Radboud University Nijmegen

UNIVERSITÀ DEGLI STUDI
DI TRENTO

DEPARTMENT OF INFORMATION ENGINEERING AND COMPUTER SCIENCE
ICT International Doctoral School

Deep Learning for Abnormal Movement Detection using Wearable Sensors:

Case Studies on Stereotypical Motor Movements in Autism and Freezing of Gait in Parkinson's Disease

Nastaran Mohammadian Rad
International Doctorate School in Information
and Communication Technologies, Università degli Studi di Trento;
Institute for Computing and Information Sciences, Radboud University;
Fondazione Bruno Kessler (FBK)

Supervisors:
Dr. Cesare Furlanello
Fondazione Bruno Kessler
Prof. dr. Elena Marchiori
Radboud University
Dr. Twan van Laarhoven
Open University of the Netherlands

May 2019

Abstract

Inertial measurement sensing technology with the capability of capturing disease-relevant data has a great potential for improving the current clinical assessments and enhancing the quality of life in patients with neuro-developmental and neuro-degenerative diseases such as autism spectrum disorders (ASD) and Parkinson's disease (PD). The current clinical assessments can be improved by developing objective tools for the disease diagnosis and continuous monitoring of patients in out of clinical settings. To this end, it is necessary to develop automatic abnormal movement detection methods with the capability of adjusting on new patients' data in real-life settings. However, achieving this goal is challenging mainly because of the inter and intra-subject variability in acquired signals and the lack of labeled data. The research presented in this thesis investigates the application of deep neural networks to address these challenges of abnormal movement detection using inertial measurement unit (IMU) sensors with case studies on stereotypical motor movements in ASD and freezing of gait in PD patients. In this direction, this thesis provides four main contributions: i) A convolutional neural network (CNN) architecture is proposed to learn discriminative features which are sufficiently robust to inter and intra-subject variability. It is further shown how the proposed CNN architecture can be used for parameter transfer learning to enhance the adaptability of the abnormal movement detection system to new data in a longitudinal study. ii) An application of recurrent neural networks and more specifically long short-term memory (LSTM) in combination with CNN is proposed in order to incorporate more the temporal dynamics of IMU signals in the process of feature learning for abnormal movement detection. iii) An ensemble learning approach is proposed to improve the detection accuracy and at the same time to reduce the variance of models. iv) In the normative modeling

framework, the problem of abnormal movement detection is redefined in the context of novelty detection and it is shown how a probabilistic denoising autoencoder can be used to learn the distribution of the normal human movements. The resulting deep normative model then is used in a novelty detection setting for unsupervised abnormal movement detection. The experimental results on three benchmark datasets collected from ASD and PD patients illustrate the high potentials of deep learning paradigm to address the crucial challenges toward real-time abnormal movement detection systems using wearable technologies.

Keywords

Wearable Sensors; Deep Learning; Novelty Detection; Parkinson's Disease; Autism Spectrum Disorders; Freezing of Gait; Stereotypical Motor Movement.

Acknowledgments

First, I would like to express my gratitude to my supervisor at Fondazione Bruno Kessler, Cesare Furlanello who gave me the chance to work with him, for letting me grow and find my way. He was always open to possibilities and provided me with his insightful comments. I am very grateful to him for all his support, guidance, and encouragement during my Ph.D.

My sincere thanks also go to my supervisor at Radboud University, Elena Marchiori who provided me with an opportunity to join Data Science lab as a joint-Ph.D. student. I am very grateful to her for the kind hospitality, encouragement, guidance, fruitful discussions, and support which helped me to successfully complete this research.

I would like to also express my sincere thanks to my supervisor at Radboud University, Twan van Laarhoven for being very supportive. I am grateful to him for the constructive comments, guidance, and discussions.

Besides my supervisors, I also would like to thank my thesis committee: Natalia Diaz Rodriguez, Peter Lucas, and Mauro Brunato for their insightful comments that improved this work significantly.

I also gratefully acknowledge the financial support during my Ph.D. from Fondazione Bruno Kessler and also the Netherlands Organization for Scientific Research (NWO) through the Radboud University.

I also would like to thank Paola Venuti who gave me the possibility of collaborating with ODFLab in the first starting months of my Ph.D. It was a wonderful experience which inspired me to continue my research on autistic children.

I would like to thank all my colleagues in both MPBA lab at Fondazione Bruno Kessler and Data Science lab at Radboud University from whom I have learned a lot. More specifically, I would like to thank Kasper Brink, Nicole Messink, and Claudia Dolci for their persistent kind support.

My final words go to my family. I would like to thank my beloved husband for his support, understanding, and love. I am grateful to him for his encouragement and for keeping me motivated during my Ph.D. Above all, no words can fully express my thankfulness to my parents for their unconditional love, the sympathetic ear and spiritual supports throughout my life.

Contents

1 Introduction **1**

2 Background **7**

2.1 Autism Spectrum Disorder 7

2.2 Parkinson 9

2.3 Wearable Sensor Technology for Human Activity Recognition 11

2.4 Pipeline for IMU Data Analysis 16

2.5 Performance Evaluation 18

2.6 Deep Neural Networks 19

2.6.1 Convolutional Neural Networks 19

2.6.2 Long Short-Term Memories 21

2.6.3 Autoencoders 23

3 Convolutional Neural Networks for Stereotypical Motor Movement Detection **25**

3.1 Introduction 25

3.1.1 Contribution 27

3.2 Methods 29

3.2.1 Notation 29

3.2.2 Feature Learning via Convolutional Neural Network 30

3.2.3 Parameter Transfer Learning via Network Pre-initialization 32

3.2.4 Experimental Materials 34

3.2.5 Experimental Setups and Evaluation 37
3.3 Results 40
3.3.1 Feature Learning Outperforms Handcrafted Features 40
3.3.2 Parameter Transfer Learning is Beneficial in Longitudinal Studies 41
3.3.3 Training on Unbalanced Data Decreases the Performance 42
3.4 Discussion 43
3.4.1 Toward Real-Time Automatic SMM Detection . . . 43
3.4.2 Limitation and Future Work 44
3.5 Conclusions 45

4 Stereotypical Motor Movement Detection in Dynamic Feature Space 47
4.1 Introduction 47
4.1.1 Contribution 48
4.2 Methods 49
4.2.1 SMM Detection via LSTM 49
4.2.2 Ensemble of the Best Base Learners 52
4.2.3 Experimental Materials 52
4.2.4 Experimental Setup 53
4.3 Results 54
4.3.1 Dynamic Feature Learning Outperforms Static Feature Learning 54
4.3.2 Ensemble of LSTMs Stabilizes the Performance . . 58
4.4 Conclusions 59

5 Novelty Detection using Deep Normative Modeling for IMU-Based Abnormal Movement Monitoring in Parkinson's Disease and Autism Spectrum Disorders 61

5.1 Introduction 61
5.1.1 Contribution 63
5.2 Related Works 64
5.3 Methods 65
5.3.1 Learning the Distribution of Normal Movements via the Denoising Autoencoder 66
5.3.2 Quantifying the Deviation from P_N 68
5.3.3 Computing the Degree of Novelty 68
5.3.4 Experimental Materials 69
5.4 Results 75
5.4.1 Normative Modeling Outperforms Reconstruction-Based and One-Class SVM in Novelty Detection . . 77
5.4.2 Novelty Detection Methods vs. Supervised Learning Methods 77
5.4.3 Effect of Dropout Level 80
5.4.4 Training Only on Normal Subjects 80
5.5 Discussion 82
5.5.1 Estimating the Prediction Uncertainty: Deep Learning vs. Gaussian Processes 82
5.5.2 Normative Modeling vs. One-Class Classification . 83
5.5.3 Toward Modeling Human Normal Daily Movements Using Wearable Sensors 84
5.5.4 Limitations and Future Work 84
5.6 Conclusions 85

6 Conclusions **87**

A Appendices **91**
A.1 Generalized Extreme Value Distribution 91

Bibliography **91**

List of Tables

3.1 Number of samples in SMM and no-SMM classes in three datasets. 37

3.2 F1-score results for SMM detection using raw, handcrafted, and static feature spaces in three benchmarked datasets. The results show that feature learning (here referred to as Static-Features) generally outperforms raw and handcrafted feature spaces. In addition, the parameter transfer learning has a positive effect on the performance of the CNN classifier. Furthermore, training the CNN classifier on unbalanced training sets causes the performance drop in feature learning and transfer learning scenarios. 43

4.1 F1-score results for SMM detection using static and dynamic feature spaces, and ensemble learning in three benchmarked datasets. The results show training the CNN classifier on unbalanced training sets causes the performance drop in the feature learning scenario. Using the LSTM network to extract dynamic features from the signal alleviates this problem to some degrees. The ensemble of LSTMs shows more stable performance compared to single LSTM classifiers. . 59

5.1 The class distribution of normal and abnormal samples and the gender of patients in three datasets. 70

5.2 The average of AUC results for novelty detection using normative modeling, reconstruction-based and one-class SVM on three benchmark datasets. 76
5.3 The average AUPR for novelty detection using normative modeling, reconstruction-based and one-class-SVM on three benchmark datasets. 79
5.4 The average of AUC results for novelty detection using normative modeling, reconstruction-based and one-class-SVM trained only on the two available normal subjects (Subjects 4 and 10) of the Daphnet-FOG dataset. 82

List of Figures

2.1 Comparison of the substantia-nigra in a healthy subject and a patient with PD [1]. 10

2.2 (a) Scanning electron microscopy (SEM) view of a MEMS accelerometer [2]; (b) A general scheme of one-axis MEMS accelerometer. 13

2.3 (a) The Coriolis acceleration [3], (b) The schematic of MEMS gyroscope with the tuning-fork configuration [4]. 15

2.4 A general pipeline for abnormal movement detection using wearable sensors. 16

2.5 An illustration of the 1-D max pooling functionality on the time-series data. 21

2.6 A schematic illustration of an RNN. 22

2.7 A schematic representation of a memory cell in LSTM (the picture from `http://colah.github.io/posts/2015-08-Understanding-`

2.8 A general architecture for an autoencoder. 24

3.1 A real-time automatic SMM detection system. Inertial Measurement Units (IMUs) can be used for data collection. The collected data can be analyzed locally or remotely to detect SMMs. In case of detecting abnormal movements, an alert is sent to a therapist, caregiver, or parents. 26

3.2 (a) The proposed architecture for SMM detection in the static feature space using a three-layer CNN. (b) The first CNN layer. This layer receives the one-second long time-series of several IMU sensors at time t, i.e., $\mathbf{X}_t$, and transfer it to the first level reduced feature map $\mathbf{M}'_t$. (c) The second CNN layer that uses the first level reduced feature map $\mathbf{M}'_t$ as its input, and transfer it to the second-level reduced feature map. (d) The third CNN layer. The reduced feature map of this layer is reshaped to the learned feature vector $\mathbf{x}'_t$ using the flattening operation. (e) The learned feature vector is fed to a fully-connected followed by a softmax layer to classify the samples to SMM and no-SMM classes. 33

3.3 (a) The configuration of the EXLs3 sensor on the right hand. (b),(c) The simulated data are collected during daily work activities in the office (e.g., sitting, writing, and typing). (d),(e) The subjects are asked to intermittently perform hand flapping upon receiving a start/stop cue from the instructor. 36

3.4 The distribution of SMM and no-SMM samples in the 2-dimensional t-SNE space for (a) raw feature space, (b) handcrafted features, (c) static feature space learned by CNN, and (d) static feature space learned by pre-initialized CNN. Feature learning increases the separability of samples in two classes compared to raw and handcrafted features. 42

4.1 The proposed architecture for SMM detection in the dynamic feature space using a long short-term memory. Each feature extraction block contains a trained three-layer CNN architecture (see Figure 3.2). Each of $\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_k$ refer to a raw data channel. 51

4.2 Comparison between the classification performances of CNN and LSTM for different time-steps (τ) and the number of neurons q, on three datasets and when an unbalanced training set is used for training the networks. The results show the superiority of the dynamic feature space over the static feature space. While the number of neurons in the LSTM unit has little effect on the performance, using around 2.5 seconds long interval is the best choice for extracting effective dynamic features from the IMU signals. 55

4.3 The distribution of SMM and no-SMM samples in the 2-dimensional t-SNE space for a) a raw feature space ($\mathbf{x}_t$), b) static feature space computed by CNN ($\mathbf{x}'_t$), and c) dynamic feature space computed by LSTM (h_t). 56

4.4 A comparison between precision and recall rates of classifiers trained on balanced/unbalanced training sets for static/dynamic feature representations. Using dynamic feature space provides an SMM detection system with higher sensitivity and specificity. 58

5.1 The proposed method for the abnormal movement detection in the test time. 66

5.2 The architecture of convolutional denoising autoencoder for (**a**) the Daphnet-FOG dataset and (**b**) the SMM dataset. Each colored box represents one layer. The type and configuration of each layer are shown inside each box. For example, Conv 64-5 denotes a convolutional layer with 64 filters and the kernel size of 5. 73
5.3 ROC curves corresponding to the reported AUCs for Subjects 1 and 6 of the Daphnet-FOG dataset in Table 5.2. . . 78
5.4 The effect of different dropout probabilities on the performance of the normative modeling method. 81

List of Acronyms

IMU	Inertial Measurement Unit
PD	Parkinson's Disease
ASD	Autism Spectrum Disorder
HAR	Human Activity Recognition
FOG	Freezing Of Gait
SMM	Stereotypical Motor Movements
MEMS	Micro-Electromechanical System
t-SNE	t-distributed Stochastic Neighbor Embedding
CNN	Convolutional Neural Network
ReLU	Rectified Linear Units
RNN	Recurrent Neural Network
LSTM	Long Short-Term Memory
EL	Ensemble Learning
DAE	Denoising Autoencoder
SVM	Support Vector Machine
RBF	Radial Basis Function
GEVD	Generalized Extreme Value Distribution
TP	True Positive
FP	False Positive
FN	False Negative
ROC	Receiver Operating Characteristic
AUC	Area Under the Curve
AUPR	Area Under the Precision-Recall Curve
MC	Monte Carlo
MTGPR	Multi-Task Gaussian Process Regression

Chapter 1

Introduction

Human activity recognition (HAR) is an emerging research field in the mobile and ubiquitous computing communities. Human movement analysis is one of the main scopes of research in the HAR field. In recent years, human movement analysis has been used in a variety of applications ranging from daily life activity recognition [5, 6, 7, 8, 9, 10] to medical and health care applications [11]. In the HAR context, a movement is defined as a change in human body style in the space-time dimensions [12, 13]. For a specific movement, patterns of these changes are consistent across different individuals, however, these normal patterns are subject to change in patients who suffer from neuro-developmental and neuro-degenerative disabilities. Thus, detecting and quantifying the deviation of body movement in patients with respect to the healthy population can provide an estimation of the disease state in patients suffering from movement disorders.

Autism spectrum disorders (ASD) and Parkinson's disease (PD) are respectively, two high-prevalent neuro-developmental and neuro-degenerative disorders manifesting specific motor symptoms [14, 15]. ASD is associated with the lack of tendency to social interactions and the presence of repetitive behaviors [14]. Stereotypical motor movements (SMM) are a major group of repetitive and aimless motions in individuals with ASD

that include body rocking, mouthing, hand flapping, and complex finger movements [16, 17, 18]. SMMs negatively affect the quality of life of ASD children by interfering with their learning process and social interactions. SMMs limit the performance of children in both learning new skills and using the learned skills. Since these behaviors are socially stigmatizing and abnormal, they cause difficulties in the social interaction with other peers at school [14, 19]. Beyond the potential negative consequences of SMMs on the quality of life of autistic children, these behaviors also increase the stress and anxiety level in parents and caregivers [20]. Thus, it is necessary to assist children with ASD by guiding efficient treatments to alleviate the intensity and frequency of SMMs. Planning a successful treatment depends on the comprehensive understanding and accurate measurement of SMMs [19]. The traditional clinical assessments for measuring SMMs rely on the paper-and-pencil rating scales, direct behavioral observation, and video-based coding [19]. However, these approaches suffer from major deficits such as subjectivity in rating, inefficiency, inaccuracy, and high cost [21, 22, 23]. These limitations highlight the importance of developing less obtrusive and more objective tools that allow continuous monitoring of SMMs over time in real-life settings.

PD is one of the most common progressive neuro-degenerative diseases with a higher prevalence in elderly people. It is estimated that around 1 million people in the USA and 10 million people worldwide suffer from this disease [24]. PD is determined by some specific motor symptoms such as tremor at rest, bradykinesia (slowness), muscle rigidity, impaired posture and balance, and freezing of gait (FOG). These motor symptoms restrict patients' daily-life activities, cause the loss of mobility, and reduce the social activities of patients [25]. For example, FOG is one of the most restrictive motor deficits in PD: it causes difficulty in generating effective forward-stepping movements and as a result, increases the risk of falling

in generally elderly patients. The current clinical assessments are based on the direct observation and patients' report [26, 27] which suffer from subjectivity in rating. Due to the rapid growth of PD in the elderly population and the limitations of the current clinical practices, it is necessary to develop methods and tools which are able to unobtrusively and accurately collect patients' movement data, measure the disease state, assess the effect of the medical interventions, and continuously follow-up patients in realistic environments.

Recent enthusiasm toward collecting accurate human movement data in real-life settings emerged a significant interest in using body-worn inertial measurement unit (IMU) sensors [28, 29]. IMUs –with the built-in accelerometer, gyroscope, and magnetometer sensors– measure the linear acceleration and the angular velocity of body movements [30]. IMU sensors are widely used in clinical applications [31]. Due to their small size and light-weights, IMU is one of the most comfortable and less obtrusive devices to collect the ecologically valid movement data from children with ASD and elderly people with PD. Furthermore, IMUs are widely integrated into mobile phones and smartwatches, thus provide an exceptional opportunity for long-term monitoring of patients in both clinical and non-clinical settings [32]. Achieving this goal, of course, demands developing algorithms that are able to accurately and reliably detect abnormal patterns in the collected human movement data beyond the clinical conditions. To this end, many research studies have applied various machine learning algorithms on the stream IMU signals from ASD and PD patients for the abnormal movement detection [33, 34, 35, 36, 19].

Despite considerable efforts toward developing an accurate abnormal movement detection system for PD and ASD patients, this still remains a challenging task due to the high inter and intra-subject variability, high-class imbalance, and lack of labeled data:

Inter-subject and intra-subject variability: one of the main challenges toward developing an automatic abnormal movement detection system is the inter and intra-subject variability. The inter-subject variability refers to different topology (the type of movement), intensity, duration, and frequency of abnormal movements between different patients. The intra-subject variability is defined as the same differences within each individual patient. These variabilities cause a decrease in the performance when the method is applied to new patients or even on the same patient over time. Thus, it is crucial to develop an abnormal movement detection algorithm that is robust to these variations in data. One possible solution to this problem is to extract a set of features that are robust to inter and intra-subject variations [11, 37, 38, 39]. The traditional feature extraction methods, so-called hand-crafted features, rely on the expert's domain knowledge. Thus, the extracted features may not contain all the important information in the data. Hence, recently there has been a huge interest in learning a high-level and at the same time lower dimensional representation of data using deep neural networks.

Lack of labeled data: another main issue that limits the application of supervised methods for abnormal movement detection in the real-life scenarios is the lack of annotated data. Data annotation is an expensive and time-consuming task, as an expert has to annotate the data in the real-time settings or to use the captured video for the offline annotation in the laboratory settings [40, 41, 42]. This issue is even more challenging in the daily-life settings where monitoring subjects' activities is impossible.

Class imbalance: class imbalance is a general issue in clinical datasets as generally, the ratio of abnormal movements to normal movements is small. In general, the class imbalance problem can be addressed simply by balancing the training data based on the number of samples in the minority class [19]. However, this results in a significant reduction in the sample size.

Therefore, finding detection models that are able to handle the imbalanced class distribution without losing the generalization performance is of high priority in the abnormal movement detection context.

The aim of this thesis is to develop effective machine learning approaches to address the aforementioned issues. The rest of the thesis is organized in 6 chapters as follows:

1. Chapter 2 provides the general background information on the discussed materials in this thesis. To this end, first, we will have an overview on the definition of autism spectrum disorders and Parkinson's disease, their symptoms, and the current clinical assessments for detecting SMMs and FOGs. Second, a brief introduction to IMU sensors –including the functionality of accelerometers and gyroscopes– and the principles of IMU data analysis are presented. This chapter is, then, finalized by a brief introduction to deep neural networks architectures and algorithms.

2. Chapter 3 presents a new convolutional neural network (CNN) architecture for automatic feature learning from IMU signals. The proposed CNN architecture is then used for parameter transfer learning to address the problem of intra and inter-subject variability in individuals with ASD in the longitudinal scenarios.

3. Chapter 4 investigates the effect of incorporating temporal information in the consecutive time intervals of IMU sensor data on the performance of SMM detection. To this end, a long short-term memory (LSTM) is combined with the CNN in order to model the temporal patterns in the sequence of multi-axes IMU signals. Further, an ensemble learning based on the best LSTM learners is proposed to provide a more stable SMM detector.

4. Chapter 5 is devoted to addressing the problem of lack of labeled data in patients with PD and individuals with ASD in real-life settings. This problem is addressed by proposing a probabilistic novelty detection approach. The proposed method allows to train a model of normal human movements in an unsupervised manner and then to detect FOG episodes in PD patients and SMMs in individuals with ASD as deviations from the learned model.

5. Finally, Chapter 6 summarizes our findings in this thesis, draws conclusions, and gives an outlook to possible future works.

Chapter 2

Background

This chapter provides the background information on all needed materials for understanding the contributions of this thesis in the next chapters. To this end, first, autism spectrum disorder (ASD) and Parkinson disease (PD), their symptoms, and their current clinical assessments are briefly described and the importance of developing automatic methods for detecting abnormal movements in ASD and PD patients is motivated. Then, the operating principles of inertial measurement units (IMU) are introduced and a general pipeline for IMU data analysis is explained. This chapter is finalized by a general introduction to employed deep neural network architectures in this thesis.

2.1 Autism Spectrum Disorder

ASD is a neuro-developmental disorder that is associated with i) the difficulty in the social communication and interaction with other people; and ii) the unusual repetitive and restricted behaviors in children [43, 44, 14]. It is estimated that the prevalence of ASD is about 1 in each 68 children in the United States [45]. The prevalence of ASD is 4 times more in males than females [14]. Despite a general agreement on the genetic source of ASD, still, the genes that predispose children to ASD have re-

mained uncertain [43, 14, 44]. Recent researches in this area suggest that the advanced age in parents and having certain genetic conditions (such as Down syndrome, fragile X syndrome, and Rett syndrome) are associated with high ASD risk [43, 44].

The symptoms of ASD gradually develop in the first three years of life. These symptoms present a high heterogeneity of types and severity over a period of time between and within individuals [46]. The difficulty in social interaction, as one of the symptoms in individuals with ASD, is determined by abnormal behavioral symptoms such as limited visual and auditory connection with surrounding people and having trouble in understanding or predicting other people's actions [46]. Individuals with ASD are also engaged in repeating certain unusual movements such as hand flapping, head rolling, finger wiggling, mouthing, and body rocking [47, 48]. These complex, repetitive, aimless, and involuntary behaviors, commonly referred to as *stereotypical motor movements* (SMMs) [49].

