of classification. Additionally, SHAP values are presented for the best models.

Tab. 5.10: The statistical analysis of audio features.

Features	pVal	pVal(FDR)	Median (PD)	Median (HC)	IQR (PD)	IQR (HC)
relF0SD (TSK7)	2.7E-0.5	0.0057	-0.0408	0.0033	0.0863	0.102
shimmer (TSK15)	4.6E-0.5	0.0057	-4.2218	3.374	12.4081	11.1142
DUV (TSK7)	7.6E-0.5	0.0062	-4.6691	-1.7888	4.7326	8.6902
relF0SD (TSK24)	0.000128	0.0078	-0.0389	0.0108	0.1234	0.0828
shimmer (TSK17)	0.00035	0.0172	-3.7772	3.4596	13.5591	11.1147
shimmer (TSK13)	0.000581	0.0237	-2.7854	2.5983	11.9312	9.9691
NSR (TSK25)	0.001787	0.0487	-0.0402	-1.3382	3.7665	2.9786
DUV (TSK8)	0.001657	0.0487	-4.6784	0.8537	16.464	13.3974
shimmer (TSK16)	0.001754	0.0487	-2.9298	1.5677	12.2635	16.6985
NSR (TSK41)	0.002379	0.0571	0.3378	-0.9651	3.1651	2.2758

Tab. 5.11: The statistical analysis of video features.

Features	pVal	pVal(FDR)	Median (PD)	Median (HC)	IQR (PD)	IQR (HC)
rsdD8 (TSK31)	0.000015	0.0733	-2.0961	1.2194	7.9407	7.0653
rsdEYE18 (TSK31)	0.000015	0.0733	-2.0961	1.2194	7.9407	7.0653
slopeM7 (TSK41)	0.000021	0.0733	0.0005	-0.0008	0.0024	0.0026
rsdD8 (TSK32)	0.000025	0.0733	-2.1191	1.4171	6.2564	8.0297
rsdEYE18 (TSK32)	0.000025	0.0733	-2.1191	1.4171	6.2564	8.0297
stdD6 (TSK32)	0.000028	0.0733	-0.0111	0.0083	0.0239	0.0337
aeEYE16 (TSK37)	0.000032	0.0733	-0.0446	0.0449	0.1445	0.1318
varD6 (TSK32)	0.000034	0.0733	-0.0015	0.0001	0.0014	0.0032
varM2 (TSK12)	0.000035	0.0733	-0.0013	-0.003	0.0010	0.0015
meanM5 (TSK18)	0.000044	0.0733	0.0416	0.1375	0.1433	0.0920

A regressed out dataset was taken under analysis by the Mann-Whitney U test with an FDR correction. The statistical analysis was conducted on 13 audio features and 550 video features for each speech exercise. The results of statistical analysis of the video and audio features are presented in Tables 5.10 and 5.11. In addition, the median and IQR for each group are provided.

Nine out of ten audio features were statistically significant ( $\alpha = 0.05$ ) when corrected with FDR. They were: relF0SD (TSK7), shimmer (TSK15), DUV (TSK7), relF0SD (TSK24), shimmer (TSK17), shimmer (TSK13), NSR (TSK25), DUV (TSK8), shimmer (TSK16). In Tables 5.6 and 5.9, the features are described in greater detail.

According to the Mann-Whitney U test, none of 10 the best-selected vide features passed the test with FDR correction. Despite this, the test was passed for them without FDR correction. They were: rsdD8 (TSK31), rsdEYE18 (TSK31), slopeM7 (TSK41), rsdD8 (TSK32), rsdEYE18 (TSK32), stdD6 (TSK32), aeEYE16 (TSK37), varD6 (TSK 32), varM2 (TSK12), meanM5 (TSK18). These details can be found in Tables 5.5, 5.9. Following this step, a ML algorithm was used to distinguish PD cases from HC cases. The results are presented for XGBoost. 10-fold stratified cross-validation was performed.

Tab. 5.12: Accuracy of PD detection from different modalities.

Modality	Accuracy (balanced)	Sensitivity	Specificity	MCC
Speech	0.77 (0.11)	0.81 (0.12)	0.73 (0.19)	0.54 (0.21)
Video	0.81 (0.13)	<b>0.88 (0.12)</b>	0.74 (0.23)	0.64 (0.24)
Multimodality	<b>0.83 (0.11)</b>	0.88 (0.13)	<b>0.78 (0.20)</b>	<b>0.68 (0.22)</b>

For the purpose of evaluating the best classification model, the models were trained on the set of all video, audio, and multimodal features thanks to the 10-

fold cross-validation. Table 5.12 presents the results of that classification. The multimodalit approach achieved the highest scores for balanced accuracy (0.83), specificity (0.78), and MCC (0.68). The sensitivity was equal to 0.88 for the video and multimodality.

The interpretabilities of those models for video, audio, and multimodality approach are presented thanks to the SHAP values in Fig. 5.10, Fig. 5.11 and Fig. 5.12, respectively.

For the two features among 10 the best were observed the positive correlation with PD disease for video modality. They were: the approximate entropy of the eyelid during the pronunciation vowel ‘a’ (aeEYE12 (TSK13)) and the slope of the time series - the skew distance of the mouth during a repeating sentence difficult to pronounce (slope M7 (TSK41)). The negative correlation was registered for the approximate entropy of the eyelid during the pronunciation of a repeated, hard-to-pronounce sentence (aeEYE16 (TSK37)) and the minimum of horizontal mouth distance during pronunciation vowel ‘e’ (minM5 (TSK4)). Additionally, negatively correlated features were: rsdEYE18 (TSK32), maxD5 (TSK31), varD9 (TSK4), meanD9 (TSK9), aeD9 (TSK4).

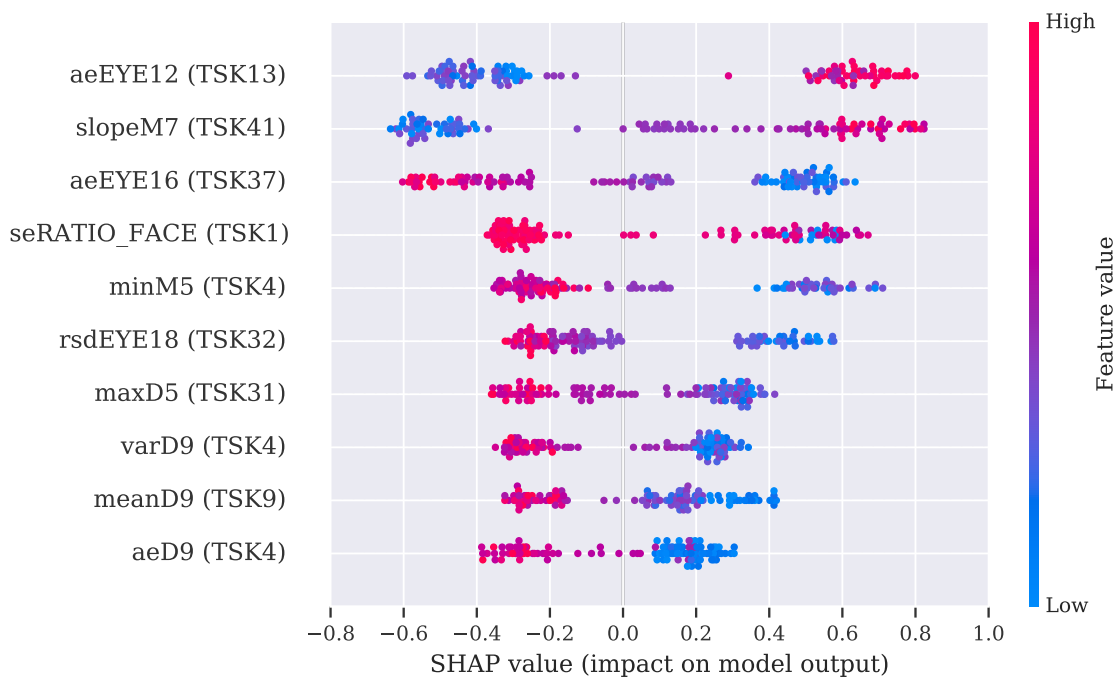


Fig. 5.10: SHAP values for the best 10 features from the video modality.

SHAP values for audio features showed negative correlation with the PD for the following features: relF0SD (TSK24), relF0SD (TSK7), shimmer (TSK15), jitter (TSK14), and shimmer (TSK17). A positive correlation was recognised among the

features based on SHAP values for NSR (TSK30) and relF1SD (TSK21). Fig. 5.11 illustrates the relationship between PD disease and audio features.

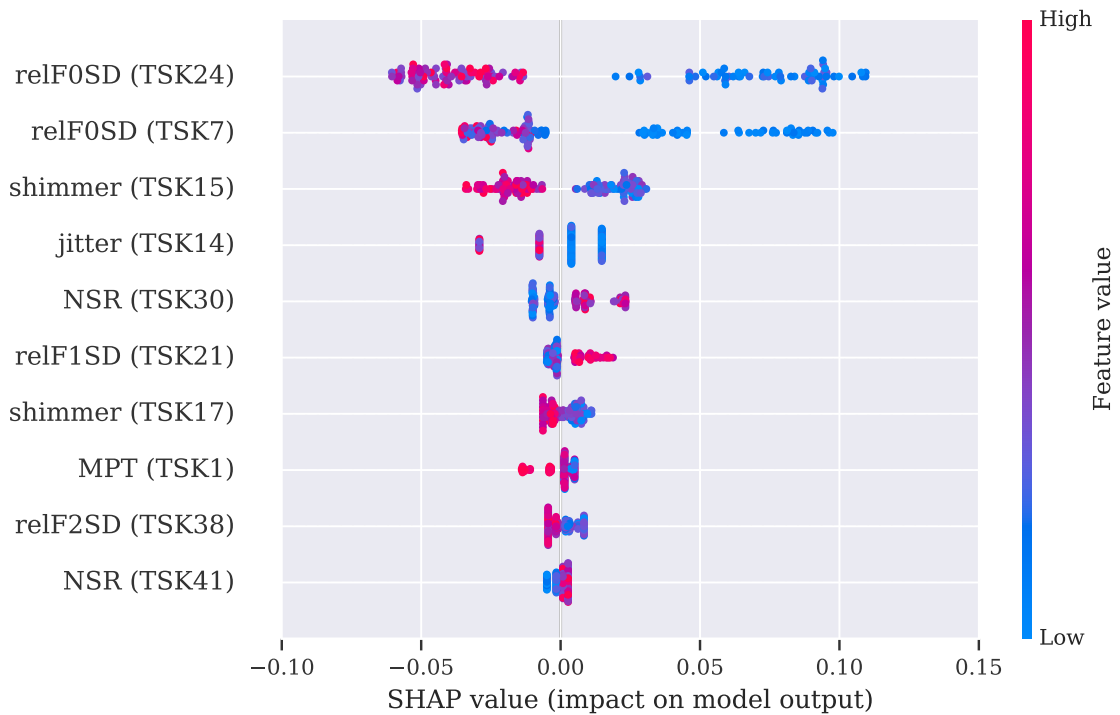


Fig. 5.11: SHAP values for the best 10 features from the audio approach.

Multimodality analysis revealed a positive correlation between PD and slopeM7 (TSK41), aeEYE12 (TSK13), minEYE20 (TSK35) and mean HNR (TSK15). A negative correlation was observed between PD and aeEYE16 (TSK37), varD9 (TSK4), rsdD8 (TSK32), maxM3 (TSK31), and varM6 (TSK12). Fig. 5.12 shows the SHAP values for the multimodality approach.

As an additional step, every exercise was classified according to its video, audio, and multimodal configuration. There are three tables presenting the results of these tests: 5.13, 5.14, 5.15, respectively.

According to the balanced accuracy, the prediction for the TSK39 was the most successful for the video modality and was equal to 0.73 (see Table 5.13). One of the tongue twisters was this task. The MCC for this same task was 0.47. TSK4 (pronouncing the vowel 'e') had the highest sensitivity of 0.81. TSK1 had the highest specificity of 0.69. The exercise consisted of the maximum phonation of 'm' in one breath.

TSK7 was classified with the highest level of balanced accuracy using audio features (see Table 5.14). This parameter was 0.68, and the task responsible for determining it was the pronunciation of the vowel 'u'. A MCC of 0.36 was obtained,

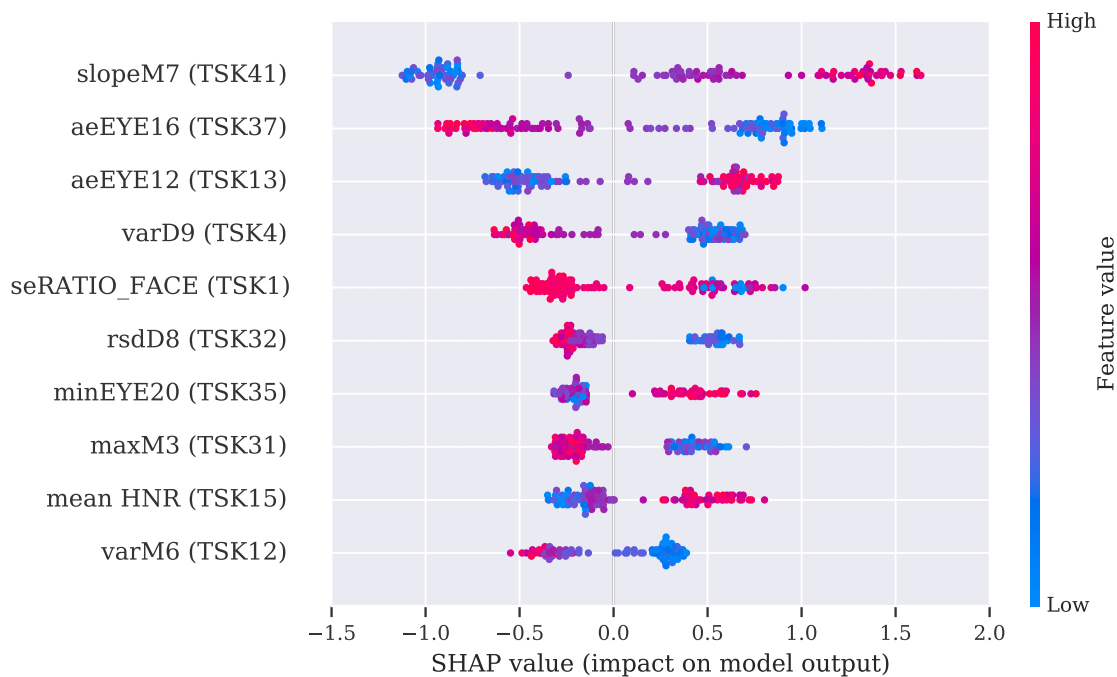


Fig. 5.12: SHAP values for the best 10 features from the multimodality.

Tab. 5.13: Video-based speech exercise accuracy.

Exercise	Accuracy (balanced)	Sensitivity	Specificity	MCC
<b>TSK39</b>	<b>0.73 (0.14)</b>	0.78 (0.17)	0.67 (0.24)	<b>0.47 (0.29)</b>
<b>TSK41</b>	0.73 (0.13)	0.79 (0.16)	0.66 (0.21)	0.47 (0.26)
<b>TSK40</b>	0.72 (0.15)	0.79 (0.17)	0.65 (0.26)	0.46 (0.30)
<b>TSK4</b>	0.72 (0.13)	<b>0.81 (0.14)</b>	0.63 (0.23)	0.46 (0.27)
<b>TSK9</b>	0.72 (0.13)	0.79 (0.15)	0.65 (0.22)	0.45 (0.26)
<b>TSK13</b>	0.71 (0.15)	0.79 (0.17)	0.62 (0.25)	0.43 (0.30)
<b>TSK23</b>	0.71 (0.15)	0.80 (0.14)	0.62 (0.25)	0.44 (0.30)
<b>TSK8</b>	0.71 (0.14)	0.80 (0.15)	0.62 (0.24)	0.44 (0.29)
<b>TSK1</b>	0.71 (0.13)	0.72 (0.18)	<b>0.69 (0.21)</b>	0.42 (0.26)
<b>TSK35</b>	0.71 (0.13)	0.78 (0.15)	0.64 (0.22)	0.42 (0.26)

while a specificity of 0.66 was achieved. A maximum sensitivity of 0.77 was registered for TSK24 (monitoring prosody).