SMMs impair the quality of life in individuals with ASD [50] as these behaviors highly interfere with learning new skills [51, 52] and with the social interaction with other peers in the school or other social communities [53]. The high prevalence of autism and the negative consequences on patients' quality of life emerge the necessity of measuring SMMs and providing possible interventions to mitigate these abnormal behaviors.

Traditional approaches to measure SMMs include paper-and-pencil rating, direct behavioral observation, and video-based coding [17, 18]. Paper and pencil rating is an interview-based approach which suffers from the subjectivity in rating and inability to accurately detect the severity, amount, and duration of SMMs [22]. In the direct behavioral observation approach, SMMs are directly observed and documented by clinicians. This is not a reliable approach as clinicians assess and measure SMMs based on a short observation. Furthermore, in high-speed movements, it is impossible for

clinicians to accurately document all SMMs episodes [21]. Video-based approaches rely on analysis of SMMs using the video capturing. This method is more precise than two previous approaches but it is impractical as a ubiquitous tool, as its application is limited to very controlled clinical settings [23].

Considering the importance of quantifying the severity and frequency of SMMs in the diagnosis and treatment [19] of children with ASD, there is a pressing need to develop automatic, accurate, and time-efficient methods for SMM detection. Recently body-worn sensor technology has offered an effective tool for collecting accurate human movement data with the possibility of time-efficient SMM detection. This technology has a great potential to enhance the current clinical assessment by continuously monitoring autistic children. However, the high variability of abnormal movements in children with ASD imposes new challenges to this field.

2.2 Parkinson

PD is the second most common neuro-degenerative disorder that affects the functioning of the motor system in the brain. The PD prevalence is estimated to be around 1% in the elderly population [15]. The etiology of PD has not been entirely understood but many research studies suggest that both genetic and non-genetic factors effectively contribute to the occurrence of PD [15]. The pathogenic process that causes PD starts with the reduction of dopamine level in the brain. Dopamine is produced by *substantia nigra* in the brain, acts as a neurotransmitter, and plays a role in initiating movements and speech. Thus, the lack of dopamine causes malfunctioning in the brain's motor system [1]. Figure 2.1 depicts the difference between the structure of substantia nigra in a healthy subject and a patient with PD.

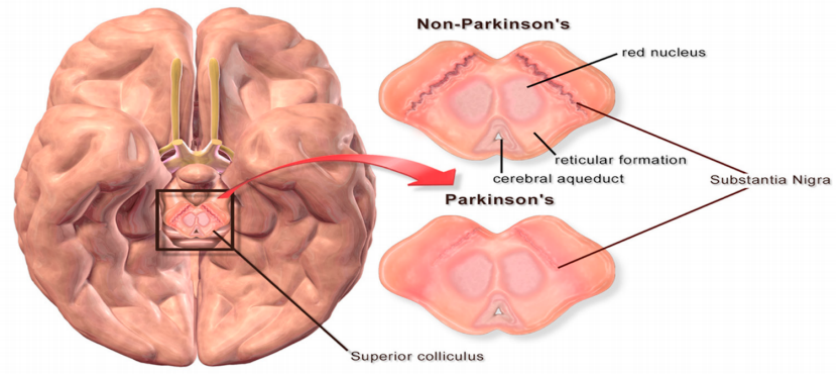

Figure 2.1: Comparison of the substantia-nigra in a healthy subject and a patient with PD [1].

Motor symptoms in PD patients appear gradually over time and may start with slight tremors in the patient's hand [54]. The most common types of motor symptoms that PD patients experience are bradykinesia (the inability to move the body swiftly), tremor at rest, muscle rigidity, impaired posture and balance, and freezing of gait (FOG) [55]. FOG is a gait disorder which occurs in the later stages of the PD [56]. FOG appears more frequently when the patient changes his/her direction or walks in narrow spaces. FOG episodes impair severely the quality of life in elderly patients as they increase the risk of falling [56, 57].

To estimate the disease severity and to optimize treatment strategies, it is necessary to accurately measure the duration and severity of abnormal movements in PD patients [58]. The traditional clinical evaluations include direct observation and patients' self-report [26, 59]. In the direct observation, clinicians assess the motor symptoms in patients directly during simple motor tasks such as getting up from a chair and walking a short distance [26]. This approach may not truly present the patients' problem, as the clinical circumstances may have an effect on the severity and occurrence of patients' abnormal motor movements [60, 59]. For example, even patients who are severely debilitated by frequent freezing episodes

at home, show fewer FOG episodes in clinical circumstances. The patients' self-report approach suffers from the subjectivity and the existence of bias in the rating [61, 59]. For example, a patient may not be aware of the relation between the off/on medical conditions and the occurrence or severity of disorder movements which may influence his/her self-report. To overcome these limitations, many research studies have focused on using wearable sensors which offer great tools for collecting accurate human movement data in realistic environments [62, 63]. Wearable sensors provide the possibility of monitoring the patients more frequently than the current clinical practices with minimum costs. Furthermore, they provide the potential for reliable and rapid feedback on patients' status. This reliable and rapid feedback has a positive effect on improving physical activities in PD patients [64, 27].

2.3 Wearable Sensor Technology for Human Activity Recognition

Accelerometers and gyroscopes are the most common wearable sensors for measuring the frequency, intensity, and duration of physical activities [65, 66, 30, 28]. These sensors in combination with magnetometer sensors are typically integrated into the same package, so called the inertial measurement unit (IMU) [30]. Due to the small size and low cost, IMUs offer the most appropriate and comfortable devices to be attached to human bodies for capturing human movement data without negatively affecting the quality of the motion of interest. In contrast to other motion sensing technologies, e.g., Kinect sensors or generic cameras, IMUs do not need any additional sensing infrastructure, which is a beneficial property especially in real-life settings [12].

Accelerometers measure the linear acceleration along 3 orthogonal axes [30].

Despite the differences in transducers and mechanisms used in their design, all type of accelerometers quantify the effect of the force applied to the sensor as a measure of the acceleration [65, 12, 67, 8]. Accelerometers are used in numerous consumer and industrial applications ranging from monitoring mechanical shock [68] to inertial navigation systems [69] and virtual reality [70]. Recently, accelerometers with the micro-electromechanical system (MEMS) technology have been widely used in mobile phones and smart watches [71]. The high resolution and sensitivity have made MEMS accelerometers suitable devices to be used for sensing human movements [72].

The most popular MEMS accelerometers employed for the human movement measurement are variable capacitive accelerometers that measure the acceleration by sensing changes in the capacitance [65]. Variable capacitive accelerometers are comprised of differential capacitors with the fixed external plates and a movable proof mass suspended between the fixed plates. The proof mass is further attached to springs. Using this configuration, the external acceleration causes a vibration in the proof mass that changes the capacity of capacitors, resulting in an output signal with an amplitude proportional to the acceleration [68, 65]. Figure 2.2 shows the structure of a capacitive MEMS accelerometer with one degree of freedom (i.e., one sensing axis).

Depending on their degree of freedom, accelerometers can be categorized into the uniaxial (1-axis), biaxial (2-axis), or tri-axial (3-axis). In the case of 3-axis accelerometers, a proof mass is shared between three oriented orthogonal capacitive sensing structure which can simultaneously measure the acceleration along X, Y, and Z axes [12]. The measured acceleration is the sum of the gravitational and movement acceleration along a sensing direction, as follow [73]:

$$a = a_g + a_l + \eta \tag{2.1}$$

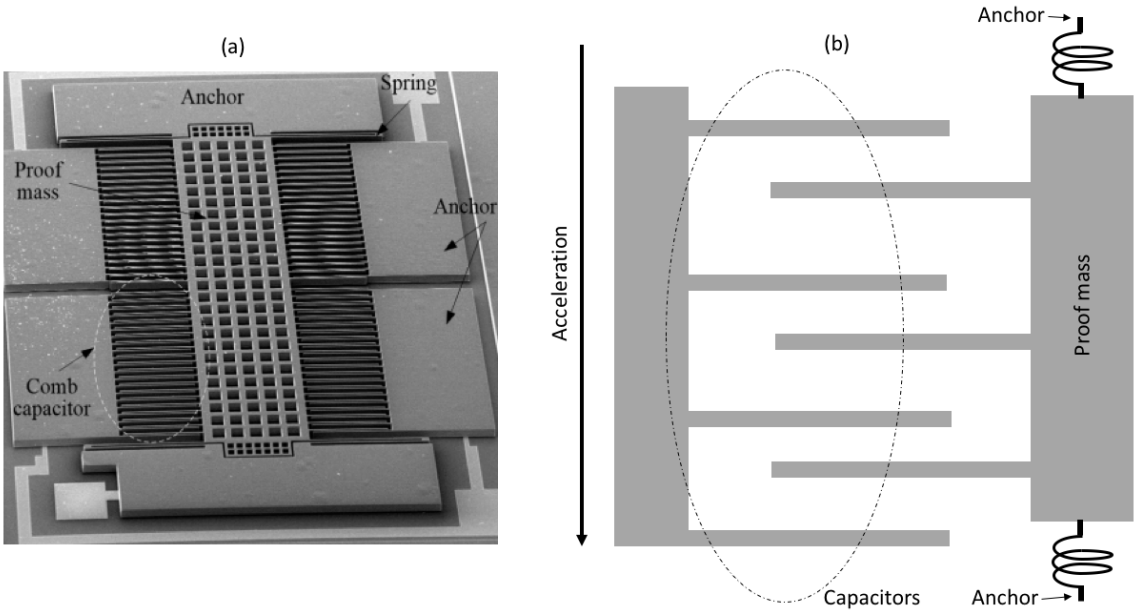

Figure 2.2: (a) Scanning electron microscopy (SEM) view of a MEMS accelerometer [2]; (b) A general scheme of one-axis MEMS accelerometer.

where a_g is the gravitational acceleration, a_l is the movement acceleration, and η is the noise. The existence of the gravitational acceleration is beneficial only when there is no movement acceleration. In this case, the gravitational acceleration helps to determine the orientation of the device related to the gravitational direction. The presence of the gravitational acceleration term during the device movement makes the analysis of the motion challenging as the measurement contains the gravitational acceleration added to the actual acceleration [74, 65].

As it was previously mentioned, accelerometers can only capture the linear acceleration of the sensing device. In order to measure the rotation of the sensing device, we need to use gyroscopes. Gyroscopes measure the angular velocity (i.e., rotational speed) and the angular displacement around a particular axis in radians per second [73].

In recent years, MEMS gyroscopes have become the most common type of gyroscopes used in the portable devices, due to their small size, af-

fordable cost, and the capability of integration with the digital interfaces into one package [3, 75]. Gyroscopes are recently used in a variety of applications such as robotic applications, navigation systems, and medical applications [68, 3].

MEMS gyroscopes employ a vibrating mechanical element to detect the angular velocity. These sensors transfer the energy between two vibration modes caused by Coriolis acceleration [3, 68]. Coriolis acceleration refers to an acceleration that appears in a rotating reference frame and is directly proportional to the angular velocity [68]. To better understand the Coriolis acceleration, assume a particle moving in the space with a velocity vector ν while an observer is anchored on the z-axis of a coordinate system watching the particle (see Figure 2.3 (a)). If the coordinate system rotates around the x-axis with an angular velocity equal to Ω, the observer sees the particle moving toward the z-axis in the rotating reference frame with the acceleration equal to $a_{coriolis} = 2\nu \times \Omega$, while there is no real force applied to the particle to move along the z-axis. All vibratory gyroscopes, in fact, operate based on this basic principle[68, 3].

The classical vibrating MEMS gyroscopes use the tuning-fork configuration [75]. As it is shown in Figure 2.3, (b) the tuning-fork is comprised of a pair of proof masses which oscillate in opposite directions but with the same amplitude. When the structure starts to rotate, the proof masses are driven at resonance along the x-axis. Then the Coriolis acceleration produced by rotation around the z-axis is sensed capacitively along the y-axis [4, 75]. The Coriolis force is proportional to the applied angular rate, from which displacements can be measured.

Magnetometers measure the strength of the magnetic field in units of Gauss [76]. Magnetometers are often used in conjunction with accelerometers and gyroscopes to improve the human movement sensing [77]. The main challenge for using magnetometers to collect the human movement

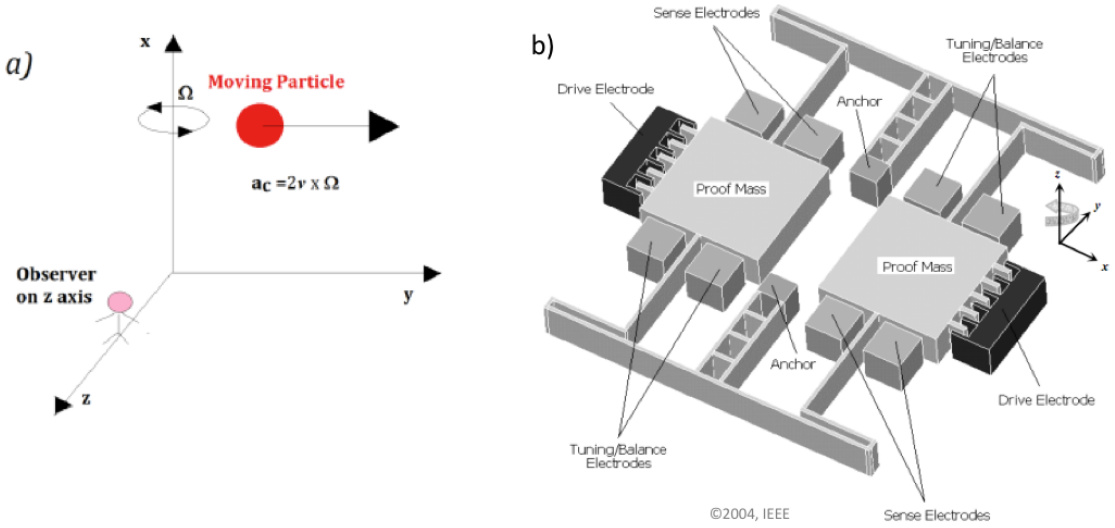

Figure 2.3: (a) The Coriolis acceleration [3], (b) The schematic of MEMS gyroscope with the tuning-fork configuration [4].

data is their high sensitivity to the magnetic field produced by other surrounding electronic devices [78].

Although body-worn sensors provide the possibility of collecting human movement data in real-life settings, the performance of these sensors highly depends on their deployment on body [79]. The position and the number of embedded wearable sensors have a decisive role in collecting high-resolution movement data [80, 81]. While attaching multiple sensors on one limb provides higher resolution data, it is burdensome for individuals with ASD who suffer from sensory sensitivities and for PD patients who have movement disability [82, 31]. Thus, finding an optimal placement to attach wearable sensors on human body parts, specifically in the clinical applications is crucial [82] and it is a subject of interest in many research studies [83, 84].

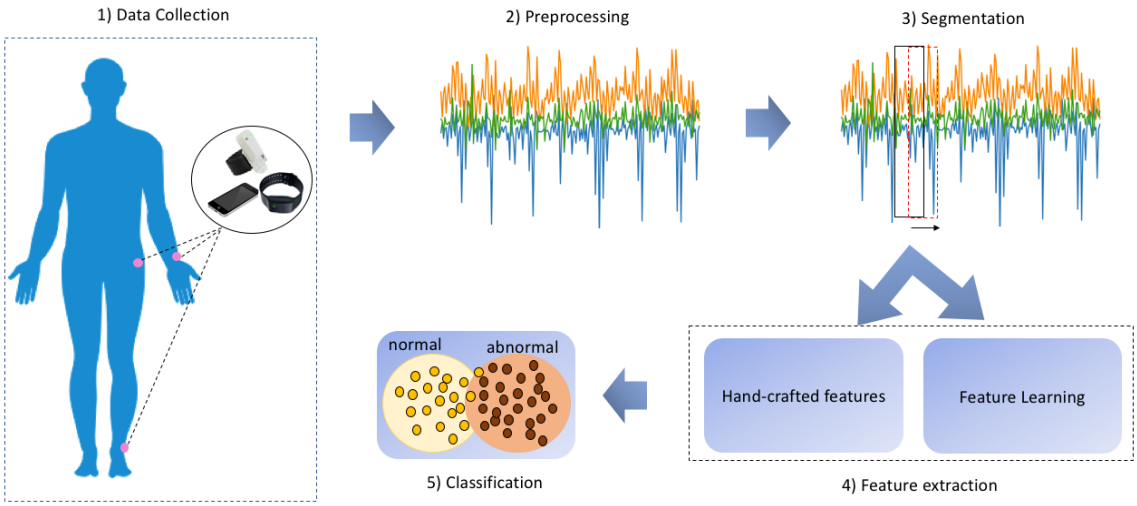

Figure 2.4: A general pipeline for abnormal movement detection using wearable sensors.

2.4 Pipeline for IMU Data Analysis

As illustrated in Figure 2.4, the general pipeline for human movement analysis using wearable sensors is comprised of 5 steps [11]: i) data collection; ii) preprocessing; iii) data segmentation; iv) feature extraction, and v) classification. In the following, we will briefly explain these steps with the main focus on abnormal movement detection.

In the data collection phase, the stream of raw data are captured using sensors attached to different parts of the body. Then in the preprocessing step, using the band-passed filtering in the frequency domain, the noise and the artifact in the signal are suppressed. Furthermore, if the data are recorded using different sensor types, the sampling rates of data are equalized to a fixed frequency rate using up-sampling or down-sampling techniques. The preprocessed data then are segmented into overlapping fixed-length intervals using the sliding-window technique. Finding the optimal length of the sliding window is a research challenge as it highly depends on the characteristics of the under-study movement, and it highly affects the performance of the system [11, 85]. The window length can be

selected based on the duration of the activity or the motion of interest. For example, for daily life activities, where there is a transition between short and long activities, a sliding window with the size around 1.5 seconds was found to be a close-to-optimal choice [86]. However, for the long periodic activities such as walking or running, the longer window size up to 32 seconds is more appropriate [87].

The segmented data are then fed to the feature extraction step. In this step, the key information from the movement data within each segment is extracted. In fact, the feature extractor maps the segmented raw data to a set of informative features for further analysis. Features can be extracted in time or frequency domain. The most popular time and frequency domain features extracted from human movement data are mean, variance, zero-crossing, energy, entropy, and fast Fourier transform peaks [5, 74, 19]. Despite the popularity of using these hand-crafted features [88], manual feature extraction suffers from restricting shortcomings: First, this approach relies on the expert's domain knowledge thus the important information in the signal may not be adequately extracted. Second, extracting these features in some cases is a time-consuming process [88, 38] that restricts their application in real-time scenarios. To address these issues, automatic feature learning methods have been recently proposed. These methods are based on training a deep neural network such as convolutional neural networks or autoencoders on the raw signal data to learn a set of discriminative features [37, 89, 90].

The last step of the IMU data analysis is the classification. In this phase, the learned/extracted features are fed to a classifier to differentiate the normal samples from the abnormal ones. Many research studies for abnormal movement detection have focused on applying supervised learning methods [39, 90, 19, 91]. These methods rely on maximizing the likelihood [92], $P(\mathbf{y}|\mathbf{X}, \theta)$, where $\mathbf{X} \in \mathbb{R}^{n \times p}$ is a given feature matrix for n samples and

p features, θ are the parameters of the classifier that can be learned by maximizing the negative-log likelihood or alternatively by minimizing empirical risk, and $\mathbf{y} \in \{0, 1\}^n$ is the corresponding target class vector for the normal and abnormal samples.

2.5 Performance Evaluation

Evaluating the performance of an abnormal movement detection system is a crucial task, as it helps to find the best parameters and components in each step of the IMU data analysis pipeline. Since data collected from ASD and PD patients are highly unbalanced, the prediction accuracy alone cannot be used as a measure for the true performance of a classifier [11]. In this case, other performance metrics such as F1-score, precision, recall, and the area under the receiver operating characteristic (ROC) curve are more appropriate metrics. In the following, the metrics that are used for performance evaluation in this thesis are briefly described.

F1-score is known as a popular performance metric when dealing with the skewed class distribution. It is a robust measure against the unbalance of test sets as it uses a harmonic average of precision and recall [93]:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

where precision and recall are respectively computed by $\frac{TP}{TP+FP}$ and $\frac{TP}{TP+FN}$. In the abnormal movement detection context, the true positive (TP) is the number of abnormal movement samples that are correctly determined as abnormal samples in the test phase. In contrast, false positive (FP) is the number of normal movement samples that are wrongly assigned to the abnormal movement class, and false negative (FN) is defined as the number of abnormal movement samples that are wrongly identified as normal movement samples.

The Area under the ROC curve (AUC) provides a summary measure of the relationship between the true positive rate and false positive rate [94] when changing the decisive threshold on the probability of each sample being an abnormal sample [11, 94]. The AUC is also a robust measure against the class imbalance.

2.6 Deep Neural Networks

Recent advances in deep learning extend applications of multilayer perceptron (MLP) theory by providing the possibility of training more than two hidden layers using backpropagation [95]. Applications of deep neural networks have recently grown in different research domains ranging from speech recognition [96, 97], to medical diagnosis applications [98] and robotics [99]. In the following, basic concepts related to the deep learning approaches that are used in this thesis are briefly reviewed.

2.6.1 Convolutional Neural Networks

Convolutional neural networks (CNN) are a type of neural networks which are suitable to be applied to data that have a grid-like topology such as time-series and images [100]. A typical CNN has a hierarchical architecture which alternates a convolution layer, a rectified linear unit (ReLU) activation function, and a pooling layer to summarize the large input spaces into a lower dimensional feature space. In the following, we briefly describe these parts and their functionality.

Convolution layer: depending on the structure of data 1D, 2D, or 3D convolutional layers might be used in a CNN architecture. As an example in a 2D-CNN case, a convolution layer receives the input 2D-data, I, convolves it with a set of K 2D-filters/kernels with the size of $m \times n$, and produces a feature map S. A discrete convolution operation is defined as follows:

$$S(i,j) = (I * K)(i,j) = \sum_m \sum_n I(m,n)K(i-m, j-n)$$

Rectified linear unit: in traditional neural networks, sigmoid or tangent hyperbolic activation functions are used to provide the non-linearity in the network. These activation functions, however, cause the vanishing gradients problem in backpropagation when the number of hidden layers increases. The *rectified* linear unit activation function provides the non-linearity in the network while it alleviates gradient vanishing problem, thus provides the possibility to use backpropagation for training very deep networks. For a given input x, the rectified linear unit function is defined as $max(0, x)$.

Pooling layer: pooling operation performs a sub-sampling of the feature map using a summary statistic such as average or maximum within a rectangular neighborhood [100]. Max-pooling and average-pooling are the two commonly used pooling functions. Figure 2.5 provides an illustration for the 1-D max pooling operation. The pooling operation provides robustness over shifts and distortions in the input space. The pooling layer reduces the spatial size of the input space, resulting in the reduction of the number of parameters, thus, the computational cost of training and prediction processes.

CNNs provide several advantages over traditional fully-connected neural networks including the possibility for sparse connectivity, parameter sharing, and equivariant representation:

Sparse connectivity: unlike the traditional neural networks with fully connected architecture, where each output unit in a given layer is connected to all input units of the next layer, CNNs benefit from sparse connectivity architecture by using convolutional kernels which have a much smaller size than the data. This property results in reducing the memory requirements

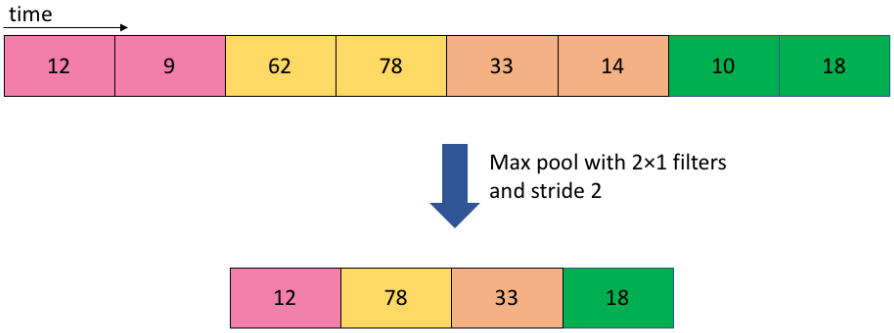

Figure 2.5: An illustration of the 1-D max pooling functionality on the time-series data.

for storing the parameters and improving the computational efficiency of the model [100].

Parameter sharing: in fully-connected neural networks, each parameter in the weight matrix is used only once when computing the output of each layer. In contrast, CNNs benefit from parameter sharing where instead of using different kernels for different locations, the same kernel can be used in all positions throughout the input data. Same as sparse connectivity, parameter sharing is also important in improving the computational efficiency of the network when dealing with large input spaces [100].

Equivariant representation: a function is considered as equivariant to a specific transformation if any changes in the input by means of that transformation causes the equivalent changes in the output. The convolution process is equivariant to translation in the input space. This is a very important property when dealing with the structured dynamic signal in space or time.

2.6.2 Long Short-Term Memories

Recurrent neural networks (RNNs) are a class of neural networks for processing the sequential data [100]. In general, RNNs benefit from the sharing of parameters through a very deep computational graph, as at each

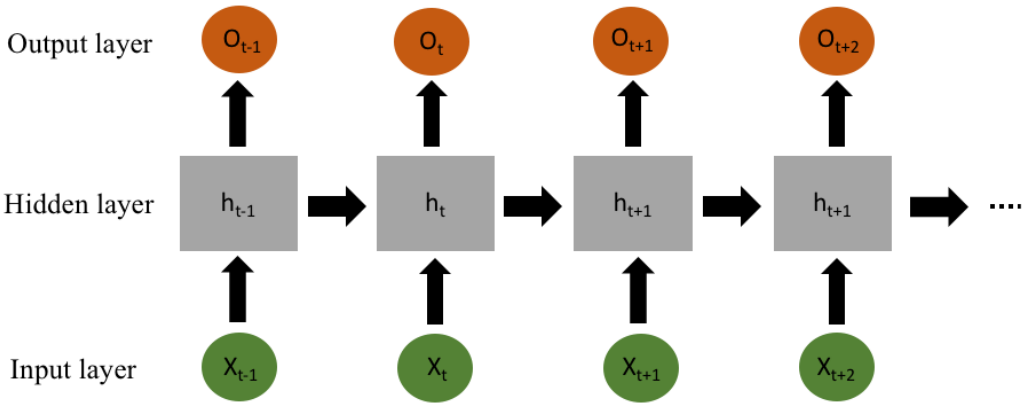

Figure 2.6: A schematic illustration of an RNN.

time step t, the subsequent hidden state from the previous state h_{t-1} and the input of the current state x_t are used to produce the output o_t (see Figure 2.6). RNNs are suitable when dealing with short-term dependencies. In long-term dependencies, however, they suffer from the problem of vanishing or exploding gradients during the training [100].

In 1997, in order to overcome this problem, Sepp Hochreiter [101] proposed an extension of RNNs called long short-term memories (LSTMs). LSTMs are able to capture the long-term dependencies and nonlinear dynamics by using a forget gate which enables the network to restart its state [100, 102]. LSTMs are comprised of memory cells that are connected to each other recurrently [100]. The state of each memory cell is updated according to the activation of the gates. As it is shown in Figure 2.7, each memory cell includes four elements: an input gate, a cell state unit, a forget gate, and an output gate. The weight of the cell state is controlled by the forget gate. In the training phase, the forget gate learns how much of information in the old state should be remained or forgotten. The input gate and output gate control the flow of information into the network. These mechanisms are more formally explained in Section 4.2. LSTMs have been successfully applied in a wide variety of applications, such as machine translation [103], speech recognition [104, 105], and human activ-

ity recognition [89].

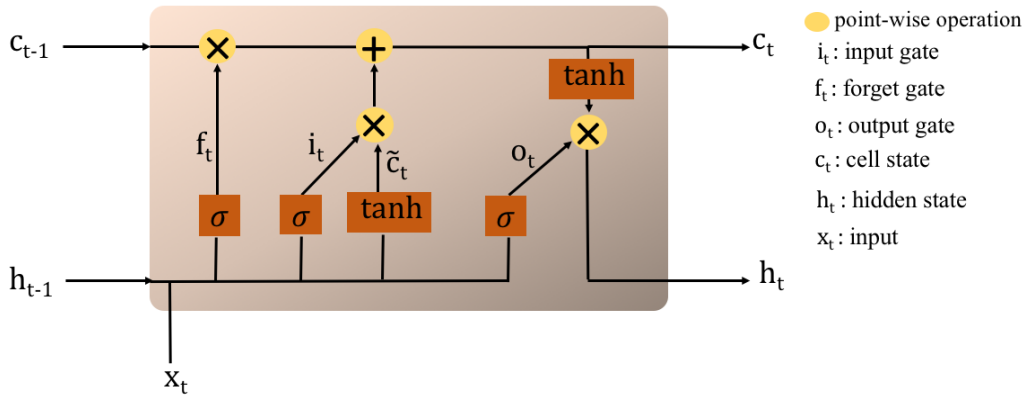

Figure 2.7: A schematic representation of a memory cell in LSTM (the picture from `http://colah.github.io/posts/2015-08-Understanding-LSTMs`).

2.6.3 Autoencoders

Autoencoders are neural networks which are trained to regenerate the input data in the output. An autoencoder is comprised of two parts: an encoder $f(x)$ which transforms the large input space to a lower dimensional feature space (also called latent codes), and a decoder $g(f(x))$ which reconstructs the input data from the latent codes in the output. Figure 2.8 shows a schematic representation of the general architecture of an autoencoder.

Autoencoders are trained to minimize a loss function, $L(x, g(f(x))$, i.e., the error between the actual input and the reconstructed input. Thus, during training, the network learns a new representation of data from the reduced latent space. This makes autoencoders successful in unsupervised feature learning, dimensionality reduction, and information retrieval applications [106, 107]. Denoising autoencoder is a variation of autoencoder that aims to learn more robust features by minimizing $L(x, g(f(\tilde{x})))$ where $\tilde{x}$ is the input data that is corrupted by some noise.

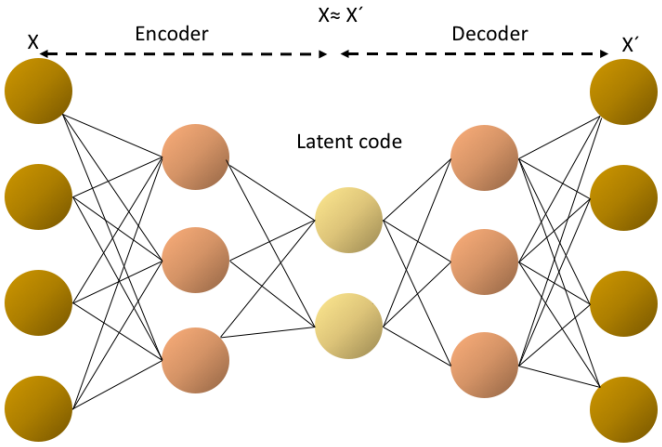

Figure 2.8: A general architecture for an autoencoder.

Chapter 3

Convolutional Neural Networks for Stereotypical Motor Movement Detection

3.1 Introduction

As stated in Section 2.1, stereotypical motor movements (SMM) have negative effects on the quality of life of children with autism. Considering the negative consequences of abnormal movements on the quality of life of autistic children and the limitations of existing methods for measuring

This chapter is based on the following publications:

1. Nastaran Mohammadian Rad, et al., "Convolutional neural network for stereotypical motor movement detection in autism." Accepted in NIPS 5th workshop of machine learning and interpretation in neuroimaging, 2015.

2. Nastaran Mohammadian Rad, and Cesare Furlanello, "Applying deep learning to stereotypical motor movement detection in autism spectrum disorders." Accepted in 16th International Conference on Data Mining Workshops (ICDMW), 2016.

3. Nastaran Mohammadian Rad, et al., "Deep learning for automatic stereotypical motor movement detection using wearable sensors in autism spectrum disorders." Accepted in Signal Processing, 2018.

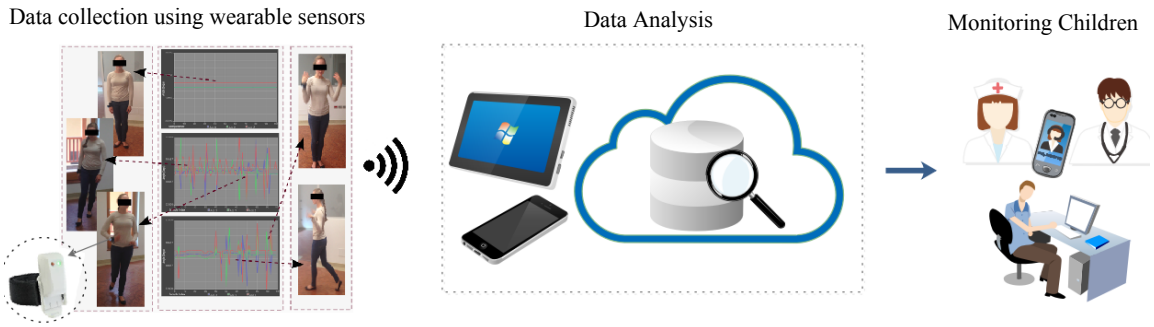

Figure 3.1: A real-time automatic SMM detection system. Inertial Measurement Units (IMUs) can be used for data collection. The collected data can be analyzed locally or remotely to detect SMMs. In case of detecting abnormal movements, an alert is sent to a therapist, caregiver, or parents.

SMMs, it is essential to develop time efficient and accurate methods for automatic SMM detection. Developing a real-time SMM detection and quantification system would be advantageous for researchers, caregivers, families, and therapists. Such a system would provide a powerful tool to evaluate the adaptation of subjects with autism spectrum disorder (ASD) to diverse life contexts within an ecologic approach. In particular, it helps to mitigate the meltdown behaviors that are anticipated by the increase in atypical behaviors. Any automatic quantification of atypical movements would indeed help caregivers and teachers to defuse the mechanism leading to stereotyped behaviors by involving children in specific activities or social interactions. Such involvement decreases the frequency of SMMs and gradually alleviates their duration and severity [108, 109]. A real-time implementation of SMM detection system (see Figure 3.1) would help therapists to evaluate the efficacy of behavioral interventions.

One of the important challenges for accurate SMM detection is to extract a set of effective and robust features from the inertial measurement unit (IMU) signal. As in many other signal processing applications, SMM detection is commonly based on extracting handcrafted features from the

IMU signals. So far, a wide variety of feature extraction methods have been used in the literature. Generally, two main types of features are extracted from the accelerometer signal [110]: i) time domain features, ii) frequency domain features. For time domain features, some statistical features such as mean, standard deviation, zero-crossing, energy, and correlation are extracted from the overlapping windows of the signal. In the case of frequency features, the discrete Fourier transform is used to estimate the power of different frequency bands. In addition, the Stockwell transform [111] is proposed by [19] for feature extraction from 3-axis accelerometer signals in order to provide better time-frequency resolution for non-stationary signals. In spite of their popularity, manual feature extraction and selection suffer from subjectivity and time inefficiency [88] that restrict the performance and also the application of SMM detection systems in real-time scenarios.

Another challenge toward developing a real-time SMM detection system is personalization due to the intra and inter-subject variability [19, 83]. This challenge, despite its crucial importance, has been undervalued [19]. Intra-subject variability is mainly due to the high variability in the intensity, duration, frequency, and topography of SMMs in each individual with ASD. Inter-subject differences are defined by the same variability across different individuals. The existence of these two types of variability within and across ASD persons motivates the necessity of developing an adaptive SMM detection algorithm that is capable to adjust to new patterns of behaviors.

3.1.1 Contribution

In this chapter, we present three main contributions: 1) robust feature learning from multi-sensor IMU signals; 2) enhancing the adaptability of SMM detection system to new data via parameter transfer learning. 3)

presenting a simulated dataset for SMM detection purpose using IMU sensors.

To achieve our first goal, we propose a new application of the deep learning paradigm in order to directly *learn* discriminating features for detecting SMM patterns. In particular, we use a convolutional neural network (CNN) [97] to bypass the commonly used feature extraction procedure. The idea of the CNN is inspired by the visual sensory system of living creatures [112]. Following this idea, LeCun et al. [113] developed a deep CNN architecture to address a pattern recognition problem in computer vision. Having fewer connections and parameters due to the weight sharing property, CNNs are easier to train compared to other deep neural networks. Currently, CNN solutions are among the best-performing systems on pattern recognition systems specifically for the handwritten character [113] and object recognition [114]. Beyond audio and image recognition systems, CNNs are successfully applied to various types of signals. Mirowski et al. [115] applied CNN on EEG signals for seizure detection. In the domain of psychophysiology, for the first time, Martinez et al. [88] proposed a model based on CNN to predict affective states of fun, excitement, anxiety, and relaxation. Their proposed model was tested on skin conductance and blood volume pulse signals. Recent studies also show the advantageous of applying CNN on accelerometer signals for human activity recognition [116, 117, 118].

To fulfill our second goal, we employ the parameter transfer learning by pre-initializing the parameters of the CNN [119]. We hypothesize that this capability can be used to transfer the prior knowledge regarding the distribution of parameters from one dataset to another dataset that are collected in a longitudinal study. If successful, our method can be employed in order to enhance the adaptability of the SMM detection system to new unseen data, thus facilitates its applications in wild real-world scenarios.

As a third contribution, we present a simulated dataset for SMM detection purpose. The dataset contains a complete 9-axes IMU signal, i.e., gyroscope and magnetometer sensor streams are also available in addition to the accelerometer signal, thus, provides a benchmark for developing multi-modal SMM detection systems.

Our experimental results on a simulated dataset and two real datasets show that feature learning via the CNN outperforms handcrafted features in SMM classification. Furthermore, it is shown that the parameter transfer learning is beneficial in enhancing the SMM detection rate when moving to a new dataset.

3.2 Methods

3.2.1 Notation

Let $\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_c \in \mathbb{R}^L$ be c time-series of length L that are recorded by a set of IMUs (e.g., accelerometer, gyroscope, and magnetometer sensors) at the sampling frequency of ν Hz. Thus, $T = L/\nu$ represents the length of the signal in seconds. We refer to each $\mathbf{S}_i$ as a data channel. Now consider $\mathbf{X}_t \in \mathbb{R}^{c \times \nu}$ for $t \in \{1, 2, \ldots, T\}$ as a sample in the raw feature space that is constructed by concatenating time-series of c data channels in a given time t. Let $y_t \in \{0, 1\}$ be the label associated to $\mathbf{X}_t$ where $y_t = 1$ corresponds to an SMM sample. In this text, we use boldface capital letters to represent matrices, boldface lowercase letters to represent vectors, and italic lowercase letters to represent scalars. We represent the matrix and element-wise multiplication between $\mathbf{A}$ and $\mathbf{B}$ matrices by $\mathbf{A} \cdot \mathbf{B}$ and $\mathbf{A} \odot \mathbf{B}$, respectively. Further, $[\mathbf{a}, \mathbf{b}]$ represents vector concatenation operation between $\mathbf{a}$ and $\mathbf{b}$ vectors.

3.2.2 Feature Learning via Convolutional Neural Network

The goal of an SMM detector is to predict the probability of being an SMM for a given sample $\mathbf{X}_t$, i.e., $P(y_t = 1 \mid \mathbf{X}_t)$. While the raw feature space ($\mathbf{X}_t$) is sensitive to intra and inter-subject differences, feature learning can provide the possibility to learn a new feature space that is robust over time and across different subjects. The aim of *feature learning* is to learn a linear or non-linear mapping function $\mathcal{F} : \mathbf{X}_t \mapsto \mathbf{x}'_t$, where $\mathbf{x}'_t \in \mathbb{R}^d$ is called the learned feature space. Then a classifier can be used in the learned feature space to estimate $P(y_t = 1 \mid \mathbf{x}'_t)$.

Convolutional Neural Networks (CNNs) offer an effective infrastructure for feature learning. CNN benefits from invariant local receptive fields, shared weights, and spatio-temporal sub-sampling features to provide robustness over shifts and distortions in the input space [97]. A classic CNN has a hierarchical architecture that alternates convolutional and pooling layers in order to summarize large input spaces with spatio-temporal relations into a lower dimensional feature space. A 1D-convolutional layer receives the input signal $\mathbf{X}_t \in \mathbb{R}^{c \times \nu}$, convolves it with a set of f filters with the length of m, $\mathbf{W} \in \mathbb{R}^{f \times c \times m}$, and produces a feature map $\mathbf{M}_t \in \mathbb{R}^{f \times \nu}$:

$$\mathbf{M}_t = \mathbf{X}_t * \mathbf{W} = \begin{bmatrix} \sum_{j=1}^{c} \sum_{i=1}^{m} w_{1,j,i} \times x_{j,1+i} & \dots & \sum_{j=1}^{c} \sum_{i=1}^{m} w_{1,j,i} \times x_{j,\nu+i} \\ \sum_{j=1}^{c} \sum_{i=1}^{m} w_{2,j,i} \times x_{j,1+i} & \dots & \sum_{j=1}^{c} \sum_{i=1}^{m} w_{2,j,i} \times x_{j,\nu+i} \\ \vdots & \vdots & \vdots \\ \sum_{j=1}^{c} \sum_{i=1}^{m} w_{f,j,i} \times x_{j,1+i} & \dots & \sum_{j=1}^{c} \sum_{i=1}^{m} w_{f,j,i} \times x_{j,\nu+i} \end{bmatrix} \tag{3.1}$$

where $*$ represents the convolution operator. The feature map is then fed to an activation function, generally the rectified linear unit (ReLU), to add non-linearity to the network and also to avoid the gradient vanishing problem [100], where:

$$\mathbf{M}_t^+ = max(\mathbf{0}^{f \times \nu}, \mathbf{M}_t). \tag{3.2}$$

Here $max(.,.)$ represents the element-wise max operation. Finally, in order to reduce the sensitivity of the output to shifts and distortions, $\mathbf{M}_t^+$ is passed through a pooling layer which performs a local averaging or sub-sampling over a pooling window with size of p elements and calculates the reduced feature map $\mathbf{M}'_t \in \mathbb{R}^{f \times \frac{\nu}{u}}$. In fact, a pooling layer reduces the resolution of a feature map by factor of $\frac{1}{u}$ where u is the stride (or step) size. Max-pooling and average-pooling are two commonly used pooling functions which compute the maximum or average value among the values in a pooling window, respectively. In average-pooling for $m'_{i,j} \in \mathbf{M}'_t$, $i \in \{1, \ldots, f\}$, and $j \in \{1, \ldots, \frac{\nu}{u}\}$, we have:

$$m'_{i,j} = \frac{1}{p} \sum_{k=1}^{p} m_{i,(j-1)\times u+k}. \tag{3.3}$$

Alternatively, in the max-pooling each element of the reduced feature map is the maximum value in a corresponding pooling window:

$$m'_{i,j} = max(m_{i,(j-1)\times u+1}, m_{i,(j-1)\times u+2}, \ldots, m_{i,(j-1)\times u+p}). \tag{3.4}$$

The reduced feature map $\mathbf{M}'_t$ can be used as the input to the next convolutional layer, i.e., $\mathbf{X}_t$ of the next layer. In general, the reduced feature map computed by stacking several convolutions, ReLU, and pooling layers is flattened as a vector before the classification step. The flattening step is performed by collapsing the rows of $\mathbf{M}'_t$ in the form of a vector. The resulting vector is called the learned feature space $\mathbf{x}'_t$ that represents a new representation of the original feature space. This new representation is typically fed to a fully connected neural network followed by a softmax layer for the classification purposes.

In this study, and for the purpose of SMM detection on the multi-sensor IMU data, we propose to use a three-layer CNN to transform the time-series of multiple sensors to a new feature space. The proposed architecture is shown in Figure 3.2. Three convolutional layers are designed to have 4, 4,

and 8 filters with the length of 9 samples (i.e., 0.1 seconds), respectively. The length of the pooling window and pooling stride are fixed to 3 ($p = 3$) and 2 ($u = 2$), respectively. The pooling stride of 2 reduces the length of feature maps by the factor of 0.5 after each pooling layer. The output of the third convolutional layer after flattening provides the learned feature vector. Then, the learned feature vector is fed to a two-layer fully-connected network with 8 and 2 neurons that are connected to a softmax layer. A dropout [120] rate of 0.5 is used in the fully connected layers to avoid the overfitting problem. Since only the information in $\mathbf{X}_t$ is used to compute $\mathbf{x}'_t$ and then predict y_t, we refer to the learned feature space via this CNN architecture as the static feature space. In fact, we do not consider the information in the consecutive time intervals as we assume that the sequence of samples over time are independent and identically distributed (i.i.d).

3.2.3 Parameter Transfer Learning via Network Pre-initialization

The quality and characteristics of recorded IMU signals vary not only from subject to subject but also from time to time in a single subject. Therefore it is important for the SMM detector system to adapt to new streams of signals in longitudinal scenarios. In this study, we explore the possibility of parameter transfer learning via network pre-initialization in order to transfer the learned patterns to the newly seen data in a different time span. In this direction, we first formalize the background theoretical concepts.