The results of the multimodal classification for each task are presented in Table 5.15. The most valuable exercise was TSK41, with a balanced accuracy of 0.74 and a MCC of 0.49. It is another tongue twister. The highest sensitivity was observed



Tab. 5.14: An assessment of the accuracy of the best speech exercises based on audio recordings.

Exercise	Accuracy (balanced)	Sensitivity	Specificity	MCC
<b>TSK7</b>	<b>0.68 (0.13)</b>	0.71 (0.15)	<b>0.66 (0.22)</b>	<b>0.36 (0.26)</b>
<b>TSK24</b>	0.67 (0.12)	<b>0.77 (0.13)</b>	0.57 (0.22)	0.35 (0.25)
<b>TSK14</b>	0.66 (0.12)	0.70 (0.15)	0.61 (0.20)	0.31 (0.24)
<b>TSK19</b>	0.66 (0.11)	0.67 (0.14)	0.65 (0.21)	0.32 (0.22)
<b>TSK15</b>	0.64 (0.12)	0.75 (0.12)	0.53 (0.23)	0.28 (0.25)
<b>TSK37</b>	0.62 (0.14)	0.64 (0.14)	0.61 (0.21)	0.24 (0.27)
<b>TSK41</b>	0.62 (0.14)	0.65 (0.15)	0.59 (0.23)	0.23 (0.27)
<b>TSK42</b>	0.62 (0.13)	0.73 (0.14)	0.51 (0.22)	0.24 (0.27)
<b>TSK11</b>	0.61 (0.13)	0.64 (0.13)	0.58 (0.21)	0.21 (0.26)
<b>TSK22</b>	0.61 (0.12)	0.66 (0.16)	0.56 (0.21)	0.22 (0.24)

Tab. 5.15: Evaluation of the accuracy of multimodal speech exercises.

Exercise	Accuracy (balanced)	Sensitivity	Specificity	MCC
<b>TSK41</b>	<b>0.74 (0.13)</b>	0.79 (0.15)	0.68 (0.22)	<b>0.49 (0.27)</b>
<b>TSK23</b>	0.73 (0.15)	<b>0.83 (0.14)</b>	0.62 (0.26)	0.47 (0.32)
<b>TSK39</b>	0.73 (0.14)	0.78 (0.17)	0.67 (0.24)	0.47 (0.29)
<b>TSK18</b>	0.73 (0.13)	0.78 (0.16)	0.68 (0.23)	0.48 (0.27)
<b>TSK40</b>	0.72 (0.16)	0.80 (0.16)	0.64 (0.25)	0.46 (0.32)
<b>TSK8</b>	0.72 (0.14)	0.81 (0.15)	0.63 (0.24)	0.45 (0.28)
<b>TSK22</b>	0.72 (0.14)	0.75 (0.17)	<b>0.69 (0.24)</b>	0.44 (0.28)
<b>TSK4</b>	0.72 (0.13)	0.82 (0.15)	0.62 (0.24)	0.46 (0.27)
<b>TSK9</b>	0.72 (0.13)	0.78 (0.16)	0.65 (0.21)	0.45 (0.25)
<b>TSK1</b>	0.71 (0.13)	0.72 (0.18)	0.69 (0.21)	0.42 (0.26)

for TSK23 (during monitoring of prosody), which was equal to 0.83. TSK22, during which the sentence was read in a declarative manner, displayed the highest specificity (0.69).

The results for the multimodality and video have been compiled in Table 5.16 to compare the results of the most powerful speech exercises. Five of the 10 exercises showed an improvement in classification, and two showed no improvement. In the multimodality approach, TSK41 (the tongue twister) demonstrated the best accuracy of 0.74.

Tab. 5.16: An analysis of the results obtained from multimodal and video approaches.

Exercise	Accuracy (balanced) for multimodality	Accuracy (balanced) for video
<b>TSK41</b>	0.74 (0.13)	0.73 (0.13)
<b>TSK23</b>	0.73 (0.15)	0.71 (0.15)
<b>TSK39</b>	0.73 (0.14)	0.73 (0.14)
<b>TSK18</b>	0.73 (0.13)	0.71 (0.12)
<b>TSK40</b>	0.72 (0.16)	0.72 (0.15)
<b>TSK8</b>	0.72 (0.14)	0.71 (0.14)
<b>TSK22</b>	0.72 (0.14)	0.70 (0.13)
<b>TSK4</b>	0.72 (0.13)	0.72 (0.13)
<b>TSK9</b>	0.72 (0.13)	0.72 (0.13)
<b>TSK1</b>	0.71 (0.13)	0.71 (0.13)

Figs. 5.13, 5.14, 5.15 presents the SHAP values for determined the most accurate classification of video, audio, and multimodality, respectively.

For the tongue twister (TSK39) (see Fig. 5.13), the video features which were positively correlated with PD are varEYEBROW3, slopeEYE13, and minEYE1. In contrast, negative correlations were found between rangeM5, maxM5, minEYEBROW3, maxD6, aeREYE\_AREA, aeEYEBROW5 and PD.

According to the audio results (see Fig. 5.14), the TSK7 speech exercise provided the best results. There is a negative correlation between DUV, relF0SD, and shimmer and PD. Mean HNR and relF1SD have a positive correlation.

For the task TSK41 - the tongue twister (see Fig. 5.15), the multimodal approach shows higher values of slopeM7, slopeD4, meanD2, and meanEYE5 for positive SHAP values. aeEYE9, aeEYE5, aeMOUTH\_AREA, slopeD7, maxRATIO\_FACE, and slopeEYE18 were lower for negative SHAP values.

## 5.2.5 Discussion and Summarisation

As part of this research, multiple scenarios were performed, including evaluating of variety of Czech speech exercises, as well as a performing multimodality approach for detecting PD. The carried-out experiments allow to find the best-fitted model based on the most appropriate features for this task. As a result of the choice of XGBoost, it is possible to train ML models capable of capturing the connection between impairments of facial muscles and observable changes in HD. The highest

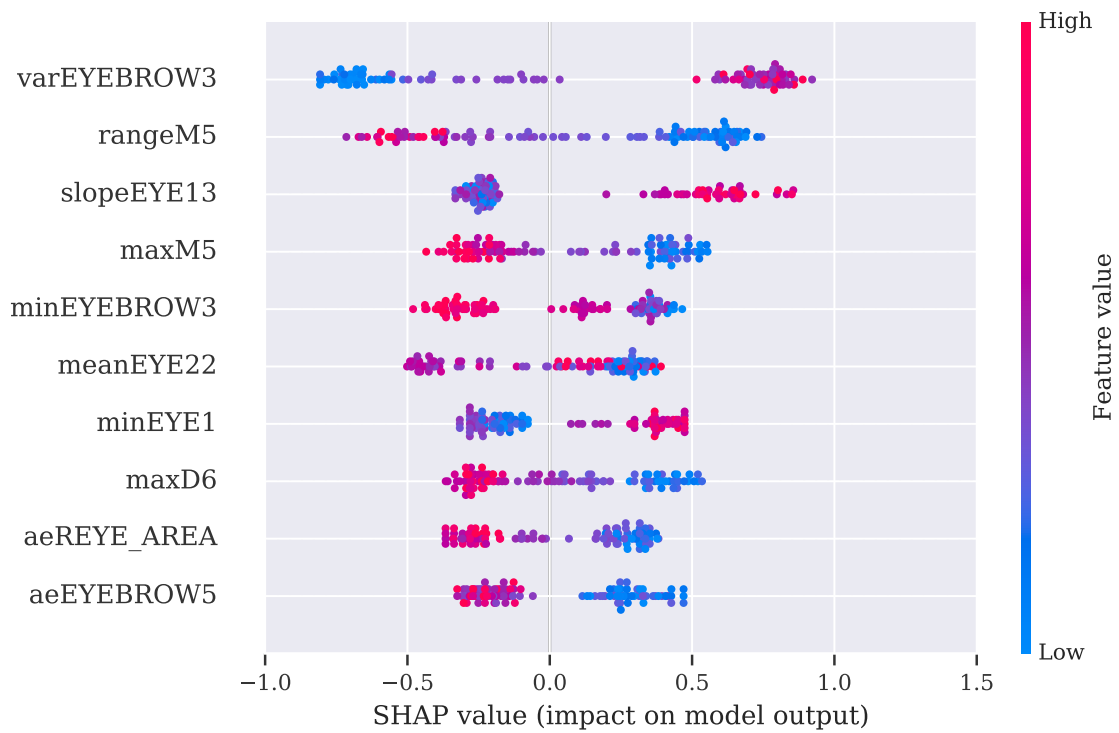


Fig. 5.13: SHAP values for the optimal video approach (TSK39).

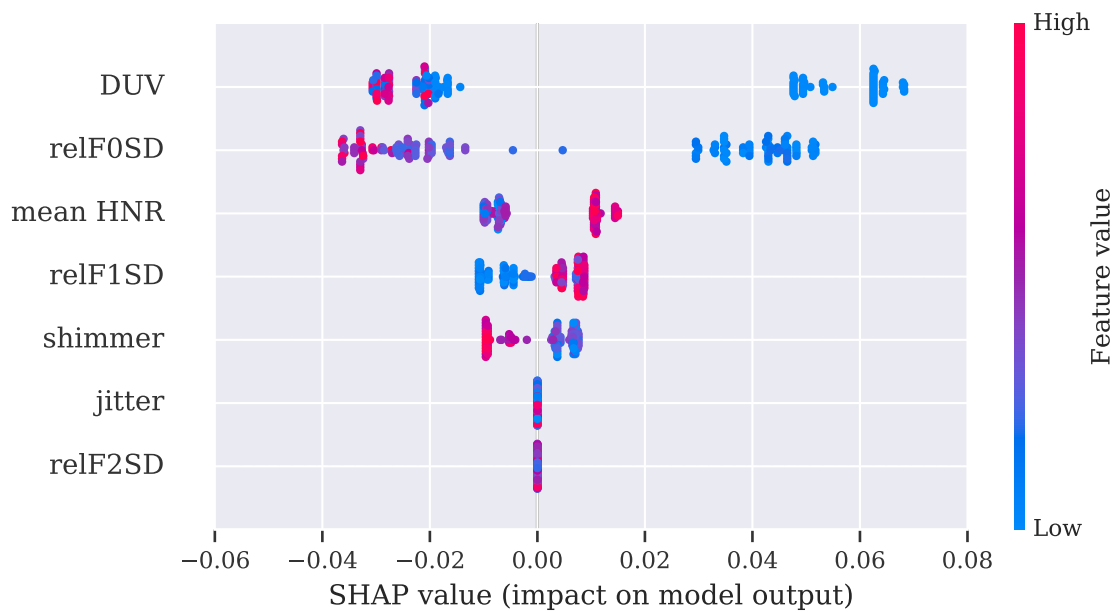


Fig. 5.14: SHAP values for the optimal audio approach (TSK7).

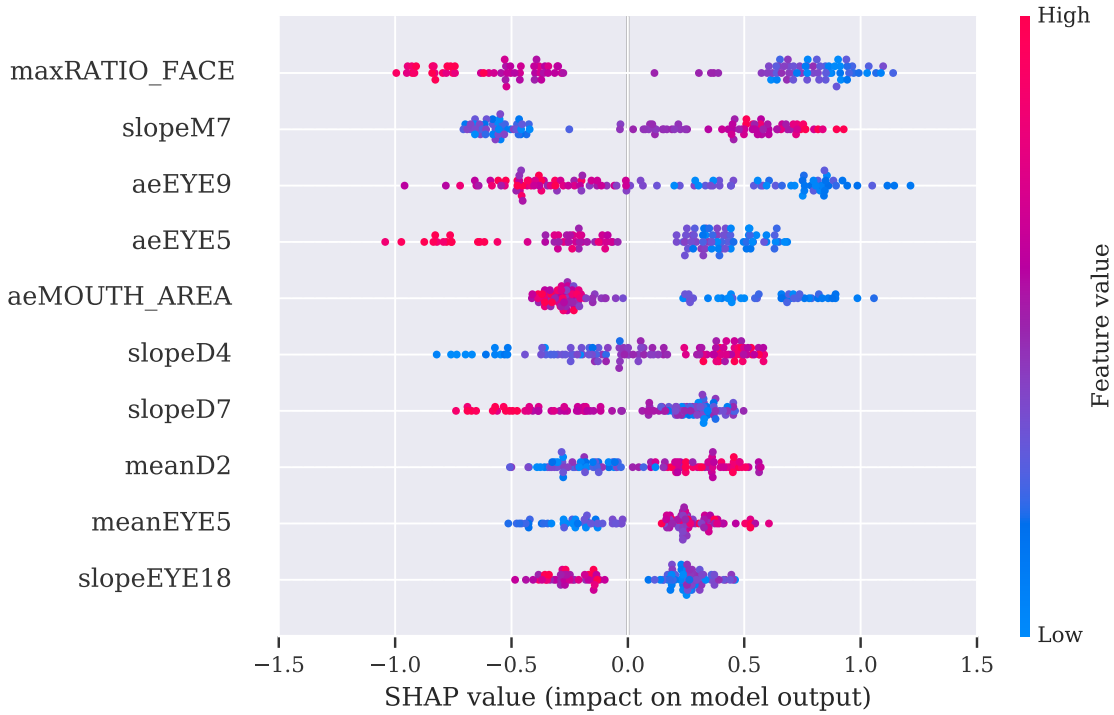


Fig. 5.15: SHAP values for the optimal multimodal approach (TSK41).

level of accuracy can be achieved as a result of the combination of various speech exercises with multimodality. The multimodality outperformed the single modality-based approach in the detection of PD. The best-achieved accuracy was equal to 0.83.

Considering the statistical analysis performed with the Mann-Whitney U test, audio features were more significant than video features after FDR correction. The most distinctive feature was shimmer across all of the exercises and was characteristic of the pronunciation of the vowels ‘i’, ‘u’, ‘a’, and ‘o’. A further important feature is the rsd of fundamental frequency when the vowel ‘u’ is pronounced and the variability in intonation is checked. Fraction of locally unvoiced frames matters for the pronunciation of vowels ‘u’ and ‘a’. Net speech rate also proves the importance of checking intonation variability.

Any of the features considered in the video analysis passed the Mann-Whitney U test with FDR correction. Nonetheless, they were close to a significant level of  $\alpha = 0.05$  and equal to 0.0733. Most of them were recognised for performing tongue twisters. Moreover, the diadochokinesis (DKK) task was valuable and the pronunciation of the vowel ‘u’. The changes in blinking rate expressed in rsd and approximate entropy were recognised as a beneficial sources of information. The slope of the records of changes in opening mouth diagonal was detected as informative,

exactly variance and std. Additionally, the mean horizontal distance of mouth was valuable for the diadochokinesis task.

The results of the experiments confirmed the statement that it would be better to use both modalities than a single one. A combination of subdatasets resulted in a higher level of specificity and accuracy. The sensitivity occurred to be the best for video modality and 0.88 of the PD cases were positively diagnosed.

The models revealed that the most informative features are tongue twisters – difficult-to-pronounce speech exercises. They were for video modality: TSK39, TSK41, TSK40 (see Table 5.13). Moreover, the pronunciation of the vowel ‘e’ caused a problem for patients (TSK4, TSK9). A similar situation was observed for the vowel ‘a’. For the audio features, the most problematic task was to pronounce the vowel ‘u’ (see Table 5.14). The vowels ‘e’ and ‘i’ created also difficulty (TSK14, TSK15). Furthermore, intonation variability was highly significant in distinguishing PD cases from HC cases (TSK24). Moreover, reading a poem by patients indicates difficulty in expression by PD patients (TSK19). Tongue twisters also worked out for this modality (TSK37, TSK41, TSK42). The combination of the video and audio modality (Table 5.15) showed that the most robust diagnosis of PD was tongue twister (TSK41). Similarly, another tongue twister was capable of distinguishing illness at a high level (TSK39, TSK40). The intonation variability offers the valuable distinction of PD cases from HC (TSK23). The diadochokinesis task was also successful in this task (TSK18). Regarding multimodality, the classification analysing the pronunciation of the vowel ‘e’ achieved a relatively high result (TSK4, TSK9). The interpretability of the models was performed thanks to the SHAP values. They indicated a positive and negative correlation between the features of the PD.