In the statistical learning theory, the goal is to learn a task $\mathcal{T}$ in a certain domain $\mathcal{D}$. A *domain* $\mathcal{D} = \{\mathcal{X}, \rho_{\mathcal{X}}\}$ is defined as a possible conjunction between an input space $\mathcal{X}$ and a marginal probability distribution $\rho_{\mathcal{X}}$. For example in the SMM detection context, the recorded IMU signal for different subjects can be considered as different domains as the marginal probability distribution $\rho_{\mathcal{X}}$ is different from one subject to another. Similarly, different domains can be defined by time in longitudinal data collection

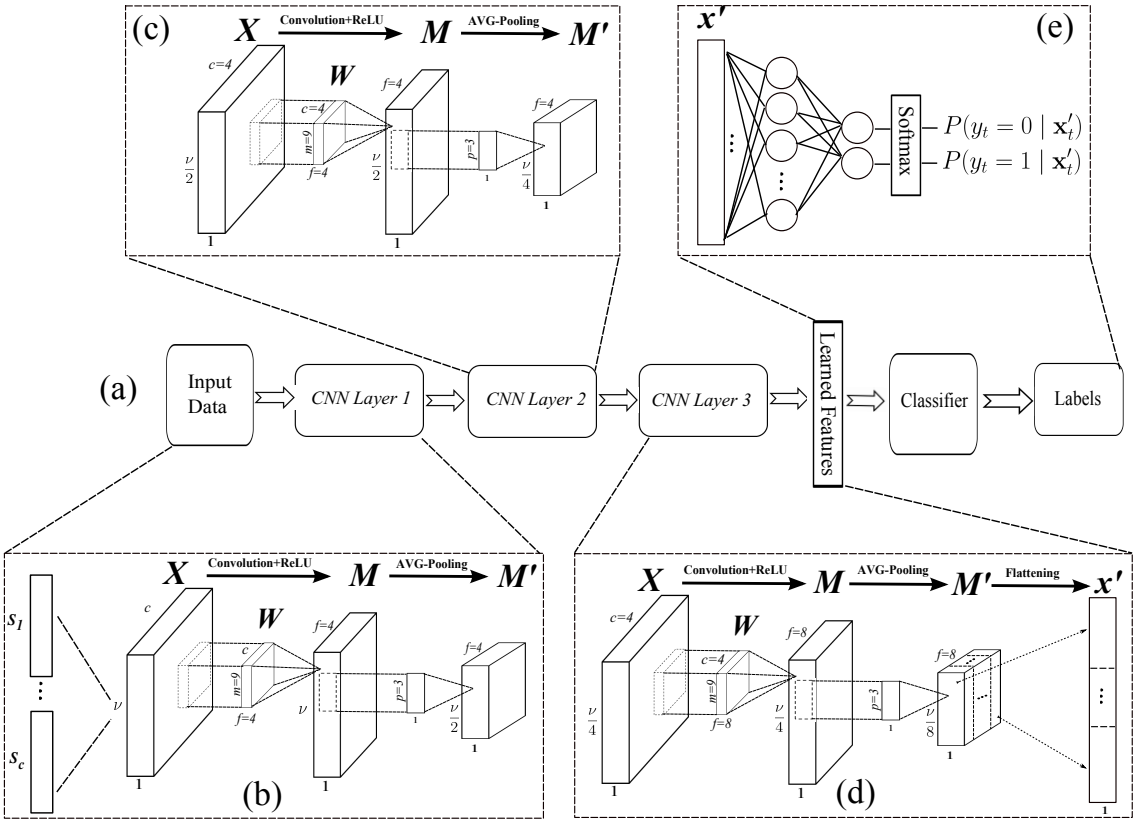

Figure 3.2: (a) The proposed architecture for SMM detection in the static feature space using a three-layer CNN. (b) The first CNN layer. This layer receives the one-second long time-series of several IMU sensors at time t, i.e., $\mathbf{X}_t$, and transfer it to the first level reduced feature map $\mathbf{M}'_t$. (c) The second CNN layer that uses the first level reduced feature map $\mathbf{M}'_t$ as its input, and transfer it to the second-level reduced feature map. (d) The third CNN layer. The reduced feature map of this layer is reshaped to the learned feature vector $\mathbf{x}'_t$ using the flattening operation. (e) The learned feature vector is fed to a fully-connected followed by a softmax layer to classify the samples to SMM and no-SMM classes.

scenarios. Given a domain $\mathcal{D}$, a *task* $\mathcal{T} = \{\mathcal{Y}, \Phi\}$ is defined as a predictive function Φ from $\mathcal{D}$ to the output space $\mathcal{Y}$. For example, in this study, Φ is the SMM detector, and $\mathcal{Y}$ represents the categorical output space of SMM and no-SMM samples. Assume $\mathcal{D}_S$, $\mathcal{D}_T$, $\mathcal{T}_S$, and $\mathcal{T}_T$ to represent the source domain, target domain, source task, and target task, respectively. *Transfer learning* aims to benefit from the knowledge in the source domain and

task in order to improve the predictive power in the target domain when $\mathcal{D}_S \neq \mathcal{D}_T$ or $\mathcal{T}_S \neq \mathcal{T}_T$ [121].

In this study, we are interested in the application of *parameter transfer learning* via pre-initializing the parameters of a deep neural network, as a well-established technique in the deep learning community, to improve the SMM prediction performance across different subjects and time intervals. To this end, we define the source domain $\mathcal{D}_S$ as the IMU signal which is collected on several subjects at the time span T_1. Similarly, the target domain $\mathcal{D}_T$ is defined as the IMU signal which is collected on several subjects at the time span T_2. Assume Φ_S be the learned predictive function, i.e., the CNN classifier, in the source domain. We use the learned parameters in Φ_S to pre-initialize the parameters of the predictive function in the target domain Φ_T. In simpler words, instead of random pre-initialization, we initialize the parameters of CNN classifier in the target domain with the learned CNN parameters in the source domain. We hypothesize that such a knowledge transfer via learned parameters improves the prediction performance in the longitudinal studies where the data are collected at different time intervals.

3.2.4 Experimental Materials

We assess the performance of the proposed methods on both simulated and real data. In the following, we describe the datasets and the procedures that are used for data preparation.

Simulated Data

In a simulation setting, 5 healthy subjects (3 females and 2 males) are asked to emulate stereotypical movements in a controlled environment. Each par-

ticipant wore an EXLs3 sensor[1], a miniaturized electronic device with the function of real-time IMU, fixed on the right wrist using a wristband (see Figure 3.3(a)). The EXLs3 sensor records three-axis accelerometer, gyroscope, and magnetometer data (it has 9 data channels in total). The sensor was set to transmit three-axis ± 16 g acceleration and ± 2000 dps angular velocity at the 100 Hz sampling rate. The participants were instructed to perform their normal working activities in the office such as sitting, writing, and typing; while intermittently performing hand flapping upon receiving a start/stop cue from the instructor (see Figure 3.3(b)-(e)). The total period of SMMs is organized somehow to keep the distribution of two classes comparable with real datasets where 27% of samples are in the SMM class (see Table 3.1). The total duration of each experiment was 30 minutes organized in three 10 minutes sessions. Real-time coding is undertaken during sessions to annotate the starting and ending time of movements. The captured data were band-pass filtered with a cut-off frequency of 0.1 Hz to remove the direct current (DC) components. Then the signal was segmented to 1 second long (i.e., 100 time-points) using a sliding window. The sliding window was moved along the time dimension with 10 time-steps resulting in 0.9 overlaps between consecutive windows [2].

[1]For the technical description see: `http://www.exelmicroel.com/elettronica_medicale-tecnologia-indossabile-exl-s3_module.html`.

[2]The collected simulated data is made publicly available at `https://gitlab.fbk.eu/MPBA/smm-detection`.

Figure 3.3: (a) The configuration of the EXLs3 sensor on the right hand. (b),(c) The simulated data are collected during daily work activities in the office (e.g., sitting, writing, and typing). (d),(e) The subjects are asked to intermittently perform hand flapping upon receiving a start/stop cue from the instructor.

Real Data

We use the data presented in [19] wherein the accelerometer data were collected from 6 subjects with autism in a longitudinal study[3]. As stated in [19], the aim of data collection in a longitudinal scenario was to investigate whether a previously trained classifier can accurately detect SMMs over time. The data were collected in the laboratory and classroom environments while the subjects wore three 3-axis wireless accelerometers and engaged in body rocking, hand flapping, or simultaneous body rocking and hand flapping. The accelerometer sensors were attached to the left and right wrists, and on the torso. Offline annotation based on a recorded video is used to annotate the data by an expert. Two separate collections are available: the first collection, here we call it *SMM-1*, was recorded by MITes sensors [122] at 60 Hz sampling frequency [123]. The second collection *SMM-2*, was recorded three years after the first recording on the same subjects using Wockets sensors with the sampling frequency of 90 Hz. In the preprocessing step, first, the sampling rate of two recordings is equalized by re-sampling the signal in *SMM-1* to 90 Hz using linear

[3]The dataset and full description of data are publicly available at `https://bitbucket.org/mhealthresearchgroup/stereotypypublicdataset-sourcecodes/downloads`.

interpolation. The cut-off high pass filter at 0.1 Hz is applied in order to remove the DC components of the signal. Similar to [19], the signal is segmented to 1-second long overlapped intervals using a sliding window. Then similar to [19], amount of overlap is set to 10 time-points resulting in 0.87 overlap between consecutive windows. Table 3.1 summarizes the number of samples in no-SMM and SMM classes for each subject. The difference in the number of samples in SMM and no-SMM classes shows the unbalanced nature of the real data, wherein SMM-1 and SMM-2 datasets 31% and 23% of samples are in the SMM class, respectively.

Table 3.1: Number of samples in SMM and no-SMM classes in three datasets.

Data	Subjects	No-SMM	SMM	All	SMM/All
Simulated Data	**Sub1**	13875	4075	17950	0.23
	Sub2	11686	6224	17910	0.35
	Sub3	13694	4246	17940	0.24
	Sub4	12428	5532	17960	0.31
	Sub5	13583	4367	17950	0.24
	Total	65266	24444	89710	**0.27**
SMM-1 [19]	**Sub1**	21292	5663	26955	0.21
	Sub2	12763	4372	17135	0.26
	Sub3	31780	2855	34635	0.08
	Sub4	10571	10243	20814	0.49
	Sub5	17782	6173	23955	0.26
	Sub6	12207	17725	29932	0.59
	Total	106395	47031	153426	**0.31**
SMM-2 [19]	**Sub1**	18729	11656	30385	0.38
	Sub2	22611	4804	27415	0.18
	Sub3	40557	268	40825	0.01
	Sub4	38796	8176	46972	0.17
	Sub5	22896	6728	29624	0.23
	Sub6	2375	11178	13553	0.82
	Total	145964	42810	188774	**0.23**

3.2.5 Experimental Setups and Evaluation

To investigate the effect of static feature learning and parameter transfer learning on the performance of SMM detection when balanced and unbalanced data is used in the training set, we conducted three experiments.

Keras library [124] is used in our implementations[4].

Experiment 1: Static Feature Learning

The main aim of this experiment is to compare the effectiveness of feature learning using a deep neural network versus raw feature space and handcrafted features in an across-subject SMM detection setting. To evaluate the effect of both feature extraction and feature learning on the SMM classification performance, first, without any feature extraction the signals in the raw feature space are used as the input to a support vector machine (SVM) classifier. In this case, all data channels of each sample $\mathbf{X}_t$ are collapsed into a feature vector (with length of $900 = 9 \times 100\ Hz$ in simulated data and $810 = 9 \times 90\ Hz$ in real datasets, where 9 is the number of data channels). Second, to evaluate the detection performance using handcrafted features we extracted all features mentioned in [19] including time, frequency, and Stockwell transform features, then, we replicated the across-subject SMM detection experiment in [19]. In this setting, we used exactly the same implementation provided by the authors [5] in the feature extraction and classification steps. Third, a CNN architecture (see Section 3.2.2) is used to learn an intermediate representation of the multisensor signal. In this experiment, all effective parameters of CNN (weights and biases) are initialized by drawing small random numbers from the normal distribution. The stochastic gradient descent with momentum (the momentum is fixed to 0.9) is used for training the network. All these steps are performed only on the training data to ensure unbiased error estimation. Due to the random initialization of weights and employing stochastic gradient descent algorithm for optimization, results can be different from one training run to another. Therefore, we repeated the whole procedure

[4]See `https://gitlab.fbk.eu/MPBA/smm-detection` to access the implemented scripts and codes.

[5]The code is available at: `https://bitbucket.org/mhealthresearchgroup/stereotypypublicdataset-sourcecodes/downloads`.

of learning and classification 10 times and the mean and standard variation over runs are reported. It is important to emphasize that, similar to [19], in all three parts of this experiment the number of samples in the minority class is used to randomly draw a balanced training set.

Experiment 2: Parameter Transfer Learning

As discussed before, deep neural networks provide the capability of parameter transfer learning via network pre-initialization. We applied this experiment only on two real datasets in order to investigate the possibility of transferring learned knowledge from one dataset to another in a longitudinal data collection setting. This experiment is similar to Experiment 1, except for the network initialization step. Instead of random initialization, here we first train the CNN on one balanced real dataset, e.g., SMM-1, and then we use the learned parameters for pre-initializing the parameters of CNN before training on another balanced real dataset, e.g., SMM-2. Similar to the previous experiment, we repeated the whole experiment 10 times to compute the standard deviation of the classification performance.

Experiment 3: Training on the Unbalanced Training Set

As explained, in Experiment 1 and 2, we balanced the training set based on the number of samples in the minority class. Even though balancing the training set improves the quality of the trained model, in fact, it suffers from some deficits: 1) by balancing the training set we impose a wrong prior assumption on the original distribution of data. As shown in Table 3.1 in real datasets around 0.3 of samples belong to SMM class, when by balancing the dataset we assume it is 0.5; 2) by balancing the training set we cannot employ the full richness of the data as we need to remove the significant amount of samples from the training set; 3) in some practical scenarios, such as real-time adaptation or classification on the sequence of

streamed data, balancing the training set is impractical. Considering these limitations, in this experiment in order to evaluate the effect of balancing on the performance of CNN model we evaluate the performance of the proposed CNN architecture in predicting SMMs when unbalanced training sets are used in the training phase.

Evaluation

In all experiments, the leave-one-subject-out scheme [6] is used for model evaluation in order to measure the robustness of the trained model against inter-subject variability. Due to the unbalanced class distributions in the test set, we computed the F1-score to evaluate the classification performance.

3.3 Results

3.3.1 Feature Learning Outperforms Handcrafted Features

The classification performances summarized in Table 3.2 compare the quality of feature learning via CNN with raw and handcrafted feature spaces on three datasets. For all three datasets, the classification performance of SMM detection on the handcrafted and learned features is higher than the classification performance on the raw feature space. This result demonstrates the importance of developing feature extraction/learning methods for detecting SMMs. Furthermore, the comparison between the results achieved by handcrafted and learned features illustrates the efficacy of feature learning over the manual feature extraction in SMM prediction. The learned feature space reaches on average 10% and 20% higher F1-score than

[6] Using the leave-one-subject-out cross-validation prevents the risk of overfitting as in this case all samples of all subjects except one subject have been used for training the model and then the trained model is evaluated on the samples of the remained subject.

the handcrafted features in case of simulated data and SMM-1, respectively, while in case of SMM-2 its performance declines by 5%. These results support the overall efficacy of feature learning versus handcrafted features in extracting robust features for across-subject SMM detection. In order to further confirm these findings, the different feature spaces were analyzed by projecting in a 2-dimensional space with the t-distributed Stochastic Neighbor Embedding (t-SNE) [125] technique (see Figure 3.4(a)-(c)). We used the average of Fisher's separability score [126] across two t-SNE dimensions to quantify the separability of samples in two classes for different feature spaces. Figure 3.4(a) shows 2D t-SNE distribution of SMM and no-SMM samples in the raw feature space, where there is a high overlap between the samples of two classes. This high overlap is also well-reflected in the low Fisher's separability score in raw feature space (0.02). Figure 3.4(b) depicts the distribution of samples of two classes in handcrafted feature space. The samples in two classes are barely separable and the Fisher's separability score is 0.03. Figure 3.4(c) displays the 2D t-SNE space for the learned features via the CNN architecture. In this case, the separability score is improved significantly to 0.10.

3.3.2 Parameter Transfer Learning is Beneficial in Longitudinal Studies

As mentioned in Section 3.2.5, the aim of our second experiment was to investigate the possibility of transferring learned knowledge from one dataset to another using parameter transfer learning. Our results in Table 3.2 shows that transferring knowledge from one dataset to another in a longitudinal study, by pre-initializing the parameters of CNN model improves the average classification performance of the SMM detectors by 4% and 21% in SMM-1 and SMM-2 datasets, respectively.

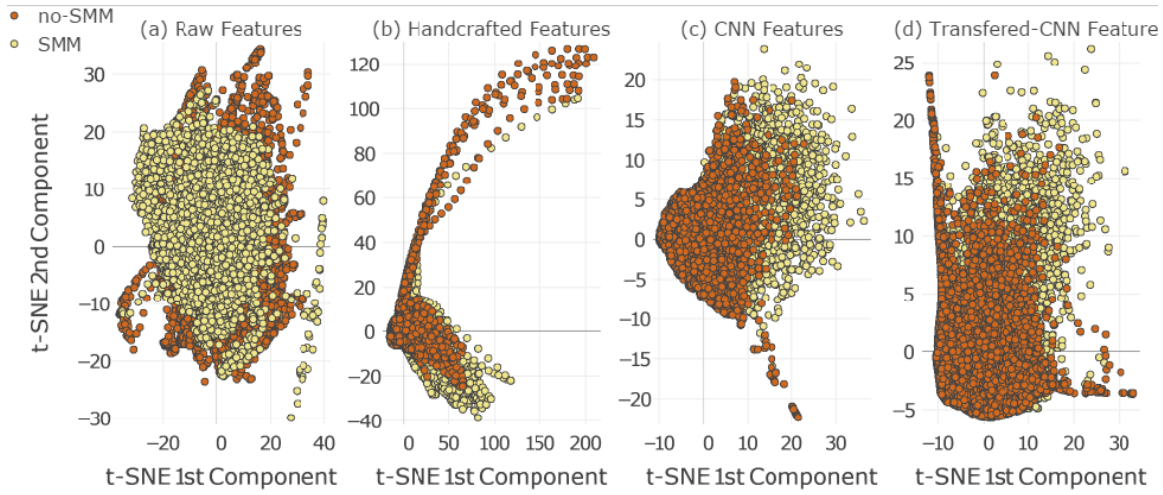

Figure 3.4: The distribution of SMM and no-SMM samples in the 2-dimensional t-SNE space for (a) raw feature space, (b) handcrafted features, (c) static feature space learned by CNN, and (d) static feature space learned by pre-initialized CNN. Feature learning increases the separability of samples in two classes compared to raw and handcrafted features.

3.3.3 Training on Unbalanced Data Decreases the Performance

The results in Table 3.2 illustrate the importance of balancing classes during training. Indeed a negative effect is observed when unbalanced training set is used in training CNN architecture in randomly initialized (feature learning) and pre-initialized (transfer learning) scenarios. The performance of SMM detection in feature learning scenario drops by 6% and 3% in SMM-1 and SMM-2 datasets, respectively. This performance drop is even more pronounced in the transfer learning scenario where we observe 8% and 20% performance drop in F1-score in the corresponding datasets.

In all experiments, we also observed very low performance for subject 3 in SMM-2 dataset which can be explained by the fact that for this subject the distribution of samples in SMM and no-SMM classes is highly skewed (see Table 3.1).

3.4 Discussion

3.4.1 Toward Real-Time Automatic SMM Detection

Developing real-time mobile applications for detecting abnormal movements such as SMMs can be considered as the final goal in the context of automatic SMM detection using wearable sensors. At the moment there are numerous challenges in real-time human activity recognition us-

Table 3.2: F1-score results for SMM detection using raw, handcrafted, and static feature spaces in three benchmarked datasets. The results show that feature learning (here referred to as Static-Features) generally outperforms raw and handcrafted feature spaces. In addition, the parameter transfer learning has a positive effect on the performance of the CNN classifier. Furthermore, training the CNN classifier on unbalanced training sets causes the performance drop in feature learning and transfer learning scenarios.

Data	**Sub**	**Balanced-training**				**Unbalanced-training**	
		Raw Features	**Handcrafted Features**	**Static-Feature Learning**	**Transfer Learning**	**Static-Feature Learning**	**Transfer Learning**
Simulated	**1**	0.29	0.71	0.78 ± 0.05	-	0.73 ± 0.13	-
	2	0.84	0.86	0.86 ± 0.03	-	0.78 ± 0.09	-
	3	0.55	0.76	0.80 ± 0.01	-	0.75 ± 0.13	-
	4	0.76	0.48	0.85 ± 0.03	-	0.73 ± 0.12	-
	5	0.38	0.77	0.79 ± 0.01	-	0.80 ± 0.04	-
	Mean	0.56	0.72	**0.82 ± 0.03**	-	0.76 ± 0.11	-
SMM-1	**1**	0.44	0.74	0.74 ± 0.02	0.71 ± 0.02	0.70 ± 0.02	0.71 ± 0.03
	2	0.32	0.37	0.75 ± 0.02	0.73 ± 0.01	0.63 ± 0.03	0.63 ± 0.04
	3	0.22	0.50	0.68 ± 0.04	0.70 ± 0.03	0.57 ± 0.08	0.59 ± 0.06
	4	0.44	0.73	0.92 ± 0.01	0.92 ± 0.00	0.88 ± 0.01	0.88 ± 0.01
	5	0.56	0.44	0.51 ± 0.04	0.68 ± 0.05	0.51 ± 0.08	0.58 ± 0.07
	6	0.56	0.46	0.90 ± 0.01	0.94 ± 0.01	0.79 ± 0.07	0.81 ± 0.09
	Mean	0.42	0.54	0.74 ± 0.03	**0.78 ± 0.03**	0.68 ± 0.06	0.70 ± 0.06
SMM-2	**1**	0.47	0.43	0.61 ± 0.11	0.68 ± 0.05	0.33 ± 0.14	0.36 ± 0.08
	2	0.23	0.26	0.20 ± 0.04	0.22 ± 0.04	0.11 ± 0.03	0.16 ± 0.04
	3	0.01	0.03	0.02 ± 0.01	0.02 ± 0.01	0.02 ± 0.01	0.02 ± 0.01
	4	0.32	0.86	0.72 ± 0.03	0.77 ± 0.02	0.71 ± 0.14	0.83 ± 0.03
	5	0.38	0.73	0.21 ± 0.09	0.75 ± 0.09	0.14 ± 0.09	0.09 ± 0.02
	6	0.50	0.07	0.36 ± 0.13	0.91 ± 0.05	0.62 ± 0.08	0.70 ± 0.16
	Mean	0.32	0.40	0.35 ± 0.08	**0.56 ± 0.05**	0.32 ± 0.1	0.36 ± 0.08

ing wearable sensors, namely [127, 128, 129]: 1) designing effective feature extraction and inference methods; 2) recognition on realistic data; 3) the adaptability of the system to new users. Addressing these issues demands a huge investment in research toward finding robust and effective features that can be extracted in a reasonable time from the stream of the IMU signal. Our proposal to learn an intermediate representation of the signal, that is robust to the signal variability of a single subject data over time and also to the across-subject variability, can be considered as an effective solution in this direction. In addition, the parameter transfer learning capability besides the possibility of incremental training (i.e., the continual training and extending the model's knowledge) of the proposed deep architecture facilitates the online adaptation of an automatic SMM detector in real-time scenarios. This finding overlooks the subject specific [130], and monolithic [131, 83] activity recognition systems, thus opening new frontiers toward adaptable activity recognition systems which are more appropriate for real-time usages.