The SHAP values for the combination of all features showed the importance of changes in eye blinking during the pronunciation vowel ‘a’ (aeEYE12 (TSK13)) for video modality (Fig. 5.10). The approximate entropy of the eye distance showed a positive relationship with PD. It means that deviations from constant values of opening eyelid were positively correlated with PD – observation of unnormal behaviour of eyes. During the longer activity, i.e., the pronunciation of tongue twister was observed a negative correlation between PD and the irregular pattern of eye behavior (aeEYE16 (TSK37)). The changes in the mouth’s ability to pronounce were observed (slopeM7 (TSK41), minM5 (TSK4)). slopeM7 (TSK41) indicates the highest slope of the time series of the skew distance of the mouth during the pronunciation of tongue twister, which could be explained by the opening mouth longer time by PD patients of the mouth. The minimal value of the horizontal distance of the mouth was negatively correlated with PD during the pronunciation of the vowel ‘e’. Therefore, the HC group is able to open their mouths widely. There was also a negative correlation between maximum movement between eyelid and eyebrow, and

PD, which indicates higher level of 'freezing' of the face for persons with PD. That was observed during the pronunciation of difficult Czech word (maxD5 (TSK31)). Furthermore, there were challenges in moving the jaw during the pronunciation of the vowel 'e' (varD9 (TSK4), meanD9 (TSK9), and aeD9 (TSK4)).

According to the speech analysis using SHAP values, there is a negative correlation between the relative standard of fundamental frequency during monitoring prosody and pronunciation of the vowel 'u' (relF0SD (TSK24, TSK7)). This dependency with PD was also alleged by [243]. In addition, shimmer and jitter were negatively correlated when participants were pronouncing vowels 'e', 'i', and 'u'. Nevertheless, the work [244] reports that the lower values are typical for women PD patients and ill men characterise higher values. This situation could be explained by the fact that the data were regressed out.

In addition, the multimodality analysis model indicates a negative correlation between rsd of the eyelid (rsdD8 (TSK32)) and hard to pronounce words. This could be explained by the fact that PD is associated with a lower blinking rate. For PD, a lower value of skew of the mouth was observed (maxM3 (TSK31)) when persons were speaking a difficult word. It was noted that the open mouth variance was lower for PD when pronouncing 'u' (varM6 (TSK12)). Furthermore, PD displayed a higher harmonic-to-noise ratio during the pronunciation of vowel 'i' (mean HNR (TSK15)). The authors in [245, 246] also confirmed it.

TSK39 was identified as the most effective speech exercise for the video (Fig. 5.13). PD was positively correlated with variation in ankle movement between eyebrows (varEYEBROW3) and its minimal value was lower for PD (minEYEBROW3). Additionally, HC had a higher chaotic movement of the distance between the horizontal outer corner of the eye and the eyebrow (aeEYEBROW5). In addition, a dependency of keeping the open mouth horizontally was detected for the parameter range to HC (rangeM5) and the maximum to HC (maxM5). HC demonstrates a greater degree of flexibility when it comes to expressing yourself. An increase in approximate entropy was observed for the right eye area in HC (aeREYE\_AREA) - this is due to higher blinking rate of HC. The slopeEYE13 parameter was also used to measure keeping longer eyes open.

The audio analysis of single speech exercises (Fig. 5.14) – pronunciation vowel 'u' (TSK7) shows the negative correlation with PD for fraction of locally unvoiced frames (DUV), relative std of fundamental frequency (relF0SD), and shimmer. The negative relationship of relF0SD was reported in [243]. The reverse phenomenon was registered for DUV and shimmer [246], most probably because of regression data out and gender issues. Whereas, the positive dependency was observable for the mean of harmonic-to-noise ratio (mean HNR) and relative std of 1st format (relF1SD). Those relations were reported in [245, 246].

The TSK41 was considered to be the most informative task for the multimodal approach (Fig. 5.15). The closer the opening mouth value to the constant value, the greater the chance that the model will identify the sample as PD. The symptoms of PD were also manifested in the pattern of eye blinking this time. HC showed a higher entropy in the movement of the eyelids (aeEYE9, aeEYE5). HC displayed a more chaotic pattern of open mouth during the pronunciation tongue twister (aeMOUTH\_AREA). The parameter slopeD4 indicates longer staying open the mouth for PD with limited movability. This conclusion is also supported by the parameter meanD2 behaviour.

The limitation of this research is the size of the dataset: 46 HC and 73 PD patients. However, the amount of participants in the experiment is quite big considering another dataset used for automatic hypomimia analysis and PD recognition (see Table 3.2). If the proposed approach is used for clinical purposes, it should be malleable on a larger dataset. Furthermore, the individuals were using glasses during the test. They were needed for the reading task of the text. Nevertheless, the outcome of the classification could have been improved if the individuals had taken them off.

To summarise, the SHAP values illustrate the changes in eye-blinking for PD patients, and the impairments for PD. There was a slower pace of movement and patients kept their eyes open for a longer period of time. Comparatively to HC, the mouth movements were limited. Patients with PD used to keep their eyes open. The stacked movement of the jaw is also characteristic for PD. During the pronunciation tongue twister, the changes in facial muscles were most evident, whereas a variety of audio patterns were observed for the pronunciation of vowels. It is typical for PD to have a lower value of relF0SD. There was a positive correlation between mean HNR and PD.

To summarize this section, the support methodology for PD detection based on multimodality, i.e., video and audio was presented. The proposed dataset is exceptional and contains 73 PD patients and 46 HC. When comparing other datasets introduced by scientists, this dataset is quite large and it uses a multimodal approach, which is a significant benefit (see Table 3.2). A total of 43 unique speech exercises were evaluated in order to identify the most reliable ones. The strength of this research is the identification of the most effective and clinically valuable speech exercise - tongue twister. Moreover, the results obtained by the XGBoost classifier were satisfactory. The multimodal approach showed that it outstands the solutions based on a single modality. There is so far a limited number of works dealing with multimodal solutions for PD detection. With the usage of a multimodal approach, the detection of PD was possible on the level 0.83 balanced accuracy. Several facial and audio features were prepared and the most significant features were selected

during the feature selection procedure, which allows models to achieve higher accuracy of PD detection. The desirable clinical interpretability was obtained thanks to the statistical analysing and SHAP values. The performed statistical analysis revealed the statistical importance of the features. The SHAP values explain the value of parameters measuring the eye blinking, the openness of the mouth, and asymmetry of the face. Those founded dependencies are confirmed by the literature [88,105,108,110]. The most important features occurred to be parameters such as variance, approximate entropy, and slope of computed biomarkers. This work explored the clinical value and various speech exercises power in the prediction of PD with the potential to be applied as a mHealth solution. The accuracy of PD detection based on the most powerful speech exercise – tongue twister and multi-modality achieved 0.74 balanced accuracy. By extending the dataset and using AUs for PD diagnosis, the accuracy and robustness of the model could be improved.

### 5.3 Conclusion

The need for the creation of more approachable and inexpensive solutions for PD detection than PET, CT, MRI, and PSG exists [63,92]. There is still a lack of techniques dedicated for mHealth and AAL. The methodologies presented in this chapter are the answer for the occurring niche and two approaches to the detection of PD were presented. The common denominator of them was the analysed symptom of PD – hypomimia. The aim was to create not only the support methodologies for PD recognition but also to identify the computational biomarkers which could be clinically interpretable.

The first methodology analysed computationally the difficulties in expressing emotions (section 5.1), whereas the second approach thanks to the created facial features explored the challenges of moving facial muscles and mimic of the participants of the experiments (section 5.2). The advantage of this research is that the anthropometric characteristics were taken into consideration during facial features design. The database used for research in sections 5.1 and 5.2 was the same. The second approach, in section 5.2, also used the audio modality and HD symptom, combined moreover with the video modality. Subsequently, here, the existing research gap was explored where there is still a lack of multimodal solutions of PD detection. Moreover, a limited number of works about hypomimia analysis for PD are already published. Furthermore, the dataset which was used for the research is unique, and it is relatively large in comparison to those presented in the literature. This dataset allows for identifying the most powerful speech exercise. Furthermore, the proposed methodologies are as objective as possible and not troublesome for patients.



In the section 5.1 and 5.2, both studies reported the difficulties in performing the tongue twister and the highest usefulness of this speech exercise for the clinical purpose of detecting PD. This speech exercise - tongue twister is the answer to the questions **RQ4.1.** and **RQ5.2.** Considering the research in which the recognition of emotion was utilised (section 5.1), the most informative emotions were identified thanks to the statistical analysis and they were fear and anger.

Furthermore, the identification of the correlation between the parameters based on emotions and PD was not earlier provided in the literature, considering SHAP values. The approximate entropy of sadness showed its importance as a feature. Additionally, the interpretability of SHAP values depicted also the positive correlation between PD and fear likewise anger (fear\_std, fear\_variance, fear\_range, fear\_max, fear\_mean, angry\_max). Whereas, the negative correlation was reported for approximate entropy of surprise and sadness for a checked tongue twister. Those identified features are the reply to the question **RQ4.2.** The prediction of PD was achieved with the XGBoost algorithm and was equal to 0.69 balanced accuracy. This accuracy is the answer to question **RQ4.3.** The other tongue twisters have the potential to increase the accuracy, according to the section 5.2.

In the second scenario 5.2, the model, which was presented for the multimodal approach, and all speech exercises could serve as a support methodology system. The best prediction was obtained for this version of the model thanks to the XGBoost and was equal to 0.83 balanced accuracy. This approach and accuracy are the answer to question **RQ5.3.** While the result for the multimodal approach for single exercise TSK41 was equal to 0.74 balanced accuracy. The multimodal solutions proved to be outperforming in contrast to single-modality approaches.

The interpretability of the models illustrated thanks to the SHAP values showed that PD manifests in blinking rate during the pronunciation of vowel ‘a’ likewise during the pronunciation of tongue twister. Mouth ability to move is higher for HC, whereas PD patients tend to keep their mouths open longer. Freezing of the eyebrow was also visible which could indicate the stiffness in facial muscles. The video features based on the combination of slope, approximate entropy, and variance occurred to be the most valuable.

Furthermore, the HC displays a more chaotic movement of the mouth during pronunciation tongue twister (TSK41). On the top of that, the stacked movement of the jaw is also characteristic of PD. The audio analysis presented also the negative relationship of relF0SD during monitoring pronunciation of the vowel ‘u’ and prosody to PD. To sum up, the audio features reveal the disease during the pronunciation of vowels, whereas the video features detect PD thanks to tongue twisters. What’s more, the biomarkers were identified and could serve clinicians. Those two above-mentioned paragraphs are the answer to the question **RQ5.1.**

In conclusion, the introduced multimodal model could be used as the solution for mHealth (section 5.2). In the case of the first scenario (section 5.1), other tongue twisters have to be tested, if there are more suitable and could obtain higher accuracy in prediction PD based on difficulties in expressing the emotions.

## 6 Final Conclusion

First and foremost, the thesis considers developing the application of wearables and ML technologies for healthcare solutions. The WHT is still expanding type of technology on the market and will consume and adapt new solutions based on wearable technology. Because of this reason, the developed technologies in this thesis are appropriate, considering the still progressing field. Additionally, the desirable solutions are those with provided interpretability. The main focus of this thesis is concentrated on used ML methods likewise also the applicability of the presented support system methodologies - i.e., their performance and interpretability.

Two main thematic issues were discussed in this thesis, i.e., the wearable solutions for COVID-19 detection likewise the application of AAL for PD. The introduced research in this thesis is focused on finding solutions to prevent and minimize the effects of those emergency problems. The common denominator in this work is the usage of ML to generate support methodologies.

The section Introduction is guiding the readers into the topic. The background about COVID-19 and its diagnosis likewise PD and its recognition, EEG analysis are presented in Chapter 2. Additionally, the thesis presents the state-of-the-art of the discussed scientific problems in Chapter 3. The summarisation of the existing approaches for the detection of COVID-19 and pandemic models with the usage of wearable and ML is described. Next, a description of how PD symptoms hypomimia, HD, and sleep disorders can be used to diagnose PD is illustrated. Moreover, the detailed representations of the recognition of other diseases based on actigraphy records and ML are presented in subsection 3.2.3. Those methods were identified because of their potential applications for PD detection based on sleep records. Furthermore, the readers could familiarise themselves with the deep learning techniques for EEG analysis, especially the potential of the novelty - ODE network.

The practical solutions for COVID-19 detection based on ML and wearables for three scenarios are presented in Chapter 4. The thesis tried to find the answer to the emergency need for screening tests in the early stage of the disease. Two datasets are the basement of those presented solutions [1,59]. The kinds of analysing signals were heart rate and the number of steps taken. The data were gathered by the Fitbit device and presented solutions are destined for this device. Moreover, the biggest contribution of the illustrated approaches is the consideration of the nature of the disease, i.e., the contagiousness of the disease and incubation period. To reduce the increase in the number of sick people, those two parameters were taken into account. Additionally, the amount of presented solutions based on ML and wearables dedicated to COVID-19 detection is limited. By the same token, there is still room for exploring this area.

As the outcome of the first experiment, the support system methodology for the emergency issue - COVID-19 detection in the early stage of illness, was presented. The best-identified model based on 5-day windows allows obtaining the prediction of 78 % accuracy for differentiating the COVID-19 cases from HC. The two time 5-day windows were chosen for each sample – time series and from them were extracted bunches of features. To emphasise, this outcome is significantly better than the one presented in the original paper [59]. Furthermore, the model treated the COVID-19 cases and Influenza as one group-identified illness with 73 % of accuracy, also for 5-day windows. The used classifier for both scenarios was k-NN. The same results for the scenario with Influenza were also achieved thanks to the GLVQ. The consensus is that this solution could serve as a screening test. An additional benefit is that the most relevant features based on statistical evaluations were identified and they were frequency- and spectral-related. Furthermore, the unquestionable advantage is that the solution could be introduced as one of eHealth approach. Smartwatches are commonly available and can be used as a helpful diagnostic tool when dealing with outbreaks of pandemics.

The second set of support methodologies aimed to distinguish COVID-19 cases from the two different types of Influenza in the early stage. Two types of Influenza were distinguished before the main pandemic and in the middle of the pandemic. The dataset intrinsically is unique because have instances of different viruses. This thesis extended the original research in [1] because it introduced the classifications, not only statistical analysis like in the [1] paper. The developed support methodologies allow distinguishing the COVID-19 cases between the Influenza in the middle of the pandemic on the level of 0.73 balanced accuracy thanks to the k-NN. Moreover, the distinction between the two types of Influenza was achieved with 0.82 balanced accuracy thanks to the GLVQ. What's more, this study proved the existing differences between COVID-19 cases and Influenza being able to recognise cases by wearable and ML for the first time.

The last support methodologies presented in this chapter are the models malleable on combinations of the two previously used datasets. The data contained the representations of the COVID-19 cases, HC, and Influenza. The introduced dataset is relatively big to those limited and presented in the literature having COVID-19 examples. The advantage of this study is the presented methodologies to differentiate cases from various classes. The COVID-19 cases were distinguished from HC on the level of 0.73 accuracy by XGBoost. Whereas, the COVID-19 cases and Influenza were differentiated from HC with 0.72 accuracy by SVM and GLVQ. The statistical analysis revealed that the highest statistical importance was registered for the features generated as the parameters of the ratio of heart rate to the variable of the number of steps.

The next presented subject in this thesis concerned computerised automatic PD detection. The outcomes of the research were described in Chapter 5. The presented methods in the thesis could be potentially applied as still lacking the mHealth methods for PD recognition. They are more inexpensive and more accessible alternatives to the common tests: PET, MRI, CT, and PSG. In addition, there still exists a research niche about hypomimia analysis for PD detection (Tab. 3.2). Moreover, not only the support methodologies are desirable but also clinical interpretability is well perceived together with recommending valid biomarkers.

The first method of PD detection used the automatic analysis of changes in expression emotion by participants. The symptom which was the foundation of this research was hypomimia. The automatic analysis with speech exercises: reading the poem and tongue twister was especially valuable because could be as much nonsubjective and not bothersome as possible (Tab. 3.2). Neural network - FER was utilised to recognise the intensity of seven emotions in each frame. To train the model, parameters expressing changes in those emotions were used. The best-obtained prediction of PD was equal to 0.69 balanced accuracy for the XGBoost classifier and tongue twister. The tongue twister was identified as a clinically valuable speech exercise. As the most informative emotion occurred to be fear and it was registered as positively correlated with PD based on SHAP values. The found captivating biomarkers were fear-related. To emphasise, such analysis was not earlier provided in the literature.

The second example of PD detection used the computerised multimodal approach analysing hypomimia and HD. There are just few papers that treated the multimodal detection of PD, especially with the participation of hypomimia symptom. The choice of the audio and video modality and combined ML methodology allowed for obtaining 0.83 balanced accuracy for the fusion of biomarkers generated for all studied speech exercises. The used classifier was XGBoost. As the most predictive and powerful speech exercise was recognised the tongue twister for the multimodal approach. Furthermore, the assumption of this study was confirmed by the results, the usage of multimodal approaches displayed to be better than those based on a single modality. By the same token, it makes sense to merge the modalities for PD detection. The tongue twister - speech exercise was identified as the most appropriate for the utility in clinical practice. A plethora of them was tested and such exploration of the suitable tasks has not been earlier made. What's more, the dataset is unique and fairly large vis-a-vis other related datasets (Tab. 3.2).

Moreover, the SHAP values were used to provide interpretability. The understandable solutions of the ML allowed identifying the features which are the most informative for the models and also were compared with the literature. The most valuable were those which capture the relation to different distances of eyelid, mouth,

likewise angles between eyebrows and others. To summarise, this approach provided a support system methodology that will be valued by clinicians because of its presented interpretability. It could broadly serve as a screening tool for PD detection and it is dedicated to mHealth solutions.

The future directions are targeting the usage of ML methodologies for PD detection based on multimodal approaches and extending databases. Furthermore, there is still space for researching COVID-19 detection thanks to the wearables, increasing databases, and extending the number of types of analysing signals.

To sum up, there were introduced ML-aided monitoring and prediction of respiratory and neurodegenerative diseases using wearables in the thesis. The first topic considered COVID-19 detection. The solutions were destined for early recognition thanks to taking into account contagiousness and incubation period. It is an incredible asset. There is still a research gap for COVID-19 detection, and the obtained outcomes in this thesis outperformed those already presented in the literature. Thereby, they introduced new solutions for diagnosis of COVID-19. Moreover, a longer discussion together with illustrating the classification between a few types of viruses, including COVID-19 was presented. Such distinctions have not been previously published. The merging of two datasets allows for having the largest such dataset. Moreover, the second topic presented the methods of PD detection. The symptom of hypomimia was the basement of the presented classification problem. The research on this topic filled the existing scientific demand. Furthermore, the HD was also evaluated. The multimodal methodology was proposed and approved as better than the single modality. The tongue twister arose as the best speech exercise which has special clinical value. The extra interpretability of the experiment was provided thanks to the statistical analysis and SHAP values. Furthermore, the computational analysis of emotion demonstrated to have potential in recognition of PD and could replace the other uncomfortable or fair subjective tests. Additionally, the transfer of probable methodology of detection PD based on discrepant sleep disorders was explored and depicted.

## Bibliography of Author

- **Justyna Skibińska**, Jiri Hosek, and Asma Channa. Wearable analytics and early diagnostic of covid-19 based on two cohorts. In 2022 14th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), volume 2022-October, pages 56–63, Valencia, October 2022. IEEE. doi:10.1109/ICUMT57764.2022.9943460.
- **Justyna Skibińska** and Radim Burget. Is it possible to distinguish covid19 cases and influenza with wearable devices? analysis with machine learning. *Journal of Advances in Information Technology*, 13(3):265–270, 2022. doi: 10.12720/jait.13.3.265-270.
- Asma Channa, Nirvana Popescu, **Justyna Skibińska**, and Radim Burget. The rise of wearable devices during the covid-19 pandemic: A systematic review. *Sensors*, 21(17):5787, 2021. doi:10.3390/s21175787.
- **Justyna Skibińska**, Radim Burget, Asma Channa, Nirvana Popescu, and Yevgeni Koucheryavy. Covid-19 diagnosis at early stage based on smart-watches and machine learning techniques. *IEEE Access*, 9:119476–119491, 2021. doi: 10.1109/ACCESS.2021.3106255.
- **Justyna Skibińska** and Radim Burget. The transferable methodologies of detection sleep disorders thanks to the actigraphy device for parkinson’s disease detection. volume 2880, Tampere, June 2021. CEUR Workshop Proceedings, CEUR-WS.
- **Justyna Skibińska** and Radim Burget. Parkinson’s disease detection based on changes of emotions during speech. In 2020 12th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), volume 2020-October, pages 124–130, Brno, October 2020. IEEE, IEEE Computer Society. doi:10.1109/ICUMT51630.2020.9222446.
- Aleksandr Ometov, Viktoriia Shubina, Lucie Klus, **Justyna Skibińska**, Salwa Saafi, Pavel Pascacio, Laura Flueratoru, Darwin Quezada Gaibor, Nadezhda Chukhno, Olga Chukhno, et al. A survey on wearable technology: History, state-of-the-art and current challenges. *Computer Networks*, 193:108074, 2021. doi:10.1016/j.comnet.2021.108074.
- **Justyna Skibińska** and Radim Burget. The application of deep learning techniques in the electroencephalogram (eeg) analysis. In 2019 XXXV Finnish Union Radio Scientifique Internationale (URSI) Convention on Radio Science, Tampere, October 2019. URSI.
- **Justyna Skibińska**, Abbas Shah, Asma Channa, Muhammad Shehram Shah Syed, Zafi Sherhan Syed, and Jiri Hosek. Foreseeing wearing-off state in Parkinson’s disease patients, a multimodal approach with the usage of ma-

chine learning and wearables. In 5th International Conference on Activity and Behavior Computing (ABC), Kaiserslautern, September 2023.

- **Justyna Skibińska** and Jiri Hosek, Computerised Analysis of Hypomimia and Hypokinetic Dysarthria for Improved Diagnosis of Parkinson’s Disease, 2023, submitted to the Heliyon journal  
.....
- Paulina Koziol, Magda K Raczkowska, **Justyna Skibińska**, Nicholas J McCollum, Sławka Urbaniak-Wasik, Czesława Paluszkiewicz, Wojciech M Kwiatek, and Tomasz P Wrobel. Denoising influence on discrete frequency classification results for quantum cascade laser based infrared microscopy. *Analytica Chimica Acta*, 1051:24–31, 2019. doi:10.1016/j.aca.2018.11.032.
- Paulina Koziol, Magda K Raczkowska, **Justyna Skibińska**, Sławka Urbaniak-Wasik, Czesława Paluszkiewicz, Wojciech Kwiatek, and Tomasz P Wrobel. Comparison of spectral and spatial denoising techniques in the context of high definition ft-ir imaging hyperspectral data. *Scientific reports*, 8(1):14351, 2018. doi:10.1038/s41598-018-32713-7.



# Bibliography

- [1] Allison Shapiro, Nicole Marinsek, Ieuan Clay, Benjamin Bradshaw, Ernesto Ramirez, Jae Min, Andrew Trister, Yuedong Wang, Tim Althoff, and Luca Foschini. Characterizing covid-19 and influenza illnesses in the real world via person-generated health data. *Patterns*, 2(1), 2021. doi:10.1016/j.patter.2020.100188.
- [2] **Justyna Skibińska**, Radim Burget, Asma Channa, Nirvana Popescu, and Yevgeni Koucheryavy. Covid-19 diagnosis at early stage based on smartwatches and machine learning techniques. *IEEE Access*, 9:119476–119491, 2021. doi:10.1109/ACCESS.2021.3106255.
- [3] Hasan K Siddiqi and Mandeep R Mehra. Covid-19 illness in native and immunosuppressed states: a clinical–therapeutic staging proposal. *The Journal of Heart and Lung Transplantation*, 39(5):405, 2020. doi:10.1016/j.healun.2020.03.012.
- [4] Rohan Gupta, Smita Kumari, Anusha Senapati, Rashmi K Ambasta, and Pravir Kumar. New era of artificial intelligence and machine learning-based detection, diagnosis, and therapeutics in parkinson’s disease. *Ageing Research Reviews*, page 102013, 2023. doi:10.1016/j.arr.2023.102013.
- [5] Guokang Zhu, Jia Li, Zi Meng, Yi Yu, Yanan Li, Xiao Tang, Yuling Dong, Guangxin Sun, Rui Zhou, Hui Wang, et al. Learning from large-scale wearable device data for predicting epidemics trend of covid-19. *Discrete Dynamics in Nature and Society*, 2020, 2020. doi:10.1155/2020/6152041.
- [6] Ge Su, Bo Lin, Jianwei Yin, Wei Luo, Renjun Xu, Jie Xu, and Kexiong Dong. Detection of hypomimia in patients with parkinson’s disease via smile videos. *Annals of Translational Medicine*, 9(16), 2021. doi:10.21037/atm-21-3457.
- [7] Ajjen Joshi, Soumya Ghosh, Sarah Gunnery, Linda Tickle-Degnen, Stan Sclaroff, and Margrit Betke. Context-sensitive prediction of facial expressivity using multimodal hierarchical bayesian neural networks. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, pages 278–285, Xi’an, June 2018. IEEE. doi:10.1109/FG.2018.00048.
- [8] Juan Camilo Vásquez-Correa, Tomas Arias-Vergara, Cristian D Rios-Urrego, Maria Schuster, Jan Ruzs, Juan Rafael Orozco-Arroyave, and Elmar Nöth. Convolutional neural networks and a transfer learning strategy to classify

- parkinson's disease from speech in three different languages. In *Iberoamerican Congress on Pattern Recognition*, pages 697–706, Havana, October 2019. Springer. doi:10.1007/978-3-030-33904-3\_66.
- [9] Akane Sano, Weixuan Chen, Daniel Lopez-Martinez, Sara Taylor, and Rosalind W Picard. Multimodal ambulatory sleep detection using lstm recurrent neural networks. *IEEE journal of biomedical and health informatics*, 23(4):1607–1617, 2018. doi:10.1109/JBHI.2018.2867619.
- [10] Patricia Amado-Caballero, Pablo Casaseca-de-la Higuera, Susana Alberola-Lopez, Jesus Maria Andres-de Llano, Jose Antonio Lopez-Villalobos, Jose Ramon Garmendia-Leiza, and Carlos Alberola-Lopez. Objective adhd diagnosis using convolutional neural networks over daily-life activity records. *IEEE Journal of Biomedical and Health Informatics*, 24(9):2690 – 2700, 2020. doi:10.1109/JBHI.2020.2964072.
- [11] Abhay Koushik, Judith Amores, and Pattie Maes. Real-time sleep staging using deep learning on a smartphone for a wearable eeg. *arXiv preprint arXiv:1811.10111*, 2018. doi:10.48550/arXiv.1811.10111.
- [12] Facial landmark extraction. <https://pypi.org/project/face-recognition/>. Accessed: 2020-11-23.
- [13] **Justyna Skibińska** and Radim Burget. The transferable methodologies of detection sleep disorders thanks to the actigraphy device for parkinson's disease detection. volume 2880, Tampere, June 2021. CEUR Workshop Proceedings, CEUR-WS.
- [14] Jiri Mekyska, Zoltan Galaz, Tomas Kiska, Vojtech Zvoncak, Jan Mucha, Zdenek Smekal, Ilona Eliasova, Milena Kostalova, Martina Mrackova, Dagmar Fiedorova, et al. Quantitative analysis of relationship between hypokinetic dysarthria and the freezing of gait in parkinson's disease. *Cognitive computation*, 10(6):1006–1018, 2018. doi:10.1007/s12559-018-9575-8.
- [15] *Global Europe 2050*, 2020 (Accessed 2020-08-18). URL: [https://ec.europa.eu/research/social-sciences/pdf/policy\\_reviews/global-europe-2050-report\\_en.pdf](https://ec.europa.eu/research/social-sciences/pdf/policy_reviews/global-europe-2050-report_en.pdf).
- [16] Marco Ciotti, Massimo Ciccozzi, Alessandro Terrinoni, Wen-Can Jiang, Cheng-Bin Wang, and Sergio Bernardini. The covid-19 pandemic. *Critical reviews in clinical laboratory sciences*, 57(6):365–388, 2020. doi:10.1080/10408363.2020.1783198.

- [17] Elham Sheikhzadeh, Shimaa Eissa, Aziah Ismail, and Mohammed Zourob. Diagnostic techniques for covid-19 and new developments. *Talanta*, 220:121392, 2020. doi:10.1016/j.talanta.2020.121392.
- [18] Bastiaan R Bloem, Emily J Henderson, E Ray Dorsey, Michael S Okun, Njideka Okubadejo, Piu Chan, John Andrejack, Sirwan KL Darweesh, and Marten Munneke. Integrated and patient-centred management of parkinson’s disease: a network model for reshaping chronic neurological care. *The Lancet Neurology*, 19(7):623 – 634, 2020. doi:10.1016/S1474-4422(20)30064-8.
- [19] Alberto J Espay, Paolo Bonato, Fatta B Nahab, Walter Maetzler, John M Dean, Jochen Klucken, Bjoern M Eskofier, Aristide Merola, Fay Horak, Anthony E Lang, et al. Technology in parkinson’s disease: challenges and opportunities. *Movement Disorders*, 31(9):1272–1282, 2016. doi:10.1002/mds.26642.
- [20] Asma Channa, Nirvana Popescu, **Justyna Skibińska**, and Radim Burget. The rise of wearable devices during the covid-19 pandemic: A systematic review. *Sensors*, 21(17):5787, 2021. doi:10.3390/s21175787.
- [21] Suranga Seneviratne, Yining Hu, Tham Nguyen, Guohao Lan, Sara Khalifa, Kanchana Thilakarathna, Mahbub Hassan, and Aruna Seneviratne. A survey of wearable devices and challenges. *IEEE Communications Surveys & Tutorials*, 19(4):2573–2620, 2017. doi:10.1109/COMST.2017.2731979.
- [22] Rozita Jamili Oskouei, Zahra MousaviLou, Zohreh Bakhtiari, and Khuda Bux Jalbani. Iot-based healthcare support system for alzheimer’s patients. *Wireless Communications and Mobile Computing*, 2020:1–15, 2020. doi:10.1155/2020/8822598.
- [23] Erico Tjoa and Cuntai Guan. A survey on explainable artificial intelligence (xai): Toward medical xai. *IEEE transactions on neural networks and learning systems*, 32(11):4793–4813, 2020. doi:10.1109/TNNLS.2020.3027314.
- [24] Helen Bridgman, Man Ting Kwong, and Jeroen HM Bergmann. Mechanical safety of embedded electronics for in-body wearables: A smart mouthguard study. *Annals of biomedical engineering*, 47(8):1725–1737, 2019. doi:10.1007/s10439-019-02267-4.
- [25] Yulia Silina and Hamed Haddadi. New directions in jewelry: a close look at emerging trends & developments in jewelry-like wearable devices. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers*,

- pages 49–56, Osaka, September 2015. Association for Computing Machinery, Inc. doi:10.1145/2802083.2808410.
- [26] Aleksandr Ometov, Viktoriia Shubina, Lucie Klus, **Justyna Skibińska**, Salwa Saafi, Pavel Pascacio, Laura Flueratoru, Darwin Quezada Gaibor, Nadezhda Chukhno, Olga Chukhno, et al. A survey on wearable technology: History, state-of-the-art and current challenges. *Computer Networks*, 193:108074, 2021. doi:10.1016/j.comnet.2021.108074.
- [27] James Hayward. *Wearable Sensors 2018-2028: Technologies, Markets & Players*. IDTechEx Ltd., 2018.
- [28] Grand View Research. Wearable technology market size, share & trends analysis report by product (wrist-wear, eye-wear & head-wear, foot-wear, neck-wear, body-wear), by application, by region, and segment forecasts, 2020–2027, 2020.
- [29] Market Research Reports. Wearable healthcare devices market. <https://www.futuremarketinsights.com/reports/wearable-healthcare-devices-market>, 2022. Accessed: 2023-02-28.
- [30] He Li, Jing Wu, Yiwen Gao, and Yao Shi. Examining individuals’ adoption of healthcare wearable devices: An empirical study from privacy calculus perspective. *International journal of medical informatics*, 88:8–17, 2016. doi:10.1016/j.ijmedinf.2015.12.010.
- [31] Asma Channa, Nirvana Popescu, et al. Managing covid-19 global pandemic with high-tech consumer wearables: A comprehensive review. In *2020 12th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, volume 2020-October, pages 222–228. IEEE, IEEE Computer Society, October 2020. doi:10.1109/ICUMT51630.2020.9222428.
- [32] Mostafa Al-Emran and Jesse M Ehrenfeld. Breaking out of the box: Wearable technology applications for detecting the spread of covid-19. *Journal of Medical Systems*, 45(2):1–2, 2021. doi:10.1007/s10916-020-01697-1.
- [33] Vladimir Tomberg, Trenton Schulz, and Sebastian Kelle. Applying universal design principles to themes for wearables. In *Universal Access in Human-Computer Interaction. Access to Interaction: 9th International Conference, UAHCI 2015, Held as Part of HCI International 2015, Los Angeles, CA, USA, August 2-7, 2015, Proceedings, Part II 9*, volume 9176, pages 550–560, Los Angeles, August 2015. Springer Verlag. doi:10.1007/978-3-319-20681-3\_52.