3.4.2 Limitation and Future Work

Even though the deep architecture introduced in this study provides a significant step toward a more accurate automatic SMM detection system in real-time scenarios, it suffers from a considerable limitation: the proposed fully supervised scheme for training the SMM detection model is problematic for its online adaptation. This problem comes from the fact that in real applications the system has no access to the labels of incoming samples during usage by a new user. Therefore, the adaptation to new unseen data should be performed only based on the input unlabeled data. This limitation motivates the study of methods for online adaptation of the system in an unsupervised manner. One possible solution in this direction is transductive transfer learning [121] where the basic assumption is that

no labeled data in the target domain are available. Therefore adopting a transductive transfer learning strategy in the adaptation phase can be considered as a possible future direction to extend this work.

3.5 Conclusions

In this study, we proposed an original application of deep learning for SMM detection in ASD children using the accelerometer, and in general IMU, sensors. To the best of our knowledge, this is the first effort toward applying deep learning paradigm for detecting SMMs. Our experimental results showed that convolutional neural networks outperform the traditional classification on the handcrafted features. This observation supports our initial hypotheses about the effectiveness of embedded feature learning and transfer learning capabilities of deep neural networks in providing more accurate SMM detection systems. This study is an early effort toward developing a real-time SMM detector. Such a system can be embedded in a mobile-based application to provide the possibility of ubiquitous SMM detection.

Chapter 4

Stereotypical Motor Movement Detection in Dynamic Feature Space

4.1 Introduction

In Chapter 3, we proposed a convolutional neural network (CNN) architecture [97] in order to directly learn discriminating features for detecting stereotypical motor movements (SMM) in inertial measurement unit (IMU) data collected from children with autism. In fact, the CNN architecture was used to transform the multi-channel IMU signal to a reduced set of discriminative features. It was shown that the proposed CNN with the feature learning and transfer learning capabilities provides more accurate SMM detectors than manual feature extraction methods. Despite

This chapter is based on the following publications:

1. Nastaran Mohammadian Rad, et al., "Stereotypical motor movement detection in dynamic feature space." Accepted in 16th International Conference on Data Mining Workshops (ICDMW), 2016.

2. Nastaran Mohammadian Rad, et al., "Deep learning for automatic stereotypical motor movement detection using wearable sensors in autism spectrum disorders." Accepted in Signal Processing, 2018.

the superior classification performance, the CNN architecture proposed in Chapter 3 did not consider the temporally structured information from the IMU signal sequences. This is a general issue in SMM detection, in which individual segments of movement data in terms of IMU signals are treated as statistically independent samples. Therefore, the possible long-term dependencies stored in the longer time intervals of the signal are ignored in the detection process while by the definition, SMMs are repetitive movements. Thus, temporal patterns stored in the sequence of samples are expected to contain valuable information for SMM detection.

4.1.1 Contribution

In this chapter, we propose a deep architecture, by stacking a long short-term memory (LSTM) [101] layer on top of the CNN architecture proposed in Chapter 3, in order to learn the dynamic feature space on the sequence of IMU data. Furthermore, we employ an ensemble of the best base LSTMs to improve the accuracy and stability of results. LSTM as a type of recurrent neural networks (RNN) has been effectively used for learning long-term temporal dependencies in sequential data such as speech recognition [132], handwriting recognition [133], and natural language modeling [134]. It is shown that LSTM has been successfully used for human activity recognition as a classical sequence analysis problem. Ordonez and Roggen [89] applied the deep architecture, by combining CNN with LSTM network, on the human activity recognition problem. Their experimental results demonstrated a significant improvement upon the state-of-the-art. Elsewhere, Hammerla et al. [90] applied a different type of deep learning algorithms including CNN and forward/bi-directional LSTMs on three public human activity datasets captured by wireless sensors. Their experimental results showed the superiority of LSTMs on the activities with natural ordering in short time periods. Considering these studies in human activ-

ity recognition, it is expected that learning the temporal patterns stored in the consecutive samples of IMU data to provide higher accurate SMM detectors.

Our experimental results on one simulated dataset and two real datasets show that including the dynamics of recorded data improves the classification performance. We experimentally demonstrate the superiority of the proposed architecture over the CNN model presented in Chapter 3, when an unbalanced training set is used in the training phase. Our results show that the false-alarm rate is indeed reduced by inducing the right prior distribution into the classification task. This finding gives the possibility of employing the whole training set in the training phase by removing the need for balancing the number of samples in two classes, thus, facilitates the application of SMM detection systems in real-time scenarios.

4.2 Methods

Here, we use the notation presented in 3.2.1 and Figure 2.7 to explain the definition and concepts about LSTMs.

4.2.1 SMM Detection via LSTM

In the context of SMM detection using wearable sensors, the aim is to predict the probability of being an SMM for a given sample $\mathbf{X}_t$, i.e., $P(y_t = 1 \mid \mathbf{X}_t)$. Since only the information in $\mathbf{X}_t$ is used to predict y_t, we refer to this kind of detector as a static detector (see Section 3.2.2). By detecting SMMs using a static detector, it is implicitly assumed that the sequence of samples over time are independent and identically distributed (i.i.d). But in reality, this assumption is not valid as the samples in consecutive time steps are highly dependent. Therefore, it is expected that accounting for this dependency would improve the performance of the SMM detector.

Following this hypothesis, we propose to use a long short-term memory (LSTM) layer to model the temporal dependency between the consecutive time steps of the recorded signal.

Let $\mathbf{x}'_t$ be a set of static features that are extracted or learned from samples in the raw feature space (i.e., from $\mathbf{X}_t$). Here we assume the CNN architecture explained in Section 3.2.2 is used to compute $\mathbf{x}'_t$. Then, let $\mathbf{c}_t \in \mathbb{R}^q$ and $\mathbf{h}_t \in \mathbb{R}^q$ to represent the cell state and output of an LSTM unit at time step t, respectively, where q is the number of neurons in the LSTM unit. We will refer to $\mathbf{h}_t$ as the *dynamic* feature space. The LSTM unit receives $\mathbf{x}'_t$, $\mathbf{h}_{t-1}$, and $\mathbf{c}_{t-1}$ as its inputs, and computes $\mathbf{c}_t$ and $\mathbf{h}_t$ as follows:

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t, \tag{4.1}$$

$$\mathbf{h}_t = \mathbf{o}_t \odot (1 - e^{-2\times\mathbf{c}_t}) \odot (1 + e^{-2\times\mathbf{c}_t})^{-1}. \tag{4.2}$$

Here $\mathbf{f}_t \in \mathbb{R}^q$ is called the forget gate vector and its elements are real numbers between 0 and 1 that decide how much information to be passed from $\mathbf{c}_{t-1}$ to $\mathbf{c}_t$. During the learning phase, the forget gate learns the forget weight matrix $\mathbf{W_f}$ and the forget bias vector $\mathbf{b_f}$. $\mathbf{f}_t$ is computed by

$$\mathbf{f}_t = (1 + e^{-(\mathbf{W_f}\cdot[\mathbf{h}_{t-1},\mathbf{x}'_t]+\mathbf{b_f})})^{-1}. \tag{4.3}$$

Using a tangent hyperbolic function, $\tilde{\mathbf{c}}_t \in \mathbb{R}^q$ provides new candidate values between -1 and 1 for $\mathbf{c}_t$ by learning $\mathbf{W_c}$ and $\mathbf{b_c}$:

$$\tilde{\mathbf{c}}_t = (1 - e^{-2\times(\mathbf{W_c}\cdot[h_{t-1},\mathbf{x}'_t]+\mathbf{b_c})}) \odot (1 + e^{-2\times(\mathbf{W_c}\cdot[h_{t-1},\mathbf{x}'_t]+\mathbf{b_c})})^{-1}, \tag{4.4}$$

where $\mathbf{i}_t \in \mathbb{R}^q$ is the input gate vector with elements between 0 and 1. These values determine the level of new information in $\tilde{\mathbf{c}}_t$ to be transferred to the cell state $\mathbf{c}_t$. $\mathbf{i}_t$ is computed based on $\mathbf{W_i}$ and $\mathbf{b_i}$ as follows:

$$\mathbf{i}_t = (1 + e^{-(\mathbf{W_i}\cdot[h_{t-1},\mathbf{x}'_t]+\mathbf{b_i})})^{-1}. \tag{4.5}$$

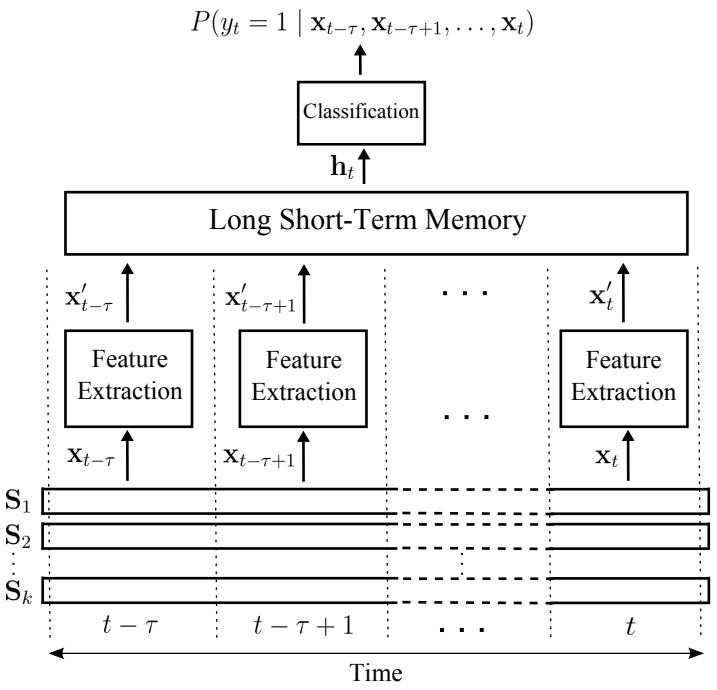

Figure 4.1: The proposed architecture for SMM detection in the dynamic feature space using a long short-term memory. Each feature extraction block contains a trained three-layer CNN architecture (see Figure 3.2). Each of $\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_k$ refer to a raw data channel.

Finally, $\mathbf{o}_t \in \mathbb{R}^q$ is the output gate vector that filters the cell state $\mathbf{c}_t$ to generate the output of the LSTM unit $\mathbf{h}_t$:

$$\mathbf{o}_t = (1 + e^{-(\mathbf{W_o} \cdot [h_{t-1}, \mathbf{x}'_t] + \mathbf{b_o})})^{-1}. \tag{4.6}$$

In this study, a fully-connected layer with dropout of 0.2 is used to transfer the output of the LSTM layer at time t, i.e., $\mathbf{h}_t$, to the input of the softmax layer $\mathbf{z}_t = [z_t^{(0)}, z_t^{(1)}]^T \in \mathbb{R}^2$:

$$P(y_t = 1 \mid \mathbf{x}_{t-\tau}, \mathbf{x}_{t-\tau+1}, \ldots, \mathbf{x}_t) = \frac{e^{z_t^{(1)}}}{e^{z_t^{(0)}} + e^{z_t^{(1)}}}, \tag{4.7}$$

where τ represents the number of previous time steps that are used as the input to the LSTM layer. Figure 4.1 presents a schematic overview of the proposed architecture.

4.2.2 Ensemble of the Best Base Learners

Due to the random initialization and using stochastic optimization algorithms on random mini-batches in training deep learning models, retraining the same model on the same training set results in heterogeneous approximations of the target classifier. This heterogeneity is the direct result of reaching different local optima in optimizing a complex non-convex error surface. One possible approach to overcome this problem is *ensemble learning* (EL) [135]. The main idea behind EL is to combine the knowledge learned by individual classifiers in order to achieve superior and stable performance. It is shown that in general, an ensemble of classifiers works better than every single classifier due to the statistical, computational, and representational reasons [136]. Considering the success of deep learning ensembles in pattern recognition and signal processing applications [137, 138, 7, 139], in this study we are interested in applying classifier selection voting approach [140] to combine an ensemble of the best base learners.

Let $(\mathbf{X}_{tr}, \mathbf{y}_{tr})$ and $(\mathbf{X}_{ts}, \mathbf{y}_{ts})$ to be the corresponding sample/target pairs in the training and test sets, respectively. Then assume $\mathcal{C} = \{c_1, c_2, \dots c_l\}$ be a set of l base learners trained on the training set. Our goal is to first find a set of b best classifiers $\mathcal{C}^* \subseteq \mathcal{C}$ based on a performance measure α on the training set, and then to combine their prediction on the test set using majority voting in the prediction phase. Algorithm 1 summarizes this approach.

4.2.3 Experimental Materials

Data and Preprocessing

We evaluate the performance of the proposed methods on both simulated and real data described in Section 3.2.4.

Algorithm 1 The training and test procedures in the majority voting on a set of b best models.

procedure TRAINING($\mathcal{C}$,$\mathbf{X}_{tr}$,$\mathbf{y}_{tr}$)
  for all $c_i \in \mathcal{C}$ **do**
    Predict $\hat{\mathbf{y}}$ using c_i on $\mathbf{X}_{tr}$.
    Evaluate $\hat{\mathbf{y}}$ and store the performance in α_i.
  end
  for $i \leftarrow 1, l$ **do**
    Store the best i classifiers in $\mathcal{C}_i^*$.
    Predict $\hat{\mathbf{y}}_1, \ldots, \hat{\mathbf{y}}_i$ using classifiers in $\mathcal{C}_i^*$ on $\mathbf{X}_{tr}$.
    Compute majority voting $\tilde{\mathbf{y}}$ on predictions in $\hat{\mathbf{y}}_1, \ldots, \hat{\mathbf{y}}_i$.
    Evaluate $\tilde{\mathbf{y}}$ and store the performance in α_i.
  end
  Find the best value for b by $b = \text{argmax}_i(\alpha_i)$.
  return $\mathcal{C}_b^*$.

procedure TEST($\mathcal{C}_b^*$,$\mathbf{X}_{ts}$, $\mathbf{y}_{ts}$)
  Predict $\hat{\mathbf{y}}_1, \ldots, \hat{\mathbf{y}}_b$ using classifiers in $\mathcal{C}_b^*$ on $\mathbf{X}_{ts}$.
  Compute majority voting $\tilde{\mathbf{y}}$ on predictions in $\hat{\mathbf{y}}_1, \ldots, \hat{\mathbf{y}}_b$.
  Evaluate $\tilde{\mathbf{y}}$ and store the performance in α.
  return α.

4.2.4 Experimental Setup

Experiment 1: Dynamic Feature Learning

In this experiment, we are interested in answering three main questions: 1) what are the advantages of learning the temporal representation of IMU signals for reliable SMM detection? 2) how long is the most informative time interval in IMU signals for detecting abnormal movements? 3) what is the optimal configuration for the LSTM unit? To answer these questions, we applied the proposed LSTM architecture in Section 4.2.1 on the three benchmark datasets with different values for τ and q, i.e., time steps and neuron number, respectively. We set $\tau = \{1, 3, 5, 10, 15, 25, 50\}$ and $q = \{5, 10, 20, 30, 40, 50\}$. The LSTM unit is trained on the extracted features by the CNN using the RMSProp [141] optimizer. The learned dynamic features via LSTM ($\mathbf{h}_t$) are classified to target classes using a softmax classifier. It is worthwhile to emphasize that, in this setting, since the order of samples in the training set matters, balancing the training set is impossible, thus we use the original unbalanced data.

Experiment 2: Ensemble of LSTMs

To explore the possible advantage of combining multiple classifiers, we used the procedure explained in Section 4.2.2 in order to combine a set of b best base learners. In this experiment, we used the LSTM models in Experiment 1 as base learners for the classification of unbalanced data. We set $l = 10$ and used the F1-score for the performance metric α in Algorithm 1. The experiment is repeated 10 times to evaluate the standard deviation over the mean prediction performance.

Evaluation

In all experiments, the leave-one-subject-out scheme was used for the model evaluation. For training the ensemble model, we used the same configurations for the LSTM models by fixing $\tau = 25$ as the best length for the time intervals and $q = 40$. Due to highly unbalanced classes, the evaluation is performed by computing F1-scores. We compare our results with the CNN results presented in Section 3.3 as the-state-of the-art solutions for SMM detection. Keras library [124] was used in our implementations, and all experiments were run on a GPU device (4x Nvidia Tesla K80).

4.3 Results

4.3.1 Dynamic Feature Learning Outperforms Static Feature Learning

Figure 4.2 compares the averaged SMM classification performance over subjects in the static feature space via the CNN (the green dashed line for plain feature learning) with the dynamic feature space via the LSTM, the latter with different values for τ (x-axis) and q (line colors). Here in all settings, an unbalanced training set is used in the training phase. The

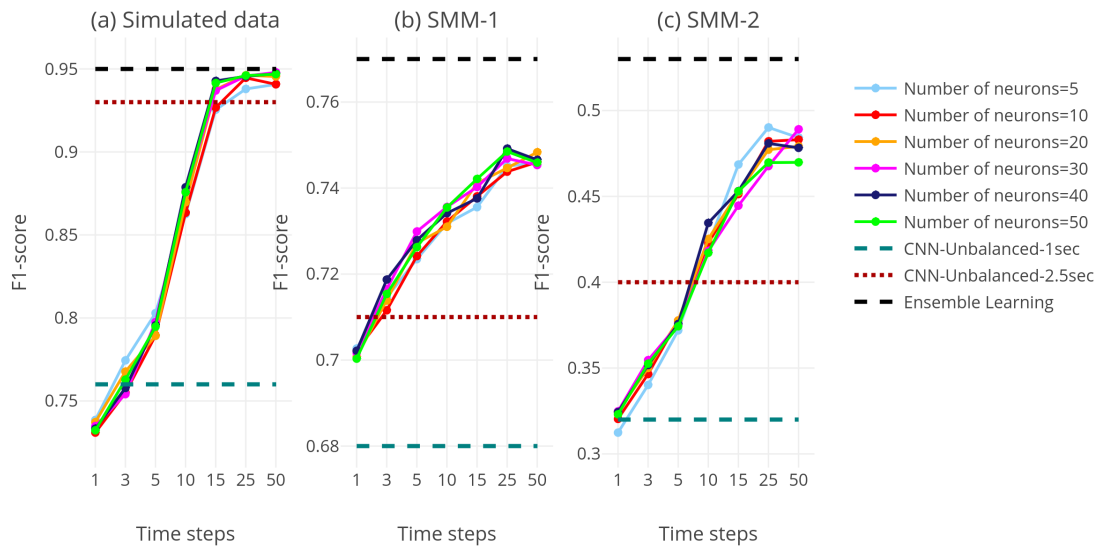

Figure 4.2: Comparison between the classification performances of CNN and LSTM for different time-steps (τ) and the number of neurons q, on three datasets and when an unbalanced training set is used for training the networks. The results show the superiority of the dynamic feature space over the static feature space. While the number of neurons in the LSTM unit has little effect on the performance, using around 2.5 seconds long interval is the best choice for extracting effective dynamic features from the IMU signals.

results on three datasets illustrate that learning the temporal representation of signals with an LSTM unit, consistently across datasets, improves the classification performance compared to the static feature learning via the CNN. The classification performance improves by increasing τ, and it reaches its highest performance around $\tau = 25$. Considering the consistency of the best τ value for different subjects and different datasets, it can be concluded that using around 25 time-steps, i.e., around 2.5 seconds long interval, for extracting dynamic features is the best choice for SMM detection purposes. On the other hand, the results show the negligible effect of the number of LSTM neurons (q) on the detection performance, thus, a value around 10 can be considered a reasonable choice.

To further benchmark the advantage of dynamic feature learning via LSTM, we used the CNN architecture for the SMM detection on the best length for the time intervals, i.e., on 2.5 seconds time intervals. The results on the three datasets are summarized in Table 4.1 and Figure 4.2 (the dotted red line). The results confirm the superiority of dynamic feature learning compared to static feature learning despite using longer time intervals for learning static features.

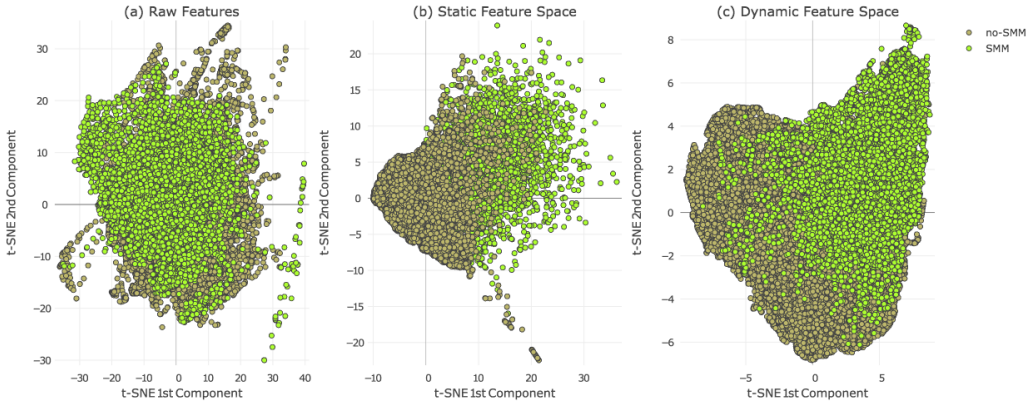

Figure 4.3: The distribution of SMM and no-SMM samples in the 2-dimensional t-SNE space for a) a raw feature space ($\mathbf{x}_t$), b) static feature space computed by CNN ($\mathbf{x}'_t$), and c) dynamic feature space computed by LSTM (h_t).

To further investigate the effect of learning the temporal representation on the separability of SMM and no-SMM samples, we employ t-distributed Stochastic Neighbor Embedding (t-SNE) [125] technique to visualize the different feature spaces in the 2-dimensional space. Then, we use the average of Fisher's separability score [126] across two t-SNE dimensions to assess the separability of classes. The effect of learning the temporal representation on the separability of SMM and no-SMM samples is shown in Figure 4.3(b). The higher Fisher's separability score (0.17) in the dynamic feature space compared to static feature spaces (0.1) can be considered as the basis for the higher classification performance of the proposed architecture, demonstrating the importance of learning dynamic features using

an LSTM based architecture.

Table 4.1 also compares the classification performance of different methods for balanced and unbalanced training sets. To facilitate comparison, here, we only present the LSTM results for $\tau = 25$ and $q = 40$. The comparison between mean classification performances over subjects shows when the static feature representation is employed (the first and the second column of Table 4.1), training the classifier on the unbalanced training set has a negative effect on the classification performance. On the other hand, employing the dynamic feature representation, computed by LSTM, improves the performances in unbalanced datasets. This observation is consistent across three datasets.

Figure 4.4 further explores the superiority of the dynamic feature representation when the training set is unbalanced. In the static feature space case, balancing the training set and enforcing the wrong prior class distribution into the classification task, despite the higher recall rate, affects negatively the precision of the classifier. In other words, the classifiers have a higher false alarm rate, which could be problematic in real-world applications. This deficit is recovered in the case of dynamic feature representation where the classifier presents higher precision rate and comparable recall with respect to static features. In fact, the LSTM-based architecture by enforcing the true prior distribution of data into the training process and, at the same time using all the recorded samples, provides an SMM detection system with higher sensitivity and specificity.

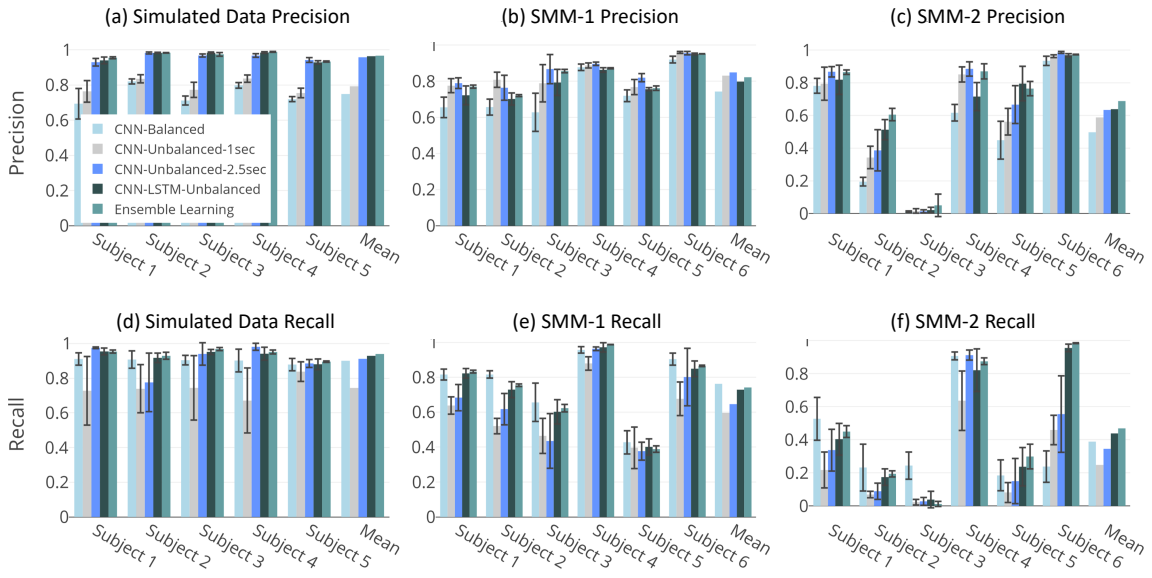

Figure 4.4: A comparison between precision and recall rates of classifiers trained on balanced/unbalanced training sets for static/dynamic feature representations. Using dynamic feature space provides an SMM detection system with higher sensitivity and specificity.

4.3.2 Ensemble of LSTMs Stabilizes the Performance

The last column of Table 4.1 summarizes the results of the ensemble approach. The results show slight boost in the mean performance compared to single LSTM classifiers, especially on SMM-2 (see dashed black line in Figure 4.2). Figure 4.4 shows that both precision and recall contribute equally to this improvement in F1-scores. In addition to the higher performance, the main advantage of EL is demonstrated by the low variability of results. This reduction in the variability is well-reflected in the reduced standard deviation around the mean performance in real datasets (2% and 4% reduction in SMM-1 and SMM-2 datasets, respectively). In other words, an ensemble of LSTMs provides more reliable SMM detector in comparison to every single LSTM classifier.

Table 4.1: F1-score results for SMM detection using static and dynamic feature spaces, and ensemble learning in three benchmarked datasets. The results show training the CNN classifier on unbalanced training sets causes the performance drop in the feature learning scenario. Using the LSTM network to extract dynamic features from the signal alleviates this problem to some degrees. The ensemble of LSTMs shows more stable performance compared to single LSTM classifiers.

Data	Sub	Static-Features Balanced [37]	Static-Features Unbalanced (1 sec) [37]	Static-Features Unbalanced (2.5 sec)	Dynamic-Features Unbalanced	Ensemble Learning
Simulated	1	0.78 ± 0.05	0.73 ± 0.13	0.95 ± 0.01	0.95 ± 0.01	0.95 ± 0.00
	2	0.86 ± 0.03	0.78 ± 0.09	0.86 ± 0.13	0.95 ± 0.01	0.96 ± 0.01
	3	0.80 ± 0.01	0.75 ± 0.13	0.95 ± 0.03	0.97 ± 0.01	0.97 ± 0.00
	4	0.85 ± 0.03	0.73 ± 0.12	0.97 ± 0.01	0.96 ± 0.02	0.97 ± 0.01
	5	0.79 ± 0.01	0.80 ± 0.04	0.91 ± 0.01	0.90 ± 0.02	0.91 ± 0.00
	Mean	0.82 ± 0.03	0.76 ± 0.11	0.93 ± 0.06	0.95 ± 0.01	$\mathbf{0.95 \pm 0.01}$
SMM-1 [19]	1	0.74 ± 0.02	0.70 ± 0.02	0.73 ± 0.04	0.77 ± 0.03	0.80 ± 0.00
	2	0.75 ± 0.02	0.63 ± 0.03	0.68 ± 0.04	0.71 ± 0.03	0.74 ± 0.00
	3	0.68 ± 0.04	0.57 ± 0.08	0.56 ± 0.13	0.68 ± 0.05	0.72 ± 0.01
	4	0.92 ± 0.01	0.88 ± 0.01	0.93 ± 0.00	0.91 ± 0.01	0.93 ± 0.00
	5	0.51 ± 0.04	0.51 ± 0.08	0.51 ± 0.04	0.52 ± 0.04	0.51 ± 0.01
	6	0.90 ± 0.01	0.79 ± 0.07	0.86 ± 0.12	0.90 ± 0.02	0.91 ± 0.00
	Mean	0.74 ± 0.03	0.68 ± 0.06	0.71 ± 0.08	0.75 ± 0.03	$\mathbf{0.77 \pm 0.01}$
SMM-2 [19]	1	0.61 ± 0.11	0.33 ± 0.14	0.47 ± 0.15	0.53 ± 0.09	0.59 ± 0.03
	2	0.20 ± 0.04	0.11 ± 0.03	0.13 ± 0.05	0.26 ± 0.06	0.29 ± 0.02
	3	0.02 ± 0.01	0.02 ± 0.01	0.02 ± 0.01	0.02 ± 0.02	0.02 ± 0.02
	4	0.72 ± 0.03	0.71 ± 0.14	0.90 ± 0.02	0.76 ± 0.09	0.87 ± 0.02
	5	0.21 ± 0.09	0.14 ± 0.09	0.23 ± 0.17	0.35 ± 0.15	0.43 ± 0.08
	6	0.36 ± 0.13	0.62 ± 0.08	0.68 ± 0.21	0.96 ± 0.01	0.98 ± 0.00
	Mean	0.35 ± 0.08	0.32 ± 0.1	0.40 ± 0.13	0.48 ± 0.08	$\mathbf{0.53 \pm 0.04}$

4.4 Conclusions

In this study, we presented an application of deep learning in SMM detection context by proposing an LSTM-based architecture for extracting dynamic features on the sequence of IMU data. Comparing with the state-of-the-art solution for SMM detection, and on both simulated and real datasets, we illustrated the advantage of employing the temporal information in improving the separability of SMM and no-SMM samples. We experimentally showed that using around 2.5 seconds long interval for extracting dynamic features is the best choice for SMM detection purposes.

Further, we showed using around 10 neurons in the LSTM unit is a reasonable choice in order to extract the dynamics of samples over time. As a side-advantage of transferring data to dynamic feature space, we experimentally demonstrated the higher performance of our method when, in a real-world setting, the distribution of samples in SMM and no-SMM classes is highly skewed. We showed, while the skewness of samples negatively affects the performance of the SMM detector in the static feature space, exploiting the temporal patterns of multi-sensor IMU signals recovers its performance. This advantage facilitates training high-performance models by exploiting whole data sequences in real-time SMM detection scenarios. We further illustrated the advantage of ensemble learning to provide more stable and reliable SMM detectors. Our effort, for the first time in the SMM detection context, demonstrated the superiority of recurrent structures in extracting discriminative temporal patterns from IMU signals.

Chapter 5

Novelty Detection using Deep Normative Modeling for IMU-Based Abnormal Movement Monitoring in Parkinson's Disease and Autism Spectrum Disorders

5.1 Introduction

In Chapter 3 and 4, as many other research studies on stereotypical motor movement (SMM) detection using wearable sensors [39, 19, 36, 142, 123], we applied supervised machine learning algorithms to classify samples of abnormal movements from normal ones. There are three main challenges in applying supervised approaches for abnormal movement detection: (i)

This chapter is based on the following paper:

1. Nastaran Mohammadian Rad, et al., "Novelty Detection using Deep Normative Modeling for IMU-Based Abnormal Movement Monitoring in Parkinsons Disease and Autism Spectrum Disorders." Accepted in Sensors, 2018.

they generally rely on the availability of labeled data while, especially in this context, data labeling is an expensive, time-consuming and subjective task [143, 144, 145], as it needs full monitoring of subjects during the data collection phase; (ii) severe class distribution skewness, where samples in the normal class severely out-represent abnormal samples in recorded data from patients with ASD and PD [146, 147]; this fact makes the classification techniques sub-optimal for these applications; (iii) the heterogeneity of non-stationary patterns in normal and abnormal movements that makes the task of finding a separating hyper-cube in classification scenarios even more cumbersome [148].

As an alternative for supervised approaches, novelty detection provides all the ingredients needed for tackling the aforementioned challenges in an unsupervised fashion. In general, novelty detection is defined as the task of learning the overall characteristics of available normal samples in the training phase and then using these characteristics to recognize novel samples that differ in some respects from the normal samples at test time [144, 149]. Based on this definition, novelty detection approaches naturally need only samples of the normal class in the training phase; hence, they do not need labeled data and are immunized against highly imbalanced class distributions. More importantly, adopting a probabilistic policy in novelty detection enables us to estimate the generative probability density function of the normal data, which can cover a wide and heterogeneous spectrum of normal samples. These advantages made novelty detection techniques very successful in many applications ranging from fraud detection [150, 151], medical diagnosis [152, 153, 154], and fault detection [155, 156], to anomaly and outlier detection in sensor networks [157, 158], video surveillance [159, 160], and text mining [161, 162].

5.1.1 Contribution

In this chapter, we adopt a probabilistic novelty detection approach based on normative modeling [163] in order to, first, model heterogeneous normal movements in Parkinson's disease (PD) and ASD and, second, to use the resulting model in a novelty detection paradigm to detect freezing of gaits (FOGs) and stereotypical motor movements (SMMs) in respectively PD and ASD patients. To this end, by assuming a multivariate normal distribution on the collected accelerometer signals of normal movements and exploiting the underlying principles of probabilistic deep neural networks [164], we extend the applications of normative modeling to unimodal datasets. In general, a normative model is constructed in the training phase by estimating a mapping function between two different data modalities, e.g., behavioral covariates and biological measurements [163]. In this study, we use the denoising autoencoder (DAE) to reconstruct the original IMU signals of normal movements from their noisy versions. In fact, the model implicitly learns the distribution of the normal movements. Using dropout layers in the DAE architecture enables us to estimate also the variance of predictions (which is necessary for normative modeling) in addition to mean predictions. We compare the proposed method with state-of-the-art supervised approaches, as well as classic one-class classification and reconstruction-based novelty detection. Our experimental results on three benchmark datasets illustrate that the proposed method provides a reasonably close performance to its supervised counterparts, whilst yielding the best performance among other competing novelty detection approaches on three benchmark datasets.

In the remaining text, we first review the state-of-the-art of novelty detection techniques for abnormal movement detection. Then, we present our proposed unsupervised novelty detection approach based on normative

modeling. We then compare our experimental results with other novelty detection and supervised methods. Finally, we discuss the advantages and limitations of the proposed method and state possible future directions.

5.2 Related Works

Recent studies on automatic SMM and FOG detection using wearable sensors have mainly focused on applying supervised machine learning and deep learning approaches, such as convolutional neural network (CNN) and long short-term memory (LSTM), to distinguish between the normal and abnormal movements [165, 166, 167, 91, 38, 82, 37]. These methods are based on extracting or learning a set of robust features from the original signals and then applying the supervised algorithms for abnormal movement detection. The main drawback of these approaches, however, is their need for labeled data. To overcome this problem, few studies have recently focused on using novelty detection methods [168, 169]. In a FOG detection application, Cola et al. [168] used a distance-based novelty detection method on accelerometer signals to detect abnormal gait patterns. Their proposed method consists of extracting a set of hand-crafted features and then applying a K-Nearest Neighbor (KNN) method. The KNN approach assumes that normal gait samples are located at the close distance from each other. Thus, a sample is determined as an abnormal sample if it is located far from its neighbors. Their proposed method achieved on average 80% accuracy for detecting abnormal gait samples. Despite the reported high accuracy rate, the high computational complexity of KNN at the test time severely limits its application in real-time applications. Elsewhere, Nguyen et al. [169] proposed a probabilistic novelty detection method for abnormal gait recognition in musculoskeletal disorders using Microsoft Kinect® sensors. Their method was based on training a Hidden Markov Model (HMM) to

model the transition of human posture states in a gait cycle. Then, to distinguish between the normal gait samples from the abnormal ones, a threshold was defined based on the mean and standard deviation of the estimated log-likelihood on normal gait samples.

Recently, deep learning approaches were also used for novelty detection applications. Erfani et al. [170] proposed a hybrid model of an autoencoder and one-class SVM for detecting anomalies in high-dimensional and large-scale datasets including a daily activity dataset. A set of learned features by autoencoders was fed to a one-class SVM in order to detect the abnormal samples. Their experimental results showed the superiority of using one-class SVM in the learned latent space rather than the original raw signal space. Autoencoders are also widely used for detecting abnormal patterns in medical images through the reconstruction error between the output of the model and the actual input [171, 172]. Novelty detection based on reconstruction error was also used by Khan and Taati [173] for fall detection using wearable sensors. The proposed approach was based on using a channel-wise ensemble of autoencoders for data reconstruction and setting a threshold on the reconstruction error to distinguish the falling instances.

5.3 Methods

In the context of abnormal movement detection using wearable sensors, novelty detection is defined as detecting atypical movements in the test phase while only normal movements are available in the training phase. In this study, we consider a probabilistic novelty detection approach consisting of the following three steps: (1) learning the distribution of normal movements using a probabilistic denoising autoencoder; (2) quantifying the deviation of each test sample from the distribution of normal movements,

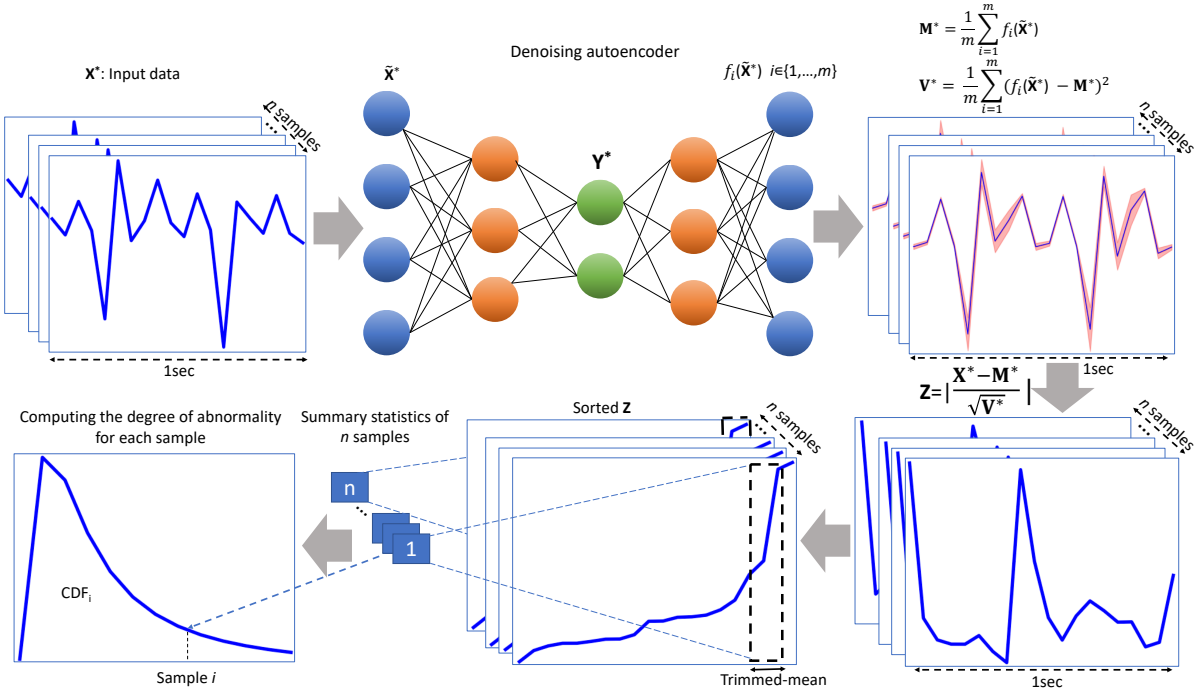

Figure 5.1: The proposed method for the abnormal movement detection in the test time.

the so-called normative probability map (NPM), in the normative modeling framework; (3) computing the degree of novelty of each test sample by fitting a generalized extreme value distribution on summary statistics of its NPM.

We formalize these three steps in the next 3 subsections. Figure 5.1 also shows the proposed method.

5.3.1 Learning the Distribution of Normal Movements via the Denoising Autoencoder

As stated in the previous section, our method starts by modeling the normal movements. To do this, we use convolutional neural networks, which are the state-of-the-art for activity recognition and movement monitoring using wearable sensors. In particular, we train a denoising autoencoder (DAE), which is a type of (autoencoding) neural network that aims to

reconstruct (denoise) its inputs from noisy samples.

More formally, given a training set $\mathbf{X}_N \in \mathbb{R}^{n_N \times p}$ consisting of n_N samples of normal movements drawn from a distribution P_N of normal movements, a trained DAE is a function f_N that has the property that $f_N(\mathbf{X} + \epsilon) \approx \mathbf{X}$ for $\mathbf{X} \in \mathbf{X}_N$. Given sufficient training data, the network generalizes to reconstruct any $\mathbf{X} \in P_N$.

How well the autoencoder is able to denoise its input is proportional to how well that input matches the distribution of the training data, in our case how well the input matches a normal movement. Hence, we can use the distance between the reconstruction of DAE and the true sample, the reconstruction error, as a measure of the likelihood $P_N(\mathbf{X})$ of the sample.

However, the neural network only produces a point estimate, that is a single possible reconstruction given a noisy input. For some features or samples, this prediction might be very accurate, while others can be much harder to reconstruct. The reconstruction error does not take this prediction uncertainty into account.

To use the prediction uncertainty properly, we use the NPM, introduced in [163]. The original NPM method used Gaussian processes to model the normal data, which also provide a variance as a measure of uncertainty. To calculate the variance of the predictions in our denoising autoencoder setup, we instead use dropout [164], to make the network nondeterministic. As shown by Gal and Ghahramani [164], using Monte Carlo sampling by applying dropout at test time provides an approximation of the posterior $P_N(\theta \mid \mathbf{X})$. After drawing m samples from the predictive distribution, we

can calculate their empirical mean and variance,

$$\mathbf{M}^* = \frac{1}{m}\sum_{i=1}^{m} f_i(\tilde{\mathbf{X}}^*) \tag{5.1a}$$

$$\mathbf{V}^* = \frac{1}{m}\sum_{i=1}^{m} (f_i(\tilde{\mathbf{X}}^*) - \mathbf{M}^*)^2. \tag{5.1b}$$

Here, f_i indicates the different variations of the autoencoder network, which are formed by applying dropout before every weight layer.

5.3.2 Quantifying the Deviation from P_N

In this study, we adapt the normative modeling framework in order to quantify the deviation of each newly-seen test sample from the distribution of normal movements P_N. In this framework, the mean and variance of the reconstruction are used to compute an NPM as follow:

$$\mathbf{Z} = \frac{\mathbf{X}^* - \mathbf{M}^*}{\sqrt{\mathbf{V}^*}}. \tag{5.2}$$

$\mathbf{Z}$ or NPM scores are in fact z-scores, quantifying the deviation of test samples in $\mathbf{X}^*$ from a reconstructed normal sample under P_N, in units of standard deviation of the predictive distribution [174]. It combines two sources of information: (1) the prediction error (difference between the true and expected predicted responses) and (2) the predictive variance of the test points.

5.3.3 Computing the Degree of Novelty

The NPM score of each test sample is a p-dimensional multivariate measure of deviations. In order to summarize these deviations into a degree of abnormality, we follow [163] and employ the generalized extreme value distribution (GEVD) [175, 176] to model the samples in the extreme tails

of P_N (see Appendix A.1 for more details). In fact, we consider that abnormal motor movements may occur as an extreme deviation from a normal pattern. As in [163], we adopt a "block maxima" approach where we compute the 90% trimmed mean of the top 1% values in Z of each sample in order to summarize the deviations as a single number. Then, to make probabilistic subject-level inferences about these deviations, we fit a GEVD on the resulting summary statistics. The cumulative density function of the resulting GEVD at a given test sample then can be used as the probability of each sample being an abnormal sample [177].

5.3.4 Experimental Materials

We compare the performance of the proposed probabilistic novelty detection approach with reconstruction-based novelty detection [149, 170], one-class support vector machine (SVM) [178] and supervised deep learning approaches on two datasets: (i) an SMM dataset collected in a longitudinal study from children with ASD [19] [1], and (ii) the Daphnet Freezing of Gait dataset collected from PD patients [179]. In the following, we detail the datasets and the preprocessing steps.

Datasets

We evaluate our proposed method on the SMM datasets described in Section 3.2.4. We follow the same preprocessing steps explained in Section 3.2.4 except that in the segmentation phase, segments with normal movement samples are selected to train the model in the novelty detection framework. Other partial normal segments were removed from the training data.

[1] The SMM dataset and the full description of the data are publicly available at `https://bitbucket.org/mhealthresearchgroup/stereotypypublicdataset-sourcecodes/downloads`.

Table 5.1: The class distribution of normal and abnormal samples and the gender of patients in three datasets.