- [34] Molly Follette Story, James L Mueller, and Ronald L Mace. The universal design file: Designing for people of all ages and abilities. 1998.
- [35] M Terzi, A Cenedese, and GA Susto. A Multivariate Symbolic Approach to Activity Recognition for Wearable Applications. *IFAC-PapersOnLine*, 50(1):15865–15870, 2017. doi:10.1016/j.ifacol.2017.08.2333.
- [36] Rachel E Stirling, Mark J Cook, David B Grayden, and Philippa J Karoly. Seizure forecasting and cyclic control of seizures. *Epilepsia*, 62(S1):S2 – S14, 2021. doi:10.1111/epi.16541.
- [37] Danilo Bzdok, Martin Krzywinski, and Naomi Altman. Machine learning: supervised methods. *Nature methods*, 15(1):5, 2018. doi:10.1038/nmeth.4551.
- [38] Mirza Mansoor Baig, Hamid GholamHosseini, Aasia A Moqem, Farhaan Mirza, and Maria Lindén. A systematic review of wearable patient monitoring systems—current challenges and opportunities for clinical adoption. *Journal of medical systems*, 41(7):115, 2017. doi:10.1007/s10916-017-0760-1.
- [39] SB Kotsiantis, Dimitris Kanellopoulos, and PE Pintelas. Data preprocessing for supervised leaning. *International Journal of Computer Science*, 1(2):111–117, 2006.
- [40] Suresh Yaram. Machine learning algorithms for document clustering and fraud detection. In *2016 International Conference on Data Science and Engineering (ICDSE)*, pages 1–6, Cochin, August 2016. IEEE. doi:10.1109/ICDSE.2016.7823950.
- [41] Jinxi Wang, Bo Hu, Xiang Li, and Zhe Yang. Gtc forest: An ensemble method for network structured data classification. In *2018 14th International Conference on Mobile Ad-Hoc and Sensor Networks (MSN)*, pages 81–85, Shenyang, December 2018. IEEE. doi:10.1109/MSN.2018.00020.
- [42] Baohua Sun, Lin Yang, Wenhan Zhang, Michael Lin, Patrick Dong, Charles Young, and Jason Dong. Supertml: Two-dimensional word embedding for the precognition on structured tabular data. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, page 2973 – 2981, Long Beach, June 2019. IEEE Computer Society. doi:10.1109/CVPRW.2019.00360.
- [43] Saurav Verma, Khushboo Jain, and Chetana Prakash. An unstructured to structured data conversion using machine learning algorithm in internet of things (iot). Available at SSRN 3563389, 2020. doi:10.2139/ssrn.3563389.

- [44] Zawar Hussain, Quan Z Sheng, and Wei Emma Zhang. A Review and Categorization of Techniques on Device-Free Human Activity Recognition. *Journal of Network and Computer Applications*, 167:102738, 2020. doi:10.1016/j.jnca.2020.102738.
- [45] L Minh Dang, Kyungbok Min, Hanxiang Wang, Md Jalil Piran, Cheol Hee Lee, and Hyeonjoon Moon. Sensor-Based and Vision-Based Human Activity Recognition: A Comprehensive Survey. *Pattern Recognition*, 108:107561, 2020. doi:10.1016/j.patcog.2020.107561.
- [46] Scott Pardoel, Jonathan Kofman, Julie Nantel, and Edward D Lemaire. Wearable-Sensor-Based Detection and Prediction of Freezing of Gait in Parkinson’s Disease: A Review. *Sensors*, 19(23):5141, 2019. doi:10.3390/s19235141.
- [47] Bjoern M Eskofier, Sunghoon I Lee, Jean-Francois Daneault, Fatemeh N Golabchi, Gabriela Ferreira-Carvalho, Gloria Vergara-Diaz, Stefano Sapienza, Gianluca Costante, Jochen Klucken, Thomas Kautz, et al. Recent Machine Learning Advancements in Sensor-Based Mobility Analysis: Deep Learning for Parkinson’s Disease Assessment. In *Proc. of 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 655–658, Orlando, August 2016. IEEE. doi:10.1109/EMBC.2016.7590787.
- [48] Muhan Zhang, Zhicheng Cui, Marion Neumann, and Yixin Chen. An End-to-End Deep Learning Architecture for Graph Classification. In *Proc. of 32nd Association for the Advancement of Artificial Intelligence (AAAI) Conference on Artificial Intelligence*, volume 32, New Orleans, April 2018. PKP Publishing Services Network. doi:10.1609/aaai.v32i1.11782.
- [49] Francisco Javier Ordóñez and Daniel Roggen. Deep Convolutional and lstm Recurrent Neural Networks for Multimodal Wearable Activity Recognition. *Sensors*, 16(1):115, 2016. doi:10.3390/s16010115.
- [50] Levent Eren, Turker Ince, and Serkan Kiranyaz. A Generic Intelligent Bearing Fault Diagnosis System Using Compact Adaptive 1D CNN Classifier. *Journal of Signal Processing Systems*, 91(2):179–189, 2019. doi:10.1007/s11265-018-1378-3.
- [51] Santosh Kumar, Wendy J Nilsen, Amy Abernethy, Audie Atienza, Kevin Patrick, Misha Pavel, William T Riley, Albert Shar, Bonnie Spring, Donna Spruijt-Metz, et al. Mobile health technology evaluation: the mhealth evidence workshop. *American journal of preventive medicine*, 45(2):228–236, 2013. doi:10.1016/j.amepre.2013.03.017.

- [52] Fadi Al-Turjman, Muhammad Hassan Nawaz, and Umit Deniz Ulusar. Intelligence in the internet of medical things era: A systematic review of current and future trends. *Computer Communications*, 150:644–660, 2020. doi:10.1016/j.comcom.2019.12.030.
- [53] Alan Godfrey. Wearables for independent living in older adults: Gait and falls. *Maturitas*, 100:16–26, 2017. doi:10.1016/j.maturitas.2017.03.317.
- [54] Shwetambara Malwade, Shabbir Syed Abdul, Mohy Uddin, Aldilas Achmad Nursetyo, Luis Fernandez-Luque, Xinxin Katie Zhu, Liezel Cilliers, Chun-Por Wong, Panagiotis Bamidis, and Yu-Chuan Jack Li. Mobile and wearable technologies in healthcare for the ageing population. *Computer methods and programs in biomedicine*, 161:233–237, 2018. doi:10.1016/j.cmpb.2018.04.026.
- [55] Donghyeog Choi, Hyunchul Choi, and Donghwa Shon. Future changes to smart home based on aal healthcare service. *Journal of Asian Architecture and Building Engineering*, 18(3):190–199, 2019. doi:10.1080/13467581.2019.1617718.
- [56] Liyakathunisa Syed, Saima Jabeen, S Manimala, and Abdullah Alsaedi. Smart healthcare framework for ambient assisted living using iomt and big data analytics techniques. *Future Generation Computer Systems*, 101:136–151, 2019. doi:10.1016/j.future.2019.06.004.
- [57] Sebastian Heinzl, Daniela Berg, Thomas Gasser, Honglei Chen, Chun Yao, Ronald B Postuma, and MDS Task Force on the Definition of Parkinson’s Disease. Update of the MDS research criteria for prodromal parkinson’s disease. *Movement Disorders*, 34(10):1464–1470, 2019. doi:10.1002/mds.27802.
- [58] Giulio Marchesini, Francesca Marchignoli, and Salvatore Petta. Evidence-based medicine and the problem of healthy volunteers. *Annals of hepatology*, 16(6):832–834, 2017. doi:10.5604/01.3001.0010.5272.
- [59] Tejaswini Mishra, Meng Wang, Ahmed A Metwally, Gireesh K Bogu, Andrew W Brooks, Amir Bahmani, Arash Alavi, Alessandra Celli, Emily Higgs, Orit Dagan-Rosenfeld, et al. Pre-symptomatic detection of covid-19 from smartwatch data. *Nature Biomedical Engineering*, 4(12):1208–1220, 2020. doi:10.1038/s41551-020-00640-6.
- [60] **Justyna Skibińska** and Asma Channa. The impact of wearables in improving silver economy. <https://projects.tuni.fi/a-wear/news/>

- the-impact-of-wearables-in-improving-silver-economy/ [Accessed 2021-10-18], 2020.
- [61] **Justyna Skibińska** and Radim Burget. Is it possible to distinguish covid-19 cases and influenza with wearable devices? analysis with machine learning. *Journal of Advances in Information Technology*, 13(3):265–270, 2022. doi:10.12720/jait.13.3.265-270.
- [62] **Justyna Skibińska**, Jiri Hosek, and Asma Channa. Wearable analytics and early diagnostic of covid-19 based on two cohorts. In *2022 14th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, volume 2022-October, pages 56–63, Valencia, October 2022. IEEE. doi:10.1109/ICUMT57764.2022.9943460.
- [63] **Justyna Skibińska** and Radim Burget. Parkinson’s disease detection based on changes of emotions during speech. In *2020 12th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, volume 2020-October, pages 124–130, Brno, October 2020. IEEE, IEEE Computer Society. doi:10.1109/ICUMT51630.2020.9222446.
- [64] Thirumalaisamy P Velavan and Christian G Meyer. The covid-19 epidemic. *Tropical medicine & international health*, 25(3):278, 2020. doi:10.1111/tmi.13383.
- [65] Cristina Menni, Ana M Valdes, Maxim B Freidin, Carole H Sudre, Long H Nguyen, David A Drew, Sajaysurya Ganesh, Thomas Varsavsky, M Jorge Cardoso, Julia S El-Sayed Moustafa, et al. Real-time tracking of self-reported symptoms to predict potential covid-19. *Nature medicine*, 26(7):1037–1040, 2020. doi:10.1038/s41591-020-0916-2.
- [66] Thomas Struyf, Jonathan J Deeks, Jacqueline Dinnes, Yemisi Takwoingi, Clare Davenport, Mariska Mg Leeflang, René Spijker, Lotty Hooft, Devy Emperador, Sabine Dittrich, et al. Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has covid-19 disease. *Cochrane Database of Systematic Reviews*, 2020(7), 2020. doi:10.1002/14651858.CD013665.
- [67] A Pryce-Roberts, M Talaei, and NP Robertson. Neurological complications of covid-19: a preliminary review. *Journal of Neurology*, 267:1870–1873, 2020. doi:10.1007/s00415-020-09941-x.



- [68] Shreyasi Gupta and Arkadeep Mitra. Challenge of post-covid era: management of cardiovascular complications in asymptomatic carriers of sars-cov-2. *Heart failure reviews*, 27(1):239 – 249, 2021. doi:10.1007/s10741-021-10076-y.
- [69] Emily Fraser. Long term respiratory complications of covid-19. *Bmj*, 370, 2020. doi:10.1136/bmj.m3001.
- [70] Qun Li, Xuhua Guan, Peng Wu, Xiaoye Wang, Lei Zhou, Yeqing Tong, Ruiqi Ren, Kathy SM Leung, Eric HY Lau, Jessica Y Wong, et al. Early transmission dynamics in wuhan, china, of novel coronavirus-infected pneumonia. *New England journal of medicine*, 382(13):1199 – 1207, 2020. doi:10.1056/NEJMoa2001316.
- [71] Xi He, Eric HY Lau, Peng Wu, Xilong Deng, Jian Wang, Xinxin Hao, Yiu Chung Lau, Jessica Y Wong, Yujuan Guan, Xinghua Tan, et al. Temporal dynamics in viral shedding and transmissibility of covid-19. *Nature medicine*, 26(5):672–675, 2020. doi:10.1038/s41591-020-0869-5.
- [72] Aravind Natarajan, Hao-Wei Su, and Conor Heneghan. Assessment of physiological signs associated with covid-19 measured using wearable devices. *NPJ digital medicine*, 3(1):1–8, 2020. doi:10.1038/s41746-020-00363-7.
- [73] W Joost Wiersinga, Andrew Rhodes, Allen C Cheng, Sharon J Peacock, and Hallie C Prescott. Pathophysiology, transmission, diagnosis, and treatment of coronavirus disease 2019 (covid-19): a review. *Jama*, 324(8):782–793, 2020. doi:10.1001/jama.2020.12839.
- [74] Rakesh Chandra Joshi, Saumya Yadav, Vinay Kumar Pathak, Hardeep Singh Malhotra, Harsh Vardhan Singh Khokhar, Anit Parihar, Neera Kohli, D Himanshu, Ravindra K Garg, Madan Lal Brahma Bhatt, et al. A deep learning-based covid-19 automatic diagnostic framework using chest x-ray images. *Biocybernetics and Biomedical Engineering*, 41(1):239 – 254, 2021. doi:10.1016/j.bbe.2021.01.002.
- [75] Dhruv R Seshadri, Evan V Davies, Ethan R Harlow, Jeffrey J Hsu, Shanina C Knighton, Timothy A Walker, James E Voos, and Colin K Drummond. Wearable sensors for covid-19: a call to action to harness our digital infrastructure for remote patient monitoring and virtual assessments. *Frontiers in Digital Health*, 2:8, 2020. doi:10.3389/fdgth.2020.00008.
- [76] Rajalakshmi Krishnamurthi, Dhanalekshmi Gopinathan, and Adarsh Kumar. Wearable devices and covid-19: state of the art, framework, and challenges.

- Emerging Technologies for Battling Covid-19*, 324:157–180, 2021. doi:10.1007/978-3-030-60039-6\_8.
- [77] Werner Poewe, Klaus Seppi, Caroline M Tanner, Glenda M Halliday, Patrik Brundin, Jens Volkman, Anette-Eleonore Schrag, and Anthony E Lang. Parkinson disease. *Nature reviews Disease primers*, 3(1):1–21, 2017. doi:10.1038/nrdp.2017.13.
- [78] Jan Stochl, Anne Boomsma, Evzen Ruzicka, Hana Brozova, and Petr Blahus. On the structure of motor symptoms of parkinson’s disease. *Movement Disorders*, 23(9):1307–1312, 2008. doi:10.1002/mds.22029.
- [79] Ronald B Postuma, Daniela Berg, Matthew Stern, Werner Poewe, C Warren Olanow, Wolfgang Oertel, José Obeso, Kenneth Marek, Irene Litvan, Anthony E Lang, et al. MDS clinical diagnostic criteria for parkinson’s disease. *Movement disorders*, 30(12):1591–1601, 2015. doi:10.1002/mds.26424.
- [80] Joseph R Duffy. *Motor Speech Disorders E-Book: Substrates, Differential Diagnosis, and Management*. Elsevier Health Sciences, 2019.
- [81] JG Kalf, BJM De Swart, BR Bloem, and Marten Munneke. Prevalence of oropharyngeal dysphagia in Parkinson’s disease: a meta-analysis. *Parkinsonism & related disorders*, 18(4):311–315, 2012. doi:10.1016/j.parkreldis.2011.11.006.
- [82] L Ricciardi, A De Angelis, L Marsili, I Faiman, P Pradhan, EA Pereira, MJ Edwards, F Morgante, and M Bologna. Hypomimia in parkinson’s disease: an axial sign responsive to levodopa. *European Journal of Neurology*, 27(12):2422 – 2429, 2020. doi:10.1111/ene.14452.
- [83] Jan Mucha, Jiri Mekyska, Zoltan Galaz, Marcos Faundez-Zanuy, Karnele Lopez-de Ipina, Vojtech Zvoncak, Tomas Kiska, Zdenek Smekal, Lubos Brabenec, and Irena Rektorova. Identification and monitoring of parkinson’s disease dysgraphia based on fractional-order derivatives of online handwriting. *Applied Sciences*, 8(12):2566, 2018. doi:10.3390/app8122566.
- [84] Luboš Brabenec, Jiří Mekyska, Z Galaz, and Irena Rektorova. Speech disorders in Parkinson’s disease: early diagnostics and effects of medication and brain stimulation. *Journal of neural transmission*, 124(3):303–334, 2017. doi:10.1007/s00702-017-1676-0.