Data	Subject	#Normal	#Abnormal	All	Abnormal/All	Gender
Daphnet-FOG [179]	**Sub1**	5714	334	6048	0.06	M
	Sub2	3918	578	4496	0.13	M
	Sub3	5488	912	6400	0.14	M
	Sub4	6592	0	6592	0	M
	Sub5	5139	1517	6656	0.23	M
	Sub6	5917	419	6336	0.07	F
	Sub7	4858	262	5120	0.05	M
	Sub8	1812	620	2432	0.25	F
	Sub9	4673	863	5536	0.16	M
	Sub10	7104	0	7104	0	F
	Total	50,482	6238	56,720	**0.11**	-
SMM-1 [19]	**Sub1**	21,292	5663	26,955	0.21	M
	Sub2	12,763	4372	17,135	0.26	M
	Sub3	31,780	2855	34,635	0.08	M
	Sub4	10,571	10,243	20,814	0.49	M
	Sub5	17,782	6173	23,955	0.26	M
	Sub6	12,207	17,725	29,932	0.59	M
	Total	106,395	47,031	153,426	**0.31**	-
SMM-2 [19]	**Sub1**	18,729	11,656	30,385	0.38	M
	Sub2	22,611	4804	27,415	0.18	M
	Sub3	40,557	268	40,825	0.01	M
	Sub4	38,796	8176	46,972	0.17	M
	Sub5	22,896	6728	29,624	0.23	M
	Sub6	2375	11,178	13,553	0.82	M
	Total	145,964	42,810	188,774	**0.23**	-

In addition to SMM dataset, we also evaluate the performance of the peoposed method on the Daphnet Freezing of Gait dataset [179]. Daphnet Freezing of Gait dataset, here referred to as Daphnet-FOG, were collected from 10 PD patients at a 64 Hz frequency rate while participants wore three 3-axis accelerometer sensors on their shank, thigh and belt. During the experiment, participants were instructed to perform walking tasks. The whole experiment was recorded with a digital video camera. Then, two physiotherapists annotated the FOG episodes using the video record-

ings. Following the preprocessing stage in [90], we first downsampled the accelerometer data to 32 Hz. The data were then segmented into 1 second-long intervals using a sliding window. The sliding window was moved along the time dimension with 10 time-steps to make overlaps between consecutive windows. Similar to SMM dataset, in the segmentation phase, only segments with normal movement samples were selected to train the model.

Table 5.1 summarizes the number of normal and abnormal samples for each subject in the SMM and Daphnet-FOG datasets. The difference in the number of samples in the abnormal and normal classes represents the unbalanced nature of data where in the SMM-1 and SMM-2 datasets, 31% and 23% of samples are in the SMM class, and in the Daphnet-FOG dataset, 11% of samples are in the FOG class.

Network Architectures

Considering their different rhythmic characteristics, we used different network architectures for the Daphnet-FOG and SMMs datasets. We adopted the CNN architecture that was proposed by Hammerla et al. [90] for the Daphnet-FOG dataset and the CNN architecture proposed by Rad et al. [39] for the SMM datasets. In the following, we detail how these architectures are manipulated to serve our purpose explained in Section 5.3 [2].

1. DAE architecture for the Daphnet-FOG dataset: The original CNN architecture in Hammerla et al. [90] was used for encoding the signal into a lower dimensional representation. This architecture contains four convolutional layers alternating convolution, batch normalization, rectified linear units (ReLU) and max-pooling layers to map the large input space to a lower dimensional feature space. A fully-connected layer is then stacked on top of the fourth convolution layer to form the encoder. We concatenate a mirror reversal of the CNN network to the

[2]The Keras library [124] is used to implement DAE and CNN architectures.

last encoder layer in order to reconstruct the input signal in a DAE architecture. In the decoding part, we replace max-pooling layers with up-sampling layers. In order to capture the model uncertainty, we placed a dropout layer before every weight layer [164]. The resulting architecture is shown in Figure 5.2a.

2. DAE architecture for SMM datasets: Similar to [39], the encoder architecture consists of three convolutional layers, which alternates convolution, batch normalization, ReLUs and average-pooling layers to transform the raw feature space into a lower dimensional set of features. A fully-connected layer is then stacked on top of the third convolution layer. The resulting latent vector is then decoded in the decoder to reconstruct the input signal. Similar to the DAE architecture for the Daphnet-FOG dataset, we concatenate a mirror reversal of the CNN network to the last layer of the encoder in order to construct the decoder. To capture the prediction uncertainty, dropout layers are used before every weight layer. The architecture and the configuration of each layer are depicted in Figure 5.2b.

Experimental Setups and Evaluation

We conducted four experiments to evaluate the performance of the proposed method against three competing approaches:

- Experiment 1, normative modeling: We followed the proposed procedure explained in Section 5.3, using the DAE architectures described in Section 5.3.4 for learning the distribution of the normal movements on the SMM-1, SMM-2 and Daphnet-FOG datasets. In this setting, models are trained in an unsupervised manner and only on the samples of normal movements. For training the DAEs, we used

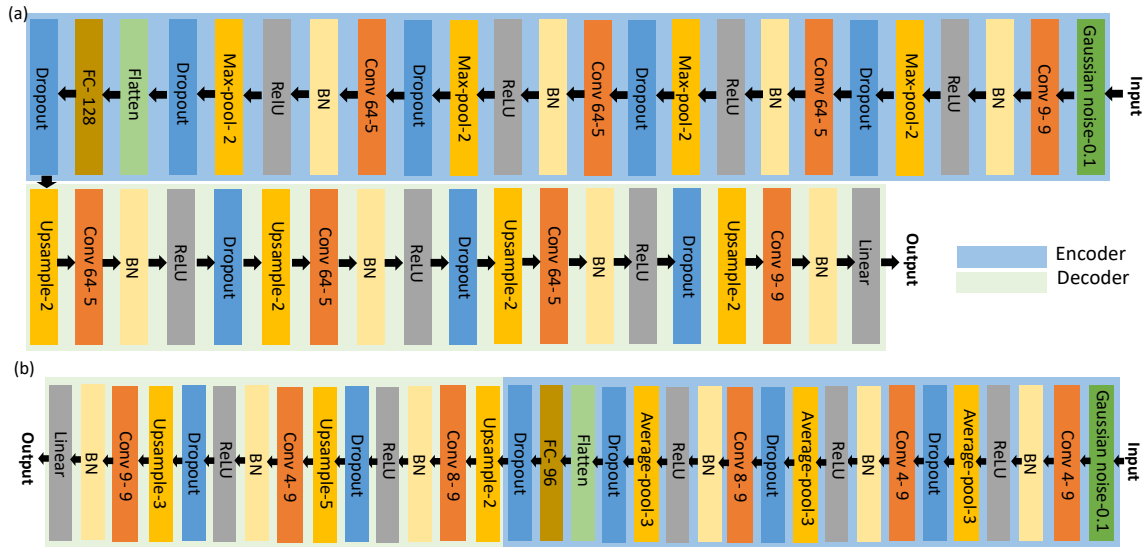

Figure 5.2: The architecture of convolutional denoising autoencoder for (**a**) the Daphnet-FOG dataset and (**b**) the SMM dataset. Each colored box represents one layer. The type and configuration of each layer are shown inside each box. For example, Conv 64-5 denotes a convolutional layer with 64 filters and the kernel size of 5.

the RMSprop optimizer to minimize the mean squared error loss function. To compute $\mathbf{M}^*$ and $\mathbf{V}^*$, we drew $m = 50$ MC samples (using dropout at test time, it is possible to randomly sample from network weights m times) from DAE predictions (the reconstructed inputs), and the mean and variance across these 50 MC samples are used to compute the $\mathbf{M}^*$ and $\mathbf{V}^*$ matrices. In all experiments, we fix the dropout level to 0.1. Later in order to investigate the effect of the dropout level on the performance of the proposed novelty detection approach, we repeat this experiment for different dropout probability levels $\delta = \{0.1, 0.2, 0.3, 0.4, 0.5\}$ and compare the results.

- Experiment 2, reconstruction-based: The goal of this experiment is to assess the effect of incorporating prediction uncertainties, i.e., $\mathbf{V}^*$, on the performance of the novelty detection system. All the experimen-

tal settings in this experiment are similar to Experiment 1, except for computing the NPMs, where we use $\mathbf{Z} = \mathbf{X}^* - \mathbf{M}^*$ instead of Equation 5.2. Since in this setting, only the reconstruction error is used to construct a model of normal movements, we refer to this experiment as "reconstruction-based".

- Experiment 3, one-class SVM: The goal in this experiment is to compare the proposed method for novelty detection with one-class classification. To this end, we train a one-class SVM model in a novelty detection setting [149, 180, 170, 181]. One-class SVM fits a hyper-sphere decision boundary on a nonlinearly-transformed feature space to include the majority of samples in the normal class and detects anomalies as deviations from the learned decision boundary. In this experiment in a similar setting used by Erfani et al. [170], we use the learned reduced-rank latent space via the DAE model, i.e., $\mathbf{Y}_N \in \mathbb{R}^{n_N \times q}$, to train a one-class SVM model. We use this model later to distinguish the normal and abnormal movements on the samples. For the one-class SVM, we employed the implementation available in the scikit-learn [182] package. We used the radial basis function (RBF) kernel with default hyperparameters [3], where $\nu = 0.5$ and $\gamma = \frac{1}{q}$.

- Experiment 4, supervised: To compare the performance of the proposed unsupervised novelty detection technique with supervised classification, we used the CNN architecture proposed in Hammerla et al. [90] and Rad et al. [39] on the Daphnet-FOG and SMM datasets, respectively, in a fully-supervised scenario.

Note that since the samples for Subjects 4 and 10 in the Daphnet-FOG dataset only contain normal movements (see Table 5.1), it is not

[3]Considering our assumption that only normal movement samples are available during the training phase, fine-tuning these hyperparameters is not possible. See Section 5.5.2 for the discussion.

possible to evaluate the benchmark approaches on these two subjects in Experiments 1–3. Thus, in an extra setting, we repeat Experiments 1–3 when only these two subjects are used in the training phase. This setting is even more close to the reality as only subjects with normal movements are available during the training phase (in this case, there is no need for the additional preprocessing procedure to select the normal segments).

In all experiments, the leave-one-subject-out cross-validation is used for the model evaluation, and the area under the receiver operating characteristic curve (ROC), i.e., AUC, is computed as the performance measure. The whole experimental procedures are repeated 5 times to report the standard deviation over the mean AUC performances.

5.4 Results

Table 5.2 summarizes single-subject and average AUC measures for the four experiments that were described in Section 5.3.4 on the Daphnet-FOG, SMM-1, and SMM-2 datasets. On the Daphnet-FOG dataset, we observed a large variance of results across subjects. In particular, the normative modeling and reconstruction-based methods achieved a much lower AUC performance on Subjects 6 and 8 than on the other subjects. These two subjects were the only females out of the 3 ones in the dataset exhibiting atypical movement behavior (see Table 5.1). A potential explanation for the lower performance is that, when training on mainly male subjects, novelty detection models, which use the reconstruction error, are unable to reconstruct normal female movement behavior correctly. On the SMM datasets, the performance was more similar across subjects, notably on the SMM-1 dataset. This could be due to the controlled setting used to collect data: while wearing the sensors, participants were observed in the lab, sitting in a comfortable chair with a familiar teacher [123]. Results

on the Daphnet-FOG dataset also indicated the presence of possible biases due to the limited size of the data from normal subjects (see also the results reported in Section 5.4.4). The public availability of larger datasets would allow a more thorough assessment of the methods for abnormal movement detection in PD and ASD, which would be highly beneficial to advance patient care and research. The results are further investigated in the following sections.

Table 5.2: The average of AUC results for novelty detection using normative modeling, reconstruction-based and one-class SVM on three benchmark datasets.

Dataset	Subject	Normative	Reconstruction	1C-SVM	Supervised
Daphnet-FOG	**Sub1**	0.87 ± 0.01	0.64 ± 0.00	0.73 ± 0.02	0.80 ± 0.02
	Sub2	0.80 ± 0.03	0.80 ± 0.00	0.54 ± 0.03	0.95 ± 0.01
	Sub3	0.87 ± 0.01	0.77 ± 0.00	0.43 ± 0.06	0.90 ± 0.02
	Sub5	0.83 ± 0.01	0.70 ± 0.00	0.60 ± 0.02	0.80 ± 0.03
	Sub6	0.60 ± 0.03	0.50 ± 0.00	0.70 ± 0.05	0.80 ± 0.04
	Sub7	0.79 ± 0.01	0.66 ± 0.00	0.67 ± 0.02	0.92 ± 0.01
	Sub8	0.64 ± 0.01	0.48 ± 0.00	0.70 ± 0.02	0.65 ± 0.03
	Sub9	0.77 ± 0.02	0.51 ± 0.00	0.62 ± 0.05	0.94 ± 0.01
	Mean	$\mathbf{0.77 \pm 0.01}$	0.63 ± 0.00	0.62 ± 0.01	0.84 ± 0.01
SMM-1	**Sub1**	0.86 ± 0.01	0.71 ± 0.00	0.32 ± 0.01	0.89 ± 0.01
	Sub2	0.85 ± 0.01	0.88 ± 0.00	0.23 ± 0.03	0.83 ± 0.02
	Sub3	0.87 ± 0.02	0.91 ± 0.00	0.22 ± 0.03	0.93 ± 0.01
	Sub4	0.88 ± 0.03	0.81 ± 0.01	0.31 ± 0.04	0.95 ± 0.00
	Sub5	0.76 ± 0.03	0.87 ± 0.01	0.26 ± 0.01	0.83 ± 0.03
	Sub6	0.88 ± 0.03	0.82 ± 0.00	0.30 ± 0.02	0.93 ± 0.01
	Mean	$\mathbf{0.85 \pm 0.01}$	0.83 ± 0.00	0.28 ± 0.01	0.89 ± 0.01
SMM-2	**Sub1**	0.69 ± 0.05	0.76 ± 0.01	0.37 ± 0.04	0.79 ± 0.07
	Sub2	0.61 ± 0.04	0.58 ± 0.00	0.46 ± 0.02	0.53 ± 0.03
	Sub3	0.62 ± 0.02	0.63 ± 0.02	0.42 ± 0.02	0.63 ± 0.05
	Sub4	0.74 ± 0.09	0.42 ± 0.02	0.49 ± 0.05	0.88 ± 0.06
	Sub5	0.65 ± 0.09	0.77 ± 0.03	0.45 ± 0.03	0.78 ± 0.04
	Sub6	0.65 ± 0.03	0.73 ± 0.02	0.44 ± 0.02	0.74 ± 0.11
	Mean	$\mathbf{0.66 \pm 0.02}$	0.65 ± 0.01	0.44 ± 0.01	0.73 ± 0.03

5.4.1 Normative Modeling Outperforms Reconstruction-Based and One-Class SVM in Novelty Detection

The comparison between results achieved by our normative modeling method and its reconstruction-based variant indicates the beneficial effect of incorporating the uncertainty of the predictions in the NPM scores for the Daphnet-FOG dataset. In this context, for all subjects, the normative modeling method outperformed the reconstruction-based one. On this dataset, normative modeling also outperformed one-class SVM on all except one subject. These results illustrate the effectiveness of normative modeling method for detecting movement disorder behavior in PD patients.

On the SMM-1 and SMM-2 datasets, normative modeling and reconstruction-based modeling methods achieved similar performance. This indicates that the uncertainty of the prediction did not significantly affect the ranking of the samples obtained using the reconstruction-error scores. On this dataset, the performance of one-class SVM was not very satisfactory. This result can be explained by the fact that one-class SVM does not rely on the properties of the distribution of the training data; rather, it fits a decision boundary on a nonlinearly-transformed feature space to include the majority of samples in the normal class and detects anomalies as samples fall outside the learned decision boundary. Therefore, the performance of this method is highly dependent on selecting proper parameters to control the size of the boundary.

5.4.2 Novelty Detection Methods vs. Supervised Learning Methods

Our experimental results in Table 5.2 demonstrate that our normative modeling method provided a reasonably close performance to its supervised

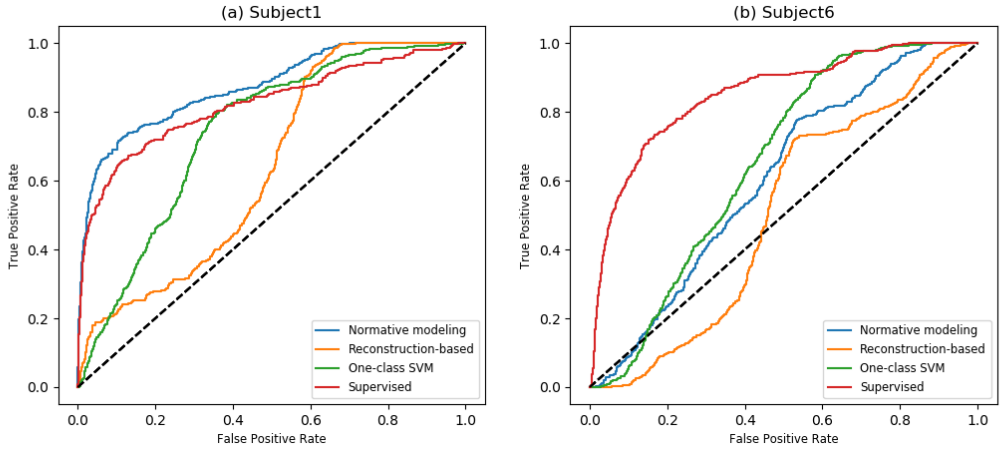

Figure 5.3: ROC curves corresponding to the reported AUCs for Subjects 1 and 6 of the Daphnet-FOG dataset in Table 5.2.

counterpart on the SMM-1 dataset and a relatively close performance to the supervised method for the SMM-2 and Daphnet-FOG datasets. In particular, on the Daphnet-FOG dataset, in two cases (Subjects 1 and 5), the normative modeling method outperformed the supervised method (with a 7% and 3% improvement, respectively). Furthermore, on the SMM-1 dataset, the reconstruction-based method outperformed the supervised method in two cases, Subjects 2 and 5, with a 5% and 4% improvement, respectively.

To get a summarized demonstration of the performance of different novelty detection methods, we consider the best and the worst normative modeling results on the Daphnet-FOG dataset, i.e., Subjects 1 and 6. ROC curves for these subjects are depicted in Figure 5.3. In Figure 5.3a, we can see that both the reconstruction-based and normative modeling methods were able to identify the most normal (negative) data for Subject 1 correctly. However, the reconstruction-based approach was not able to find the most likely abnormal movement (positive) samples. Figure 5.3b shows the results for Subject 6. Here, around 1/4 of the samples were clearly identified as normal by most methods; however, the other samples could

not be distinguished. In the normative modeling method, both positive and negative samples were assigned a high likelihood of being abnormal, perhaps because the normal movements for this subject differed too much from those in the training data.

Table 5.3: The average AUPR for novelty detection using normative modeling, reconstruction-based and one-class-SVM on three benchmark datasets.

Dataset	Subject	Normative	Reconstruction	1C-SVM	Supervised
Daphnet-FOG	**Sub1**	0.48 ± 0.02	0.11 ± 0.00	0.11 ± 0.02	0.38 ± 0.02
	Sub2	0.32 ± 0.03	0.51 ± 0.01	0.13 ± 0.01	0.79 ± 0.03
	Sub3	0.51 ± 0.01	0.35 ± 0.00	0.12 ± 0.02	0.55 ± 0.02
	Sub5	0.54 ± 0.01	0.36 ± 0.00	0.28 ± 0.02	0.59 ± 0.03
	Sub6	0.08 ± 0.01	0.06 ± 0.00	0.10 ± 0.02	0.33 ± 0.06
	Sub7	0.22 ± 0.01	0.09 ± 0.00	0.09 ± 0.01	0.49 ± 0.04
	Sub8	0.38 ± 0.01	0.26 ± 0.00	0.43 ± 0.05	0.34 ± 0.04
	Sub9	0.42 ± 0.03	0.19 ± 0.00	0.21 ± 0.05	0.71 ± 0.05
	Mean	$\mathbf{0.37 \pm 0.01}$	0.24 ± 0.00	0.18 ± 0.01	0.52 ± 0.01
SMM-1	**Sub1**	0.69 ± 0.01	0.56 ± 0.00	0.15 ± 0.00	0.76 ± 0.02
	Sub2	0.63 ± 0.02	0.76 ± 0.00	0.17 ± 0.01	0.72 ± 0.03
	Sub3	0.57 ± 0.03	0.60 ± 0.01	0.05 ± 0.00	0.70 ± 0.04
	Sub4	0.86 ± 0.02	0.78 ± 0.01	0.39 ± 0.02	0.93 ± 0.00
	Sub5	0.50 ± 0.05	0.75 ± 0.01	0.18 ± 0.00	0.67 ± 0.04
	Sub6	0.90 ± 0.02	0.81 ± 0.01	0.47 ± 0.01	0.95 ± 0.01
	Mean	0.69 ± 0.01	$\mathbf{0.71 \pm 0.00}$	0.23 ± 0.00	0.79 ± 0.01
SMM-2	**Sub1**	0.56 ± 0.05	0.65 ± 0.01	0.33 ± 0.03	0.71 ± 0.08
	Sub2	0.23 ± 0.02	0.20 ± 0.00	0.16 ± 0.01	0.22 ± 0.01
	Sub3	0.01 ± 0.00	0.02 ± 0.00	0.01 ± 0.00	0.02 ± 0.01
	Sub4	0.37 ± 0.07	0.20 ± 0.00	0.18 ± 0.02	0.66 ± 0.16
	Sub5	0.33 ± 0.10	0.42 ± 0.04	0.22 ± 0.02	0.48 ± 0.07
	Sub6	0.87 ± 0.01	0.90 ± 0.01	0.78 ± 0.01	0.91 ± 0.04
	Mean	$\mathbf{0.40 \pm 0.02}$	0.40 ± 0.01	0.28 ± 0.01	0.50 ± 0.03

Since the datasets presented in this study are highly skewed, especially the Daphnet-FOG dataset, in addition to AUC, we also evaluated the per-

formance of the methods using the area under the precision-recall curve (AUPR) [183]. Compared to AUC, the AUPR score places more weight on the highly ranked predictions by each method. As is shown in Table 5.3, on the Daphnet-FOG dataset, the normative modeling method achieved a higher average AUPR than other novelty detection methods. For some subjects, in particular Subject 6, all of the novelty detection methods showed low performances. We believe this is because this subject was too different from the training data, and hence, none of the methods found clear FOG signals, which can also be seen in Figure 5.3b. On the SMM datasets, normative modeling and reconstruction-based methods achieved comparable performance in terms of AUPR, while both clearly outperformed one-class SVMs. The AUPR scores for the autoencoder-based methods were quite high on this dataset, which indicates that they were able to find clear instances of SMM behaviors in all subjects correctly.

5.4.3 Effect of Dropout Level

Figure 5.4 depicts the effect of different dropout probabilities on the performance of the normative modeling method on the SMM-1, SMM-2 and Daphnet-FOG datasets with the leave-one-subject-out scheme. As is shown in Figure 5.4, using the different dropout probabilities had a negligible effect on the performance of the normative modeling method for the SMM and Daphnet-FOG datasets. Thus, a value between 0.1 and 0.4 can be used as the dropout probability level without a significant drop in the performance.