- [85] Claudio De Stefano, Francesco Fontanella, Donato Impedovo, Giuseppe Pirlo, and Alessandra Scotto di Freca. Handwriting analysis to support neurodegenerative diseases diagnosis: A review. *Pattern Recognition Letters*, 121:37–45, 2019. doi:10.1016/j.patrec.2018.05.013.
- [86] Ahmed A Moustafa, Srinivasa Chakravarthy, Joseph R Phillips, Ankur Gupta, Szabolcs Keri, Bertalan Polner, Michael J Frank, and Marjan Jahanshahi. Motor symptoms in parkinson’s disease: A unified framework. *Neuroscience & Biobehavioral Reviews*, 68:727–740, 2016. doi:10.1016/j.neubiorev.2016.07.010.
- [87] Sigurlaug Sveinbjornsdottir. The clinical symptoms of parkinson’s disease. *Journal of neurochemistry*, 139:318–324, 2016. doi:10.1111/jnc.13691.
- [88] Andrea Bandini, Silvia Orlandi, Hugo Jair Escalante, Fabio Giovannelli, Massimo Cincotta, Carlos A Reyes-Garcia, Paola Vanni, Gaetano Zaccara, and Claudia Manfredi. Analysis of facial expressions in parkinson’s disease through video-based automatic methods. *Journal of neuroscience methods*, 281:7–20, 2017. doi:10.1016/j.jneumeth.2017.02.006.
- [89] Pedro Gómez Vilda, Jiri Mekyska, Andrés Gómez Rodellar, Daniel Palacios Alonso, V Rodellar Biarge, and Agustin Alvarez Marquina. Monitoring parkinson disease from speech articulation kinematics. *Loquens: revista española de ciencias del habla*, (4):2, 2017. doi:10.3989/loquens.2017.036.
- [90] Ronald F Pfeiffer. Non-motor symptoms in parkinson’s disease. *Parkinsonism & related disorders*, 22:S119–S122, 2016. doi:10.1016/j.parkreldis.2015.09.004.
- [91] R Prashanth and Sumantra Dutta Roy. Early detection of parkinson’s disease through patient questionnaire and predictive modelling. *International journal of medical informatics*, 119:75–87, 2018. doi:10.1016/j.ijmedinf.2018.09.008.
- [92] Ziad Obermeyer and Ezekiel J Emanuel. Predicting the future—big data, machine learning, and clinical medicine. *The New England journal of medicine*, 375(13):1216, 2016. doi:10.1056/NEJMp1606181.
- [93] Barbara S Connolly and Anthony E Lang. Pharmacological treatment of parkinson disease: a review. *Jama*, 311(16):1670–1683, 2014. doi:10.1001/jama.2014.3654.

- [94] Jeff M Bronstein, Michele Tagliati, Ron L Alterman, Andres M Lozano, Jens Volkmann, Alessandro Stefani, Fay B Horak, Michael S Okun, Kelly D Foote, Paul Krack, et al. Deep brain stimulation for parkinson disease: an expert consensus and review of key issues. *Archives of neurology*, 68(2):165–165, 2011. doi:10.1001/archneurol.2010.260.
- [95] Bo Mohr Morberg, Anne Sofie Malling, Bente Rona Jensen, Ole Gredal, Lene Wermuth, and Per Bech. The hawthorne effect as a pre-placebo expectation in Parkinsons disease patients participating in a randomized placebo-controlled clinical study. *Nordic journal of psychiatry*, 72(6):442–446, 2018. doi:10.1080/08039488.2018.1468480.
- [96] Amir Hossein Poorjam, Mathew Shaji Kavalekalam, Liming Shi, Jordan P Raykov, Jesper Rindom Jensen, Max A Little, and Mads Græsbøll Christensen. Automatic quality control and enhancement for voice-based remote Parkinson’s disease detection. *Speech Communication*, 127:1–16, 2021. doi:10.1016/j.specom.2020.12.007.
- [97] Jan Rusz, Jan Hlavnička, Tereza Tykalová, Michal Novotný, Petr Dušek, Karel Šonka, and Evžen Ržička. Smartphone allows capture of speech abnormalities associated with high risk of developing parkinson’s disease. *IEEE transactions on neural systems and rehabilitation engineering*, 26(8):1495–1507, 2018. doi:10.1109/TNSRE.2018.2851787.
- [98] Athanasios Tsanas, Max A Little, and Lorraine O Ramig. Remote assessment of Parkinson’s disease symptom severity using the simulated cellular mobile telephone network. *IEEE Access*, 9:11024 – 11036, 2021. doi:10.1109/ACCESS.2021.3050524.
- [99] Juan Rafael Orozco-Arroyave, Juan Camilo Vásquez-Correa, Philipp Klumpp, Paula Andrea Pérez-Toro, Daniel Escobar-Grisales, Nils Roth, Cristian David Ríos-Urrego, Martin Strauss, Helber Andrés Carvajal-Castaño, Sebastian Bayerl, et al. Apkinson: the smartphone application for telemonitoring Parkinson’s patients through speech, gait and hands movement. *Neurodegenerative Disease Management*, 10(3):137–157, 2020. doi:10.2217/nmt-2019-0037.
- [100] Oliver Y Chén, Florian Lipsmeier, Huy Phan, John Prince, Kirsten I Taylor, Christian Gossens, Michael Lindemann, and Maarten De Vos. Building a machine-learning framework to remotely assess Parkinson’s disease using smartphones. *IEEE Transactions on Biomedical Engineering*, 67(12):3491–3500, 2020. doi:10.1109/TBME.2020.2988942.

- [101] The early stage detection - definition. <https://www.online-medical-dictionary.org/definitions-e/early-diagnosis.html> [Accessed 2022-01-18], April 2021.
- [102] John Noel Victorino, Yuko Shibata, Sozo Inoue, and Tomohiro Shibata. Predicting wearing-off of parkinson's disease patients using a wrist-worn fitness tracker and a smartphone: A case study. *Applied Sciences*, 11(16):7354, 2021. doi:10.3390/app11167354.
- [103] Joseph Jankovic. Parkinson's disease: clinical features and diagnosis. *Journal of neurology, neurosurgery & psychiatry*, 79(4):368–376, 2008. doi:10.1136/jnnp.2007.131045.
- [104] Martin Rajnoha, Jiri Mekyska, Radim Burget, Ilona Eliasova, Milena Kostalova, and Irena Rektorova. Towards identification of hypomimia in parkinson's disease based on face recognition methods. In *2018 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, volume 2018-November, pages 1–4, Moscow, November 2018. IEEE. doi:10.1109/ICUMT.2018.8631249.
- [105] Pedro Gómez-Vilda, Jiri Mekyska, José M Ferrández, Daniel Palacios-Alonso, Andrés Gómez-Rodellar, Victoria Rodellar-Biarge, Zoltan Galaz, Zdenek Smekal, Ilona Eliasova, Milena Kostalova, et al. Parkinson disease detection from speech articulation neuromechanics. *Frontiers in neuroinformatics*, 11:56, 2017. doi:10.3389/fninf.2017.00056.
- [106] Seyed-Mohammad Fereshtehnejad, Örjan Skogar, and Johan Lökk. Evolution of orofacial symptoms and disease progression in idiopathic parkinson's disease: Longitudinal data from the jönköping parkinson registry. *Parkinson's disease*, 2017, 2017. doi:10.1155/2017/7802819.
- [107] Gwenda Simons, Marcia C SMITH Pasqualini, Vasudevi Reddy, and Julia Wood. Emotional and nonemotional facial expressions in people with parkinson's disease. *Journal of the International Neuropsychological Society*, 10(4):521–535, 2004. doi:10.1017/S135561770410413X.
- [108] Akshada Shinde, Rashmi Atre, Anchal Singh Guleria, Radhika Nibandhe, and Revati Shriram. Facial features based prediction of parkinson's disease. In *2018 3rd International Conference for Convergence in Technology (I2CT)*, pages 1–5, Pune, Pune 2018. IEEE. doi:10.1109/I2CT.2018.8529466.
- [109] Athina Grammatikopoulou, Nikos Grammalidis, Sevasti Bostantjopoulou, and Zoe Katsarou. Detecting hypomimia symptoms by selfie photo analysis: for

- early parkinson disease detection. In *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, pages 517–522, Rhodes, June 2019. Association for Computing Machinery. doi:10.1145/3316782.3322756.
- [110] Dawn Bowers, Kimberly Miller, Wendelyn Bosch, Didem Gokcay, Otto Pedraza, Utaka Springer, and Michael Okun. Faces of emotion in parkinsons disease: micro-expressivity and bradykinesia during voluntary facial expressions. *Journal of the International Neuropsychological Society: JINS*, 12(6):765, 2006. doi:10.1017/S135561770606111X.
- [111] Lucia Ricciardi, Federica Visco-Comandini, Roberto Erro, Francesca Morgante, Matteo Bologna, Alfonso Fasano, Diego Ricciardi, Mark J Edwards, and James Kilner. Facial emotion recognition and expression in parkinson’s disease: an emotional mirror mechanism? *PLoS One*, 12(1):e0169110, 2017. doi:10.1371/journal.pone.0169110.
- [112] Dag Aarsland, Byron Creese, Marios Politis, K Ray Chaudhuri, Daniel Weintraub, Clive Ballard, et al. Cognitive decline in parkinson disease. *Nature Reviews Neurology*, 13(4):217–231, 2017. doi:10.1038/nrneuro1.2017.27.
- [113] Ulupi Sitoresmi et al. Tongue twisters in pronunciation class. In *International Conference on Teacher Training and Education*, volume 1, Surakarta, November 2016. Sebelas Maret University.
- [114] Jan Hlavnička, Roman Čmejla, Tereza Tykalová, Karel Šonka, Evžen Ržička, and Jan Rusz. Automated analysis of connected speech reveals early biomarkers of Parkinson’s disease in patients with rapid eye movement sleep behaviour disorder. *Scientific reports*, 7(1):1–13, 2017. doi:10.1038/s41598-017-00047-5.
- [115] Jan Rusz, Jan Hlavnička, Tereza Tykalová, Jitka Bušková, Olga Ulmanová, Evžen Ržička, and Karel Šonka. Quantitative assessment of motor speech abnormalities in idiopathic rapid eye movement sleep behaviour disorder. *Sleep medicine*, 19:141–147, 2016. doi:10.1016/j.sleep.2015.07.030.
- [116] Aileen K Ho, Robert Iansek, Caterina Marigliani, John L Bradshaw, and Sandra Gates. Speech impairment in a large sample of patients with Parkinson’s disease. *Behavioural neurology*, 11(3):131–137, 1998.
- [117] Caroline Moreau and Serge Pinto. Misconceptions about speech impairment in parkinson’s disease. *Movement Disorders*, 34(10):1471–1475, 2019. doi:10.1002/mds.27791.

- [118] Laureano Moro-Velazquez and Najim Dehak. A review of the use of prosodic aspects of speech for the automatic detection and assessment of Parkinson's disease. In *Automatic Assessment of Parkinsonian Speech Workshop*, volume 1295, pages 42–59, Cambridge, September 2020. Springer Science and Business Media Deutschland GmbH. doi:10.1007/978-3-030-65654-6\_3.
- [119] Laureano Moro-Velazquez, Jorge A Gomez-Garcia, Julian D Arias-Londoño, Najim Dehak, and Juan I Godino-Llorente. Advances in Parkinson's disease detection and assessment using voice and speech: A review of the articulatory and phonatory aspects. *Biomedical Signal Processing and Control*, 66:102418, 2021. doi:10.1016/j.bspc.2021.102418.
- [120] Vineet Prasad and Cary A Brown. A pilot study to determine the consistency of simultaneous sleep actigraphy measurements comparing all four limbs of patients with parkinson disease. *Geriatrics*, 3(1):1, 2018. doi:10.3390/geriatrics3010001.
- [121] Alex Iranzo, José Luis Molinuevo, Joan Santamaría, Mónica Serradell, María José Martí, Francesc Valldeoriola, and Eduard Tolosa. Rapid-eye-movement sleep behaviour disorder as an early marker for a neurodegenerative disorder: a descriptive study. *The Lancet Neurology*, 5(7):572–577, 2006. doi:10.1016/S1474-4422(06)70476-8.
- [122] K Ray Chaudhuri and Yogini Naidu. Early parkinson's disease and non-motor issues. *Journal of neurology*, 255(5):33–38, 2008. doi:10.1007/s00415-008-5006-1.
- [123] Ignacio Perez-Pozuelo, Bing Zhai, Joao Palotti, Raghvendra Mall, Michaël Au-petit, Juan M Garcia-Gomez, Shahrad Taheri, Yu Guan, and Luis Fernandez-Luque. The future of sleep health: a data-driven revolution in sleep science and medicine. *NPJ digital medicine*, 3(1):1–15, 2020. doi:10.1038/s41746-020-0244-4.
- [124] Michael A Grandner and Mary E Rosenberger. Actigraphic sleep tracking and wearables: Historical context, scientific applications and guidelines, limitations, and considerations for commercial sleep devices. In *Sleep and health*, pages 147–157. Elsevier, 2019. doi:10.1016/B978-0-12-815373-4.00012-5.
- [125] Jessica Vensel Rundo and Ralph Downey III. Polysomnography. *Handbook of clinical neurology*, 160:381–392, 2019. doi:10.1016/B978-0-444-64032-1.00025-4.

- [126] Paul L Nunez, Ramesh Srinivasan, et al. *Electric fields of the brain: the neurophysics of EEG*. Oxford University Press, 2006. doi:10.1093/acprof:oso/9780195050387.001.0001.
- [127] Piotr Augustyniak. *Elektroniczna aparatura medyczna*. Wydawnictwo AGH, 2015.
- [128] Francesco Carlo Morabito, Maurizio Campolo, Cosimo Ieracitano, Javad Mohammad Ebadi, Lilla Bonanno, Alessia Bramanti, Simona Desalvo, Nadia Mammone, and Placido Bramanti. Deep convolutional neural networks for classification of mild cognitive impaired and alzheimer’s disease patients from scalp eeg recordings. In *2016 IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI)*, pages 1–6, Bologna, September 2016. IEEE. doi:10.1109/RTSI.2016.7740576.
- [129] Siavash Sakhavi and Cuntai Guan. Convolutional neural network-based transfer learning and knowledge distillation using multi-subject data in motor imagery bci. In *2017 8th International IEEE/EMBS Conference on Neural Engineering (NER)*, pages 588–591, Shanghai, May 2017. IEEE. doi:10.1109/NER.2017.8008420.
- [130] Sachin S Talathi. Deep recurrent neural networks for seizure detection and early seizure detection systems. *arXiv preprint arXiv:1706.03283*, 2017. doi:10.48550/arXiv.1706.03283.
- [131] Xiu An, Deping Kuang, Xiaojiao Guo, Yilu Zhao, and Lianghua He. A deep learning method for classification of eeg data based on motor imagery. In *International Conference on Intelligent Computing*, pages 203–210. Springer International Publishing, August 2014. doi:10.1007/978-3-319-09330-7\_25.
- [132] Kay Gregor Hartmann, Robin Tibor Schirrmester, and Tonio Ball. Eeg-gan: Generative adversarial networks for electroencephalographic (eeg) brain signals. *arXiv preprint arXiv:1806.01875*, 2018. doi:10.48550/arXiv.1806.01875.
- [133] Sławomir Opałka, Bartłomiej Stasiak, Dominik Szajerman, and Adam Wojciechowski. Multi-channel convolutional neural networks architecture feeding for effective eeg mental tasks classification. *Sensors*, 18(10):3451, 2018. doi:10.3390/s18103451.



- [134] Ye Yuan, Guangxu Xun, Kebin Jia, and Aidong Zhang. A multi-view deep learning framework for eeg seizure detection. *IEEE journal of biomedical and health informatics*, 23(1):83–94, 2018. doi:10.1109/JBHI.2018.2871678.
- [135] Yannick Roy, Hubert Banville, Isabela Albuquerque, Alexandre Gramfort, Tiago H Falk, and Jocelyn Faubert. Deep learning-based electroencephalography analysis: a systematic review. *Journal of neural engineering*, 16(5):051001, 2019. doi:10.1088/1741-2552/ab260c.
- [136] Farhan Fuad Abir, Khalid Alyafei, Muhammad EH Chowdhury, Amith Khandakar, Rashid Ahmed, Muhammad Maqsd Hossain, Sakib Mahmud, Ashiqur Rahman, Tareq O Abbas, Susu M Zughailer, et al. Pcovnet: A presymptomatic covid-19 detection framework using deep learning model using wearables data. *Computers in biology and medicine*, 147:105682, 2022. doi:10.1016/j.combiomed.2022.105682.
- [137] Hyeong Rae Cho, Jin Hyun Kim, Hye Rin Yoon, Yong Seop Han, Tae Seen Kang, Hyunju Choi, and Seunghwan Lee. Machine learning-based optimization of pre-symptomatic covid-19 detection through smartwatch. *Scientific Reports*, 12(1):1–15, 2022. doi:10.1038/s41598-022-11329-y.
- [138] Benjamin L Smarr, Kirstin Aschbacher, Sarah M Fisher, Anoushka Chowdhary, Stephan Dilchert, Karena Puldon, Adam Rao, Frederick M Hecht, and Ashley E Mason. Feasibility of continuous fever monitoring using wearable devices. *Scientific reports*, 10(1):1–11, 2020. doi:10.1038/s41598-020-78355-6.
- [139] Ashley E Mason, Frederick M Hecht, Shakti K Davis, Joseph L Natale, Wendy Hartogenesis, Natalie Damaso, Kajal T Claypool, Stephan Dilchert, Subhasis Dasgupta, Shweta Purawat, et al. Detection of covid-19 using multimodal data from a wearable device: results from the first tempredict study. *Scientific reports*, 12(1):1–15, 2022. doi:10.1038/s41598-022-07314-0.
- [140] Shayan Hassantabar, Novati Stefano, Vishweshwar Ghanakota, Alessandra Ferrari, Gregory N Nicola, Raffaele Bruno, Ignazio R Marino, Kenza Hamidouche, and Niraj K Jha. Coviddeep: Sars-cov-2/covid-19 test based on wearable medical sensors and efficient neural networks. *IEEE Transactions on Consumer Electronics*, 67(4):244–256, 2021. doi:10.1109/TCE.2021.3130228.
- [141] R Karthickraja, R Kumar, S Kirubakaran, R Manikandan, et al. Covid-19 prediction and symptom analysis using wearable sensors and iot. *International*

- Journal of Pervasive Computing and Communications*, 18(5):499 – 507, 2020. doi:10.1108/IJPCC-09-2020-0146.
- [142] Dean J Miller, John V Capodilupo, Michele Lastella, Charli Sargent, Gregory D Roach, Victoria H Lee, and Emily R Capodilupo. Analyzing changes in respiratory rate to predict the risk of covid-19 infection. *PloS one*, 15(12):e0243693, 2020. doi:10.1371/journal.pone.0243693.
- [143] Talha Burak Alakus and Ibrahim Turkoglu. Comparison of deep learning approaches to predict covid-19 infection. *Chaos, Solitons & Fractals*, 140:110120, 2020. doi:10.1016/j.chaos.2020.110120.
- [144] Dylan M Richards, MacKenzie J Tweardy, Steven R Steinhubl, David W Chestek, Terry L Vanden Hoek, Karen A Larimer, and Stephan W Wegerich. Wearable sensor derived decompensation index for continuous remote monitoring of covid-19 diagnosed patients. *NPJ digital medicine*, 4(1):1–11, 2021. doi:10.1038/s41746-021-00527-z.
- [145] Chun Ka Wong, Deborah Tip Yin Ho, Anthony Raymond Tam, Mi Zhou, Yuk Ming Lau, Milky Oi Yan Tang, Raymond Cheuk Fung Tong, Kuldeep Singh Rajput, Gengbo Chen, Soon Chee Chan, et al. Artificial intelligence mobile health platform for early detection of covid-19 in quarantine subjects using a wearable biosensor: protocol for a randomised controlled trial. *BMJ open*, 10(7):e038555, 2020. doi:10.1136/bmjopen-2020-038555.
- [146] Hyoyoung Jeong, John A Rogers, and Shuai Xu. Continuous on-body sensing for the covid-19 pandemic: Gaps and opportunities. *Science Advances*, 6(36):eabd4794, 2020. doi:10.1126/sciadv.abd4794.
- [147] Liang Pan, Cong Wang, Haoran Jin, Jie Li, Le Yang, Yuanjin Zheng, Yonggang Wen, Ban Hock Tan, Xian Jun Loh, and Xiaodong Chen. Lab-on-mask for remote respiratory monitoring. *ACS Materials Letters*, 2(9):1178–1181, 2020. doi:10.1021/acsmaterialslett.0c00299.
- [148] Harry J Davies, Ian Williams, Nicholas S Peters, and Danilo P Mandic. In-ear spo2: A tool for wearable, unobtrusive monitoring of hypoxaemia in covid-19. *Sensors*, 20(17):1 – 12, 2020. doi:10.3390/s20174879.
- [149] Radovan Stojanović, Andrej Škraba, and Budimir Lutovac. A headset like wearable device to track covid-19 symptoms. In *2020 9th Mediterranean Conference on Embedded Computing (MECO)*, pages 1–4, Budva, June 2020. IEEE. doi:10.1109/MECO49872.2020.9134211.

- [150] Jason S Chinitz, Rajat Goyal, Donna Chelle Morales, Melissa Harding, Samy Selim, and Laurence M Epstein. Use of a smartwatch for assessment of the qt interval in outpatients with coronavirus disease 2019. *The Journal of innovations in cardiac rhythm management*, 11(9):4219, 2020. doi:10.19102/icrm.2020.1100904.
- [151] Hasan Fleyeh and Jerker Westin. Extracting body landmarks from videos for parkinson gait analysis. In *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*, volume 2019-June, pages 379–384, Cordoba, June 2019. IEEE. doi:10.1109/CBMS.2019.00082.
- [152] Erika Rovini, Carlo Maremmani, and Filippo Cavallo. How wearable sensors can support parkinson’s disease diagnosis and treatment: a systematic review. *Frontiers in neuroscience*, 11:555, 2017. doi:10.3389/fnins.2017.00555.
- [153] Asma Channa, Giuseppe Ruggeri, Nadia Mammone, Rares-Cristian Ifrim, Antonio Iera, and Nirvana Popescu. Parkinson’s disease severity estimation using deep learning and cloud technology. In *2022 IEEE International Conference on Omni-layer Intelligent Systems (COINS)*, pages 1–7, Barcelona, August 2022. IEEE. doi:10.1109/COINS54846.2022.9854945.
- [154] Andrés Gómez-Rodellar, Agustín Álvarez-Marquina, Jiri Mekyska, Daniel Palacios-Alonso, Djamila Meghraoui, and Pedro Gómez-Vilda. Performance of articulation kinetic distributions vs mfccs in parkinson’s detection from vowel utterances. In *Neural Approaches to Dynamics of Signal Exchanges*, volume 151, pages 431–441. Springer, 2020. doi:10.1007/978-981-13-8950-4\_38.
- [155] Juan Camilo Vásquez-Correa, Tomas Arias-Vergara, Juan Rafael Orozco-Arroyave, Björn Eskofier, Jochen Klucken, and Elmar Nöth. Multimodal assessment of parkinson’s disease: a deep learning approach. *IEEE journal of biomedical and health informatics*, 23(4):1618–1630, 2018. doi:10.1109/JBHI.2018.2866873.
- [156] Wee Shin Lim, Shu-I Chiu, Meng-Ciao Wu, Shu-Fen Tsai, Pu-He Wang, Kun-Pei Lin, Yung-Ming Chen, Pei-Ling Peng, Yung-Yaw Chen, Jyh-Shing Roger Jang, et al. An integrated biometric voice and facial features for early detection of parkinson’s disease. *npj Parkinson’s Disease*, 8(1):145, 2022. doi:10.1038/s41531-022-00414-8.
- [157] John Archila, Antoine Manzanera, and Fabio Martínez. A multimodal parkinson quantification by fusing eye and gait motion patterns, using covariance

- descriptors, from non-invasive computer vision. *Computer Methods and Programs in Biomedicine*, 215:106607, 2022. doi:10.1016/j.cmpb.2021.106607.
- [158] M Katsikitis and I Pilowsky. A study of facial expression in parkinson’s disease using a novel microcomputer-based method. *Journal of Neurology, Neurosurgery & Psychiatry*, 51(3):362–366, 1988. doi:10.1136/jnnp.51.3.362.
- [159] L Ricciardi, A De Angelis, L Marsili, I Faiman, P Pradhan, EA Pereira, MJ Edwards, F Morgante, and M Bologna. Hypomimia in parkinson’s disease: an axial sign responsive to levodopa. *European Journal of Neurology*, 2020. doi:10.1111/ene.14452.
- [160] Nomi Vinokurov, David Arkadir, Eduard Linetsky, Hagai Bergman, and Daphna Weinshall. Quantifying hypomimia in parkinson patients using a depth camera. In *International Symposium on Pervasive Computing Paradigms for Mental Health*, volume 604, pages 63–71, Milan, September 2015. Springer Verlag. doi:10.1007/978-3-319-32270-4\_7.
- [161] Bhakti Sonawane and Priyanka Sharma. Review of automated emotion-based quantification of facial expression in parkinson’s patients. *environment*, 37(5):1151 – 1167. doi:10.1007/s00371-020-01859-9.
- [162] Peng Wu, Isabel Gonzalez, Georgios Patsis, Dongmei Jiang, Hichem Sahli, Eric Kerckhofs, and Marie Vandekerckhove. Objectifying facial expressivity assessment of parkinson’s patients: preliminary study. *Computational and mathematical methods in medicine*, 2014, 2014. doi:10.1155/2014/427826.
- [163] Parekh Payal and Mahesh M Goyani. A comprehensive study on face recognition: methods and challenges. *The Imaging Science Journal*, 68(2):114–127, 2020. doi:10.1080/13682199.2020.1738741.
- [164] Marcin Kolodziej, Andrzej Majkowski, Remigiusz J Rak, Pawel Tarnowski, and Tomasz Pielaszkiewicz. Analysis of facial features for the use of emotion recognition. In *19th International Conference Computational Problems of Electrical Engineering*, pages 1–4, Banska Stiavnica, September 2018. IEEE. doi:10.1109/CPEE.2018.8507137.
- [165] Michal Novotny, Tereza Tykalova, Hana Ruzickova, Evzen Ruzicka, Petr Dusek, and Jan Rusz. Automated video-based assessment of facial bradykinesia in de-novo parkinson’s disease. *NPJ digital medicine*, 5(1):1–8, 2022. doi:10.1038/s41746-022-00642-5.

- [166] Mohammad Rafayet Ali, Taylor Myers, Ellen Wagner, Harshil Ratnu, E Dorsey, and Ehsan Hoque. Facial expressions can detect parkinson’s disease: preliminary evidence from videos collected online. *NPJ digital medicine*, 4(1):1–4, 2021. doi:10.1038/s41746-021-00502-8.
- [167] Carlo Maremmani, Roberto Monastero, Giovanni Orlandi, Stefano Salvadori, Aldo Pieroni, Roberta Baschi, Alessandro Pecori, Cristina Dolciotti, Giulia Berchina, Erika Rovini, et al. Objective assessment of blinking and facial expressions in parkinson’s disease using a vertical electro-oculogram and facial surface electromyography. *Physiological measurement*, 40(6):065005, 2019. doi:10.1088/1361-6579/ab1c05.
- [168] Ge Su, Bo Lin, Wei Luo, Jianwei Yin, Shuiguang Deng, Honghao Gao, and Renjun Xu. Hypomimia recognition in parkinson’s disease with semantic features. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 17(3s):1–20, 2021. doi:10.1145/3476778.
- [169] Avner Abrami, Steven Gunzler, Camilla Kilbane, Rachel Ostrand, Bryan Ho, and Guillermo Cecchi. Automated computer vision assessment of hypomimia in parkinson disease: Proof-of-principle pilot study. *Journal of Medical Internet Research*, 23(2):e21037, 2021. doi:10.2196/21037.
- [170] Fer dataset. <https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data>, 2013 (Accessed August 13, 2020).
- [171] Abhinav Agrawal and Namita Mittal. Using cnn for facial expression recognition: a study of the effects of kernel size and number of filters on accuracy. *The Visual Computer*, 36(2):405–412, 2020. doi:10.1007/s00371-019-01630-9.
- [172] Yichuan Tang. Deep learning using linear support vector machines. *arXiv preprint arXiv:1306.0239*, 2013. doi:10.48550/arXiv.1306.0239.
- [173] Yanling Gan, Jingying Chen, and Luhui Xu. Facial expression recognition boosted by soft label with a diverse ensemble. *Pattern Recognition Letters*, 125:105–112, 2019. doi:10.1016/j.patrec.2019.04.002.
- [174] Ajjen Joshi, Linda Tickle-Degnen, Sarah Gunnery, Terry Ellis, and Margrit Betke. Predicting active facial expressivity in people with parkinson’s disease. In *Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, volume 29-June-2016, pages 1–4, Corfu, June 2016. Association for Computing Machinery. doi:10.1145/2910674.2910686.

- [175] George T Gitchel, Paul A Wetzal, and Mark S Baron. Pervasive ocular tremor in patients with parkinson disease. *Archives of neurology*, 69(8):1011–1017, 2012. doi:10.1001/archneurol.2012.70.
- [176] Jan Ruzs, Tereza Tykalova, Lorraine O Ramig, and Elina Tripoliti. Guidelines for speech recording and acoustic analyses in dysarthrias of movement disorders. *Movement Disorders*, 36(4):803 – 814, 2020. doi:10.1002/mds.28465.
- [177] Ina Kodrasi and Hervé Boudlard. Spectro-temporal sparsity characterization for dysarthric speech detection. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28:1210–1222, 2020. doi:10.1109/TASLP.2020.2985066.
- [178] Juan Rafael Orozco-Arroyave, Julián David Arias-Londoño, Jesús Francisco Vargas-Bonilla, María Claudia Gonzalez-Rátiva, and Elmar Nöth. New spanish speech corpus database for the analysis of people suffering from parkinson’s disease. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC’14)*, pages 342–347, Reykjavik, May 2014.
- [179] Laureano Moro-Velazquez, Jorge Andres Gomez-Garcia, Juan Ignacio Godino-Llorente, Jesús Villalba, Jan Ruzs, Stephanie Shattuck-Hufnagel, and Najim Dehak. A forced gaussians based methodology for the differential evaluation of parkinson’s disease by means of speech processing. *Biomedical Signal Processing and Control*, 48:205–220, 2019. doi:10.1016/j.bspc.2018.10.020.
- [180] Pedro Gómez, Jirí Mekyska, Andrés Gómez, D Palacios, Victoria Rodellar, and A Álvarez. Characterization of Parkinson’s disease dysarthria in terms of speech articulation kinematics. *Biomedical Signal Processing and Control*, 52:312–320, 2019. doi:10.1016/j.bspc.2019.04.029.
- [181] JI Godino-Llorente, S Shattuck-Hufnagel, JY Choi, L Moro-Velázquez, and JA Gómez-García. Towards the identification of idiopathic parkinson’s disease from the speech. new articulatory kinetic biomarkers. *PloS one*, 12(12):e0189583, 2017. doi:10.1371/journal.pone.0189583.
- [182] Riccardo Miotto, Fei Wang, Shuang Wang, Xiaoqian Jiang, and Joel T Dudley. Deep learning for healthcare: review, opportunities and challenges. *Briefings in bioinformatics*, 19(6):1236–1246, 2018. doi:10.1093/bib/bbx044.
- [183] Daniele Ravi, Charence Wong, Fani Deligianni, Melissa Berthelot, Javier Andreu-Perez, Benny Lo, and Guang-Zhong Yang. Deep learning for health informatics. *IEEE journal of biomedical and health informatics*, 21(1):4–21, 2016. doi:10.1109/JBHI.2016.2636665.

- [184] Juan Camilo Vásquez-Correa, Juan Rafael Orozco-Arroyave, and Elmar Nöth. Convolutional neural network to model articulation impairments in patients with parkinson’s disease. In *INTERSPEECH*, volume 2017-August, pages 314–318, Stockholm, August 2017. International Speech Communication Association. doi:10.21437/Interspeech.2017-1078.
- [185] Changqin Quan, Kang Ren, and Zhiwei Luo. A deep learning based method for parkinson’s disease detection using dynamic features of speech. *IEEE Access*, 9:10239–10252, 2021. doi:10.1109/ACCESS.2021.3051432.
- [186] Hakan Gunduz. Deep learning-based parkinson’s disease classification using vocal feature sets. *IEEE Access*, 7:115540–115551, 2019. doi:10.1109/ACCESS.2019.2936564.
- [187] Gabriel Solana-Lavalle and Roberto Rosas-Romero. Analysis of voice as an assisting tool for detection of parkinson’s disease and its subsequent clinical interpretation. *Biomedical Signal Processing and Control*, 66:102415, 2021. doi:10.1016/j.bspc.2021.102415.
- [188] Quoc Cuong Ngo, Mohammad Abdul Motin, Nemuel Daniel Pah, Peter Drotár, Peter Kempster, and Dinesh Kumar. Computerized analysis of speech and voice for parkinson’s disease: A systematic review. *Computer Methods and Programs in Biomedicine*, 226, 2022. doi:10.1016/j.cmpb.2022.107133.
- [189] Jan Ruzs, Tereza Tykalová, Michal Novotný, David Zogala, Evžen Ržička, and Petr Dušek. Automated speech analysis in early untreated parkinson’s disease: Relation to gender and dopaminergic transporter imaging. *European Journal of Neurology*, 29(1):81–90, 2022. doi:10.1111/ene.15099.
- [190] Katie L Stone and Sonia Ancoli-Israel. Actigraphy. In *Principles and practice of sleep medicine*, pages 1671–1678. Elsevier, 2017.
- [191] Meredith A Ray, Shawn D Youngstedt, Hongmei Zhang, Sara Wagner Robb, Brook E Harmon, Girardin Jean-Louis, Bo Cai, Thomas G Hurley, James R Hébert, Richard K Bogan, et al. Examination of wrist and hip actigraphy using a novel sleep estimation procedure. *Sleep Science*, 7(2):74–81, 2014. doi:10.1016/j.slsci.2014.09.007.
- [192] Marek Mikulec. SYSTÉM ZABEZPEČENÉHO PŘENOSU A ZPRACOVÁNÍ DAT Z AKTIGRAFU. Master’s thesis, Brno University of Technology, the Czech Republic, 2020.

- [193] Harneet K Walia and Reena Mehra. Practical aspects of actigraphy and approaches in clinical and research domains. In *Handbook of clinical neurology*, volume 160, pages 371–379. Elsevier, 2019. doi:10.1016/B978-0-444-64032-1.00024-2.
- [194] Hiroshi Kataoka, Keigo Saeki, Norio Kurumatani, Kazuma Sugie, and Kenji Obayashi. Objective sleep measures between patients with parkinson’s disease and community-based older adults. *Sleep Medicine*, 68:110–114, 2020. doi:10.1016/j.sleep.2019.09.010.
- [195] Jirada Sringean, Chanawat Anan, Chusak Thanawattano, and Roongroj Bhidayasiri. Time for a strategy in night-time dopaminergic therapy? an objective sensor-based analysis of nocturnal hypokinesia and sleeping positions in parkinson’s disease. *Journal of the Neurological Sciences*, 373:244–248, 2017. doi:10.1016/j.jns.2016.12.045.
- [196] Roongroj Bhidayasiri and Claudia Trenkwalder. Getting a good night sleep? the importance of recognizing and treating nocturnal hypokinesia in parkinson’s disease. *Parkinsonism & related disorders*, 50:10–18, 2018. doi:10.1016/j.parkreldis.2018.01.008.
- [197] Gianluca Aloï, Giuseppe Caliciuri, Giancarlo Fortino, Raffaele Gravina, Pasquale Pace, Wilma Russo, and Claudio Savaglio. A mobile multi-technology gateway to enable iot interoperability. In *2016 IEEE First International Conference on Internet-of-Things Design and Implementation (IoTDI)*, pages 259–264, Berlin, April 2016. IEEE. doi:10.1109/IoTDI.2015.29.
- [198] Ting Liang and Yong J Yuan. Wearable medical monitoring systems based on wireless networks: A review. *IEEE Sensors Journal*, 16(23):8186–8199, 2016. doi:10.1109/JSEN.2016.2597312.
- [199] Salvatore Tedesco, John Barton, and Brendan O’Flynn. A review of activity trackers for senior citizens: Research perspectives, commercial landscape and the role of the insurance industry. *Sensors*, 17(6):1277, 2017. doi:10.3390/s17061277.
- [200] Ritu Dhull, Dheeraj Chava, Deepala Vineeth Kumar, Kantipudi MVV Prasad, Gaurav Samudrala, and M Vijay Bhargav. Pandemic stabilizer using smart-watch. In *2020 International Conference on Decision Aid Sciences and Application (DASA)*, pages 860–866, Virtual, Sakheer, November 2020. IEEE. doi:10.1109/DASA51403.2020.9317056.



- [201] Malathi Devarajan and Logesh Ravi. Intelligent cyber-physical system for an efficient detection of parkinson disease using fog computing. *Multi-media Tools and Applications*, 78(23):32695–32719, 2019. doi:10.1007/s11042-018-6898-0.
- [202] Roisin McNaney, Emmanuel Tseklevs, and Jonathan Synnott. Future opportunities for iot to support people with parkinson’s. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–15, Honolulu HI USA, April 2020. Association for Computing Machinery. doi:10.1145/3313831.3376871.
- [203] Di Pan, Rohit Dhall, Abraham Lieberman, and Diana B Petitti. A mobile cloud-based parkinson’s disease assessment system for home-based monitoring. *JMIR mHealth and uHealth*, 3(1):e29, 2015. doi:10.2196/mhealth.3956.
- [204] Vahid Farrahi, Maisa Niemelä, Maarit Kangas, Raija Korpelainen, and Timo Jämsä. Calibration and validation of accelerometer-based activity monitors: A systematic review of machine-learning approaches. *Gait & posture*, 68:285–299, 2019. doi:10.1016/j.gaitpost.2018.12.003.
- [205] Martin Benka Wallén, Håkan Nero, Erika Franzén, and Maria Hagströmer. Comparison of two accelerometer filter settings in individuals with parkinson’s disease. *Physiological measurement*, 35(11):2287, 2014. doi:10.1088/0967-3334/35/11/2287.
- [206] Mohammad Abu Alsheikh, Ahmed Selim, Dusit Niyato, Linda Doyle, Shaowei Lin, and Hwee-Pink Tan. Deep activity recognition models with triaxial accelerometers. *arXiv preprint arXiv:1511.04664*, 2015. doi:10.48550/arXiv.1511.04664.
- [207] Aria Khademi, Yasser El-Manzalawy, Orfeu M Buxton, and Vasant Honavar. Toward personalized sleep-wake prediction from actigraphy. In *2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, pages 414–417, Las Vegas, March 2018. IEEE. doi:10.1109/BHI.2018.8333456.
- [208] AP Zaretskiy, KS Mityagin, VS Tarasov, and DN Moroz. Periodic limb movements detection through actigraphy signal analysis. In *2019 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon)*, pages 1–5, Vladivostok, October 2019. IEEE. doi:10.1109/FarEastCon.2019.8934102.

- [209] Taeheum Cho, Unang Sunarya, Minsoo Yeo, Bosun Hwang, Yong Seo Koo, and Cheolsoo Park. Deep-actinet: End-to-end deep learning architecture for automatic sleep-wake detection using wrist actigraphy. *Electronics*, 8(12):1461, 2019. doi:10.3390/electronics8121461.
- [210] Jia Li, Yu Rong, Helen Meng, Zhihui Lu, Timothy Kwok, and Hong Cheng. Tatc: Predicting alzheimer’s disease with actigraphy data. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 509–518, London, August 2018. Association for Computing Machinery. doi:10.1145/3219819.3219831.
- [211] Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In *Advances in neural information processing systems*, volume 31, pages 6571–6583, Montréal, December 2018. Curran Associates, Inc.
- [212] Colin Lea, Michael D Flynn, Rene Vidal, Austin Reiter, and Gregory D Hager. Temporal convolutional networks for action segmentation and detection. In *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 2017-January, pages 156–165, Honolulu, July 2017. IEEE. doi:10.1109/CVPR.2017.113.
- [213] Lena Granovsky, Gabi Shalev, Nancy Yacovzada, Yotam Frank, and Shai Fine. Actigraphy-based sleep/wake pattern detection using convolutional neural networks. *arXiv preprint arXiv:1802.07945*, 2018. doi:10.48550/arXiv.1802.07945.
- [214] Hanrui Zhang, Kaiwen Deng, Hongyang Li, Roger L Albin, and Yuanfang Guan. Deep learning identifies digital biomarkers for self-reported parkinson’s disease. *Patterns*, 1(3), 2020. doi:10.1016/j.patter.2020.100042.
- [215] Niamh O’Mahony, Blanca Florentino-Liano, Juan J Carballo, Enrique Baca-García, and Antonio Artés Rodríguez. Objective diagnosis of adhd using imus. *Medical engineering & physics*, 36(7):922–926, 2014. doi:10.1016/j.medengphy.2014.02.023.
- [216] **Justyna Skibińska** and Radim Burget. The application of deep learning techniques in the electroencephalogram (eeg) analysis. In *2019 XXXV Finnish Union Radio Scientifique Internationale (URSI) Convention on Radio Science*, Tampere, October 2019. URSI.

- [217] David Ahmedt-Aristizabal, Clinton Fookes, Kien Nguyen, and Sridha Sridharan. Deep classification of epileptic signals. In *2018 40th Annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 332–335, Honolulu, July 2018. IEEE. doi:10.1109/EMBC.2018.8512249.
- [218] Akara Supratak, Hao Dong, Chao Wu, and Yike Guo. Deepsleepnet: A model for automatic sleep stage scoring based on raw single-channel eeg. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(11):1998–2008, 2017. doi:10.1109/TNSRE.2017.2721116.
- [219] Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces. *Journal of neural engineering*, 15(5):056013, 2018. doi:10.1088/1741-2552/aace8c.
- [220] Dongwei Chen, Rui Miao, Zhaoyong Deng, Na Han, and Chunjian Deng. Sparse granger causality analysis model based on sensors correlation for emotion recognition classification in electroencephalography. *Frontiers in computational neuroscience*, 15, 2021. doi:10.3389/fncom.2021.684373.
- [221] Narayan Puthanmadam Subramaniam, Reik V Donner, Davide Caron, Gabriella Panuccio, and Jari Hyttinen. Causal coupling inference from multivariate time series based on ordinal partition transition networks. *Nonlinear Dynamics*, 105(1):555–578, 2021. doi:10.1007/s11071-021-06610-0.
- [222] Hadi Banaee, Mobyen Uddin Ahmed, and Amy Loutfi. Data mining for wearable sensors in health monitoring systems: a review of recent trends and challenges. *Sensors*, 13(12):17472–17500, 2013. doi:10.3390/s131217472.
- [223] Philip Schmidt, Attila Reiss, Robert Dürichen, and Kristof Van Laerhoven. Wearable-based affect recognition—a review. *Sensors*, 19(19):4079, 2019. doi:10.3390/s19194079.
- [224] Frédéric Li, Kimiaki Shirahama, Muhammad Adeel Nisar, Lukas Köping, and Marcin Grzegorzek. Comparison of feature learning methods for human activity recognition using wearable sensors. *Sensors*, 18(2):679, 2018. doi:10.3390/s18020679.
- [225] Diane J Cook and Narayanan C Krishnan. *Activity learning: discovering, recognizing, and predicting human behavior from sensor data*. John Wiley & Sons, 2015. doi:10.1002/9781119010258.

- [226] Marília Barandas, Duarte Folgado, Letícia Fernandes, Sara Santos, Mariana Abreu, Patrícia Bota, Hui Liu, Tanja Schultz, and Hugo Gamboa. Tsfel: Time series feature extraction library. *SoftwareX*, 11:100456, 2020. doi: 10.1016/j.softx.2020.100456.
- [227] Time series feature extraction library. <https://tsfel.readthedocs.io/en/latest/>. Accessed: 2021-03-18.
- [228] Siti Agrippina Alodia Yusuf and Risanuri Hidayat. Mfcc feature extraction and knn classification in ecg signals. In *2019 6th International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE)*, pages 1–5, Semarang, September 2019. IEEE. doi:10.1109/ICITACEE.2019.8904285.
- [229] Uichin Lee, Kyungsik Han, Hyunsung Cho, Kyong-Mee Chung, Hwajung Hong, Sung-Ju Lee, Youngtae Noh, Sooyoung Park, and John M Carroll. Intelligent positive computing with mobile, wearable, and iot devices: Literature review and research directions. *Ad Hoc Networks*, 83:8–24, 2019. doi:10.1016/j.adhoc.2018.08.021.
- [230] Andrea Bizzego, Giulio Gabrieli, Cesare Furlanello, and Gianluca Esposito. Comparison of wearable and clinical devices for acquisition of peripheral nervous system signals. *Sensors*, 20(23):6778, 2020. doi:10.3390/s20236778.
- [231] Jerry Chen, Maysam Abbod, and Jiann-Shing Shieh. Pain and stress detection using wearable sensors and devices—a review. *Sensors*, 21(4):1030, 2021. doi: 10.3390/s21041030.
- [232] SM Israfil, Md Moklesur Rahman Sarker, Parisa Tamannur Rashid, Ali Azam Talukder, Khandkar Ali Kawsar, Farzana Khan, Selina Akhter, Chit Laa Poh, Isa Naina Mohamed, and Long Chiau Ming. Clinical characteristics and diagnostic challenges of covid- 19: an update from the global perspective. *Frontiers in public health*, 8:955, 2021. doi:10.3389/fpubh.2020.567395.
- [233] Justin Shenk. *Facial expression recognition*, 2020 (Accessed August 11, 2020). URL: <https://github.com/justinshenk/fer>.
- [234] Ian J Goodfellow, Dumitru Erhan, Pierre Luc Carrier, Aaron Courville, Mehdi Mirza, Ben Hamner, Will Cukierski, Yichuan Tang, David Thaler, Dong-Hyun Lee, et al. Challenges in representation learning: A report on three machine learning contests. In *International conference on neural information processing*, volume 8228 LNCS, pages 117–124. Springer, Springer, November 2013. doi:10.1007/978-3-642-42051-1\_16.

- [235] Iván de Paz Centeno. *Face detection*, 2020 (Accessed August 11, 2020). URL: <https://github.com/ipazc/mtcnn/>.
- [236] Lukas Snoek, Steven Miletic, and H Steven Scholte. How to control for confounds in decoding analyses of neuroimaging data. *Neuroimage*, 184:741–760, 2019. doi:10.1016/j.neuroimage.2018.09.074.
- [237] Mohamad Amin Pourhoseingholi, Ahmad Reza Baghestani, and Mohsen Vahedi. How to control confounding effects by statistical analysis. *Gastroenterology and hepatology from bed to bench*, 5(2):79, 2012.
- [238] Carlos Alonso-Martinez, Marcos Faundez-Zanuy, and Jiri Mekyska. A comparative study of in-air trajectories at short and long distances in on-line handwriting. *Cognitive computation*, 9:712–720, 2017. doi:10.1007/s12559-017-9501-5.
- [239] Stratified cross-validation. <https://towardsdatascience.com/what-is-stratified-cross-validation-in-machine-learning-8844f3e7ae8e>. Accessed: 2023-01-23.
- [240] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, San Francisco, August 2016. Association for Computing Machinery. doi:10.1145/2939672.2939785.
- [241] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, December 2017.
- [242] Shap library. <https://shap.readthedocs.io/en/latest/>. Accessed: 2022-11-23.
- [243] Sabine Skodda, Heiko Rinsche, and Uwe Schlegel. Progression of dysprosody in parkinson’s disease over time—a longitudinal study. *Movement disorders: official journal of the Movement Disorder Society*, 24(5):716–722, 2009. doi:10.1002/mds.22430.
- [244] Hamid Azadi, Mohammad-R Akbarzadeh-T, Ali Shoeibi, and Hamid Reza Kobravi. Evaluating the effect of parkinson’s disease on jitter and shimmer speech features. *Advanced Biomedical Research*, 10(1), 2021. doi:10.4103/abr.abr\_254\_21.

- [245] Andrea Fernández Martínez. Identification of patients at the risk of lewy body diseases based on acoustic analysis of speech. In *Proceedings of the 25th Conference STUDENT EEICT 2019 [online]*, pages 50–53, 2019.
- [246] Jan Ruzs, Marika Megrelishvili, Cecilia Bonnet, Michael Okujava, Hana Brožová, Irine Khatiasvili, Madona Sekhniashvili, Marina Janelidze, Eduardo Tolosa, and Evžen Ržička. A distinct variant of mixed dysarthria reflects parkinsonism and dystonia due to ephedrone abuse. *Journal of Neural Transmission*, 121(6):655–664, 2014. doi:10.1007/s00702-014-1158-6.