5.4.4 Training Only on Normal Subjects

It is interesting to investigate how our novelty detection methods perform when only data from subjects without atypical movement behavior are

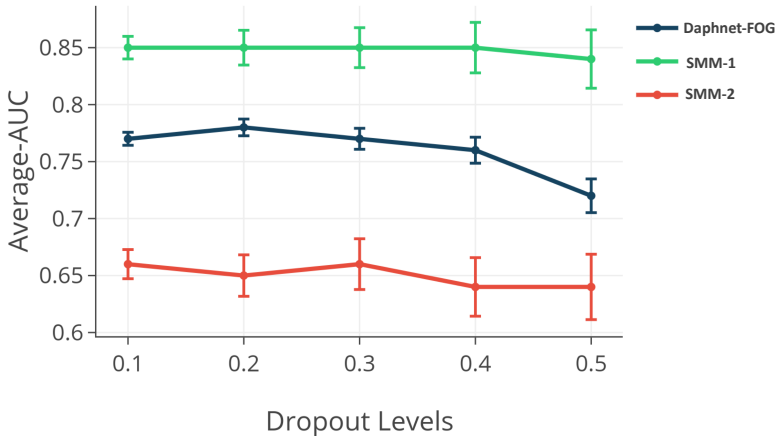

Figure 5.4: The effect of different dropout probabilities on the performance of the normative modeling method.

present in the training set. In this setting, the expert interaction and preprocessing time were reduced. Therefore, in this experiment, we trained the considered novelty detection models only on two normal subjects, i.e., Subjects 4 and 10 in the Daphnet-FOG dataset (see Table 5.1). Results of this experiment are shown in Table 5.4. As expected, there is a drop in the average performance compared to the results of Experiment 1 (see the results for Daphnet-FOG dataset in Table 5.2), which is likely due to the limited training data with just two subjects. Interestingly, in this setting, the normative modeling method improved its performance on Subject 2 (0.92 average AUC), showing that the normal movement behavior of this subject was closer to that of Subjects 4 and 10 than to that of the other subjects. Overall, the results of normative modeling and reconstruction-based methods decreased when using less data, while the results of one-class SVM did not change significantly, indicating that the latter method is incapable of exploiting information from more subjects.

Table 5.4: The average of AUC results for novelty detection using normative modeling, reconstruction-based and one-class-SVM trained only on the two available normal subjects (Subjects 4 and 10) of the Daphnet-FOG dataset.

Dataset	Subject	Normative	Reconstruction	1C-SVM
Daphnet-FOG	**Sub1**	0.81 ± 0.01	0.63 ± 0.00	0.76 ± 0.00
	Sub2	0.92 ± 0.00	0.81 ± 0.00	0.65 ± 0.02
	Sub3	0.75 ± 0.08	0.67 ± 0.01	0.41 ± 0.05
	Sub5	0.82 ± 0.01	0.69 ± 0.00	0.63 ± 0.02
	Sub6	0.51 ± 0.05	0.41 ± 0.01	0.77 ± 0.03
	Sub7	0.67 ± 0.03	0.53 ± 0.00	0.65 ± 0.01
	Sub8	0.41 ± 0.03	0.41 ± 0.01	0.71 ± 0.03
	Sub9	0.58 ± 0.09	0.44 ± 0.01	0.67 ± 0.04
	Mean	$\mathbf{0.68 \pm 0.02}$	0.57 ± 0.00	0.65 ± 0.01

5.5 Discussion

5.5.1 Estimating the Prediction Uncertainty: Deep Learning vs. Gaussian Processes

Considering our multi-variate Gaussian assumption on the distribution of the IMU signal of normal movements, multi-task Gaussian process regression (MTGPR) [184] seemed to be a natural choice for estimating the structured prediction uncertainty in normative modeling. However, MTGPR comes with extra computational overheads in time and space ($\mathbf{O}(n_N^3 p^3)$ and $\mathbf{O}(n_N^2 p^2)$ here n_N referes to the number of normal samples and p referes to the number of features) when computing the inverse cross-covariance matrices in the optimization and prediction phases. This problem is even more pronounced when dealing with multi-subject IMU-based abnormal movement detection when generally n_N is in order of 10^5 to 10^6. Despite extensive studies to reduce these computational barriers [185, 186, 187, 154],

the overall efficiency of the proposed approaches remained far below the minimum requirements in our target applications. To overcome this problem, in this study, we proposed to replace the MTGPR with a probabilistic DAE architecture for estimating the prediction uncertainties in the normative modeling framework. As supported by our experimental results, the estimated prediction uncertainties via DAE edged the novelty detection performance in comparison with the reconstruction-based approach. Our contribution facilitates the application of normative modeling on the large datasets (with large n_N or p) in the big-data era.

5.5.2 Normative Modeling vs. One-Class Classification

One-class classification [188] and more specifically one-class SVM is a common choice for solving novelty detection problems [149, 180, 181]. It is shown that one-class SVMs achieve poor performance on high-dimensional datasets, while a combination of a feature extraction method such as deep belief networks with one-class SVM enhances the performance of such novelty detection methods [170]. However, the prediction performance of one-class SVM is highly sensitive to its hyperparameters (e.g., in the case of RBF kernel ν and γ), especially on noisy data. This fact is well demonstrated in our experiments, where one-class SVM performed better when trained only on normal subjects; data, i.e., less noisy data (compare the results in Tables 5.2 and 5.4). Therefore, fine-tuning of one-class SVM hyperparameters is necessary; however, this is only possible if we have access to the labeled validation data during the model selection phase. This limitation leaves the only option of using default parameters when dealing with non-labeled data, which results in sub-optimal performances. The proposed deep normative modeling approach for novelty detection overcomes this barrier, as our experiments on three benchmark datasets show that its only hyperparameter, i.e., the dropout level, can be set to 0.1–0.4 without

a significant drop in the prediction performance.

5.5.3 Toward Modeling Human Normal Daily Movements Using Wearable Sensors

A possible solution to tackle the common issues in the supervised abnormal movement classification is to define the problem of abnormal movement detection in an unsupervised framework and try to assemble a probabilistic model of human normal daily movements. If successful, then in, for example, a novelty detection scenario, any large deviation from this model can be considered as an abnormal movement for the diagnosis and treatment of patients with motor deficiencies. Of course, learning a realistic representation of all possible human movements is very challenging due to the large set of possible movements, inter- and intra-subject heterogeneity and the prevalence of noisy samples. The proposed deep normative modeling method provides an early, but effective step toward this direction as it provides all the needed ingredients for modeling heterogeneous normal human movements in an unsupervised fashion.

5.5.4 Limitations and Future Work

Using DAE for learning P_N limits the application of the proposed method only to distance-based novelty detection approaches in the original and latent space; hence, it is not applicable in the density-based novelty detection [171]. This is because the DAE model is by nature unable to determine the density of normal data in the latent space. To address this problem, one possible future direction is to use generative alternative models instead of DAE such as variational autoencoders [189], adversarial autoencoders [190] or generative adversarial networks [191]. Another future direction is to use the proposed framework for implementing a real-time mobile application

for abnormal movement detection. The proposed DAE-based normative modeling approach, unlike its MTGPR-based alternatives, does not need to store huge inverse covariance matrices at the test time. Adding to this the low computational complexity of DAE at the prediction phase (just matrix multiplications and summations) and high potential for parallel programming (for computing MC repetitions), the proposed method offers a very well-suited approach for online mobile novelty detection applications.

5.6 Conclusions

In this study, we addressed the problem of automatic abnormal movement detection in ASD and PD patients in a novelty detection framework. In the normative modeling framework, we used a convolutional denoising autoencoder to learn the distribution of the normal human movements from the accelerometer signals. We showed how the normative modeling framework can be employed to quantify the deviation of each unseen sample from the normal movement samples. We demonstrated empirically that our proposed method outperforms two other baseline novelty detection methods on the SMM and Daphnet-FOG datasets. Our method: (i) overcomes the high computational complexities of estimating the prediction uncertainties in multi-task normative modeling, thus facilitating its application to large datasets in the big-data era; (ii) unlike the common one-class classification setting, our method relaxes the need for having access to the labeled validation data during the model selection phase; and more importantly, (iii) our method provides the first step toward modeling human normal daily movements using wearable sensors. The proposed approach gathers all the required ingredients for implementing a real-time mobile application for abnormal movement detection in the future.

Chapter 6

Conclusions

Developing a real-time mobile application for detecting abnormal movements in Parkinson's disease (PD) and autism spectrum disorder (ASD) patients can be considered as an ultimate goal in the context of automatic abnormal movement detection using wearable sensors. Developing such an application that can be incrementally adapted to new patients' data or on the same patient as his/her kinematics change in the real-life settings, could both advance research studies and provide new intervention tools that help researchers, clinicians, and caregivers to continuously monitor, better understand, and cope with abnormal behaviors in ASD and PD patients. Achieving this goal is, however, challenging due to the high inter and intra-subject variability in the acquired inertial measurement unit (IMU) data and having no access to the labels of collected data in real-life settings. The main focus of the research presented in this thesis was on addressing these particular challenges in automatic abnormal movement detection.

In this thesis, the problem of inter and intra-subject variability is attacked by using feature learning and transfer learning capabilities of deep neural networks. In particular, a new application of convolutional neural network (CNN) is proposed to learn an intermediate representation of

IMU signals in individuals with ASD. The proposed architecture is further used to transfer the knowledge, learned from one stereotypical motor movement (SMM) dataset to another dataset via the network initialization in a longitudinal scenario. The experimental results on one simulated and two real datasets indicate that i) substituting hand-crafted features with automatic feature learning using CNN results in more discriminative and robust features in across subject SMM detection scenarios; and ii) parameter transfer learning improves the adaptability and as a result, enhances the performance of the SMM detection system when applied to newly seen dataset in a longitudinal setting. Although the proposed deep neural network provides an accurate SMM detection model, the proposed fully supervised scheme for training the SMM detection model is problematic in the real-time adaptation as the adaptation to new unseen data should be performed only based on the input unlabeled data. This limitation motivates future researches on the online adaptation of the system in an unsupervised manner. Transductive transfer learning with the possibility of incremental training can possibly facilitate the online adaptation of an automatic abnormal movement detector in real-time scenarios.

In order to benefit more from the temporal fluctuations of IMU signals, a combination of long short-term memories (LSTM) with a CNN architecture is proposed to learn the dynamic features on a sequence of IMU signals. The experimental results demonstrate i) employing the temporal information improves the separability of SMM and no-SMM samples; ii) while the skewness of samples negatively affects the performance of the SMM detector in the static feature space, exploiting the temporal patterns of multi-sensor IMU signals recovers its performance. Thus, learning temporal information of IMU signals facilitates the application of SMM detection in more realistic environments where the incoming samples are highly skewed. Furthermore, in order to improve the reliability and sta-

bility of the SMM detector, an ensemble learning algorithm based on the best LSTM learners is proposed. The experimental results show a slight improvement in the mean performance compared to single LSTM classifiers. In addition, the ensemble of LSTMs provides a more reliable SMM detector in comparison to every single LSTM classifier by reducing the variability of results.

The problem of lack of labeled data in clinical applications of wearable sensors is addressed by redefining the abnormal movement detection problem in a novelty detection paradigm using the deep normative modeling method. To this end, heterogeneous normal movements in PD and ASD patients are modeled using a probabilistic denoising autoencoder. Then the resulting normative model is used to detect freezing of gaits (FOG) and SMMs in respectively PD and ASD patients. The experimental results on three benchmark datasets illustrate the effectiveness of the proposed method, which outperforms one-class SVM and the reconstruction-based novelty detection approaches. The proposed fully unsupervised novelty detection method achieved a relatively close performance to its supervised counterparts.

Possible future direction from this thesis can be summarized as follows:

1. **Evaluating our deep learning methods on a large dataset collected in the realistic conditions**: In this thesis, the deep learning methods are trained and evaluated on several benchmark IMU datasets collected from ASD and PD patients in controlled clinical conditions. To ensure the generalization across real-world datasets, it is necessary to evaluate the proposed methods on the large IMU datasets collected from patients in uncontrolled and wild real-life environments.

2. **A real-time implementation of abnormal movement detec-**

tion system: The proposed deep normative modeling method provides the possibility of modeling the normal human movement data during daily life activities without the need for having access to labeled samples. Furthermore, this method has the low computational complexity which makes it as an appropriate method to be implemented in the real-time abnormal detection movement detection in neuro-developmental and neuro-degenerative patients.

3. **Incorporating physiological data to predict abnormal movements**: Although IMU sensors are well-suited for the abnormal movement detection in ASD and PD patients, the collected data using these sensors do not always provide adequate information to achieve a reliable assessment on the abnormal movements out of clinical settings. For example, it is sometimes difficult to detect from the IMU data alone if a slowness of movement is due to bradykinesia or as a result of fatigue, anxiety, or other environmental factors. Since changes in physiological behaviors might occur during and before the onset of abnormal movements, thus monitoring and detecting these changes can help to have a higher abnormal movement detection performance. Incorporating physiological data captured by electrocardiography (ECG) and Electromyographic (EMG) sensors in the detection system provide the possibility of predicting the onset of abnormal movements rather than just detecting them.

Appendix A

Appendices

A.1 Generalized Extreme Value Distribution

For a random variable $x \in \mathbb{R}$, the cumulative distribution function of the GEVD, i.e., $F(x)$, is defined as below [176]:

$$F(x) = \begin{cases} e^{-[1+\xi(x-\mu)/\sigma]^{-1/\xi}}, & \xi \neq 0 \\ e^{-e^{[-(x-\mu)/\sigma]}}, & \xi = 0 \end{cases}, \tag{A.1}$$

$\mu \in \mathbb{R}$ and $\sigma > 0$ are the location and scale parameter, respectively. $\xi \in \mathbb{R}$ is the shape parameter and depending on whether $\xi < 0$, $\xi = 0$ or $\xi > 0$ the distribution follows the special cases of GEVD, namely Weibull, Gumbel and Fréchet, respectively.

Bibliography

[1] Ram Deepak Gottapu and Cihan H Dagli. Analysis of Parkinson's disease data. *Procedia Computer Science*, 140:334–341, 2018.

[2] Jiangbo He, Wu Zhou, Huijun Yu, Xiaoping He, and Peng Peng. Structural designing of a MEMS capacitive accelerometer for low temperature coefficient and high linearity. *Sensors*, 18(2):643, 2018.

[3] Vittorio Passaro, Antonello Cuccovillo, Lorenzo Vaiani, Martino De Carlo, and Carlo Edoardo Campanella. Gyroscope technology and applications: A review in the industrial perspective. *Sensors*, 17 (10):2284, 2017.

[4] Ajit Sharma, Faisal M Zaman, Babak V Amini, and Farrokh Ayazi. A high-Q in-plane SOI tuning fork gyroscope. In *Sensors, 2004. Proceedings of IEEE*, pages 467–470. IEEE, 2004.

[5] Nishkam Ravi, Nikhil Dandekar, Preetham Mysore, and Michael L Littman. Activity recognition from accelerometer data. In *Aaai*, volume 5, pages 1541–1546, 2005.

[6] Alireza Abedin Varamin, Ehsan Abbasnejad, Qinfeng Shi, Damith C Ranasinghe, and Hamid Rezatofighi. Deep auto-set: A deep auto-encoder-set network for activity recognition using wearables. 2018.

[7] Yu Guan and Thomas Ploetz. Ensembles of deep LSTM learn-

ers for activity recognition using wearables. *arXiv preprint arXiv:1703.09370*, 2017.

[8] Natalia Díaz-Rodríguez, Olmo Cadahía, Manuel Cuéllar, Johan Lilius, and Miguel Calvo-Flores. Handling real-world context awareness, uncertainty and vagueness in real-time human activity tracking and recognition with a fuzzy ontology-based hybrid method. *Sensors*, 14 (10):18131–18171, 2014.

[9] Natalia Díaz Rodríguez, Robin Wikström, Johan Lilius, Manuel Pegalajar Cuéllar, and Miguel Delgado Calvo Flores. Understanding movement and interaction: an ontology for kinect-based 3D depth sensors. In *Ubiquitous computing and ambient intelligence. Context-awareness and context-driven interaction*, pages 254–261. Springer, 2013.

[10] Natalia Díaz Rodríguez, Manuel P Cuéllar, Johan Lilius, and Miguel Delgado Calvo-Flores. A fuzzy ontology for semantic modelling and recognition of human behaviour. *Knowledge-Based Systems*, 66:46–60, 2014.

[11] Andreas Bulling, Ulf Blanke, and Bernt Schiele. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys (CSUR)*, 46(3):33, 2014.

[12] Nils Yannick Hammerla. Activity recognition in naturalistic environments using body-worn sensors. 2015.

[13] Aaron F Bobick. Movement, activity and action: the role of knowledge in the perception of motion. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 352(1358):1257–1265, 1997.

[14] Catherine Lord, Edwin H Cook, Bennett L Leventhal, and David G Amaral. Autism spectrum disorders. *Autism: The Science of Mental Health*, 28(2):217, 2013.

[15] Lonneke ML De Lau and Monique MB Breteler. Epidemiology of Parkinson's disease. *The Lancet Neurology*, 5(6):525–535, 2006.

[16] Victor Lotter. Epidemiology of autistic conditions in young children. *Social psychiatry*, 1(3):124–135, 1966.

[17] Michael G Aman, Nirbhay N Singh, Alistair W Stewart, and Carolyn J Field. Psychometric characteristics of the aberrant behavior checklist. *American journal of mental deficiency*, 1985.

[18] Sharon L Foster and John D Cone. Design and use of direct observation. *Handbook of behavioral assessment*, 2:253–324, 1986.

[19] Matthew S Goodwin, Marzieh Haghighi, Qu Tang, Murat Akcakaya, Deniz Erdogmus, and Stephen Intille. Moving towards a real-time system for automatically recognizing stereotypical motor movements in individuals on the autism spectrum using wireless accelerometry. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 861–872. ACM, 2014.

[20] Luc Lecavalier, Sarah Leone, and James Wiltz. The impact of behaviour problems on caregiver stress in young people with autism spectrum disorders. *Journal of Intellectual Disability Research*, 50 (3):172–183, 2006.

[21] Nicole Ciotti Gardenier, Rebecca MacDonald, and Gina Green. Comparison of direct observational methods for measuring stereotypic behavior in children with autism spectrum disorders. *Research in Developmental Disabilities*, 25(2):99–118, 2004.

[22] David AM Pyles, Mary M Riordan, and Jon S Bailey. The stereotypy analysis: An instrument for examining environmental variables associated with differential rates of stereotypic behavior. *Research in Developmental Disabilities*, 18(1):11–38, 1997.

[23] Johnny L Matson and Marie Nebel-Schwalm. Assessing challenging behaviors in children with autism spectrum disorders: A review. *Research in Developmental Disabilities*, 28(6):567–579, 2007.

[24] Parkinson's disease foundation. `http://www.pdf.org/`. Accessed: 2019-10-22.

[25] Bastiaan R Bloem, Jeffrey M Hausdorff, Jasper E Visser, and Nir Giladi. Falls and freezing of gait in Parkinson's disease: a review of two interconnected, episodic phenomena. *Movement disorders: official journal of the Movement Disorder Society*, 19(8):871–884, 2004.

[26] Joel S Perlmutter. Assessment of Parkinson disease manifestations. *Current protocols in neuroscience*, 49(1):10–1, 2009.

[27] Shen-Yang Lim and Ai Huey Tan. Historical perspective: The pros and cons of conventional outcome measures in Parkinson's disease. *Parkinsonism & related disorders*, 46:S47–S52, 2018.

[28] Natalia Díaz-Rodríguez, Stefan Grönroos, Frank Wickström, Johan Lilius, Henk Eertink, Andreas Braun, Paul Dillen, James Crowley, and Jan Alexandersson. An ontology for wearables data interoperability and ambient assisted living application development. In *Recent Developments and the New Direction in Soft-Computing Foundations and Applications*, pages 559–568. Springer, 2018.

[29] Natalia Díaz-Rodríguez, Aki Härmä, Rim Helaoui, Ignacio Huitzil, Fernando Bobillo, and Umberto Straccia. Couch potato or gym ad-

dict? semantic lifestyle profiling with wearables and fuzzy knowledge graphs.

[30] Marco Iosa, Pietro Picerno, Stefano Paolucci, and Giovanni Morone. Wearable inertial sensors for human movement analysis. *Expert review of medical devices*, 13(7):641–659, 2016.

[31] Dongni Johansson, Kristina Malmgren, and Margit Alt Murphy. Wearable sensors for clinical applications in epilepsy, Parkinson's disease, and stroke: a mixed-methods systematic review. *Journal of neurology*, pages 1–13, 2018.

[32] Lara Allet, Ruud H Knols, Kei Shirato, and Eling D de Bruin. Wearable systems for monitoring mobility-related activities in chronic disease: a systematic review. *Sensors*, 10(10):9026–9052, 2010.

[33] Tracy Westeyn, Kristin Vadas, Xuehai Bian, Thad Starner, and Gregory D Abowd. Recognizing mimicked autistic self-stimulatory behaviors using HMMs. In *Ninth IEEE International Symposium on Wearable Computers (ISWC'05)*, pages 164–167. IEEE, 2005.

[34] Cheol-Hong Min and Ahmed H Tewfik. Automatic characterization and detection of behavioral patterns using linear predictive coding of accelerometer sensor data. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pages 220–223. IEEE, 2010.

[35] Hong Min Cheol and Ahmed H Tewfik. Novel pattern detection in children with autism spectrum disorder using iterative subspace identification. In *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2266–2269. IEEE, 2010.

[36] Fahd Albinali, Matthew S Goodwin, and Stephen Intille. Detecting stereotypical motor movements in the classroom using accelerometry and pattern recognition algorithms. *Pervasive and Mobile Computing*, 8(1):103–114, 2012.

[37] Nastaran Mohammadian Rad and Cesare Furlanello. Applying deep learning to stereotypical motor movement detection in autism spectrum disorders. In *Data Mining Workshops (ICDMW), 2016 IEEE 16th International Conference on*, pages 1235–1242. IEEE, 2016.

[38] Nastaran Mohammadian Rad, Andrea Bizzego, Seyed Mostafa Kia, Giuseppe Jurman, Paola Venuti, and Cesare Furlanello. Convolutional neural network for stereotypical motor movement detection in autism. *arXiv preprint arXiv:1511.01865*, 2015.

[39] Nastaran Mohammadian Rad, Seyed Mostafa Kia, Calogero Zarbo, Twan van Laarhoven, Giuseppe Jurman, Paola Venuti, Elena Marchiori, and Cesare Furlanello. Deep learning for automatic stereotypical motor movement detection using wearable sensors in autism spectrum disorders. *Signal Processing*, 144:180–191, 2018.

[40] Daniel Roggen, Kilian Forster, Alberto Calatroni, Thomas Holleczek, Yu Fang, Gerhard Troster, Alois Ferscha, Clemens Holzmann, Andreas Riener, Paul Lukowicz, et al. OPPORTUNITY: Towards opportunistic activity and context recognition systems. In *World of Wireless, Mobile and Multimedia Networks & Workshops, 2009. WoWMoM 2009. IEEE International Symposium on a*, pages 1–6. IEEE, 2009.

[41] Nastaran Mohammadian Rad, Twan van Laarhoven, Cesare Furlanello, and Elena Marchiori. Novelty detection using deep normative modeling for IMU-based abnormal movement monitoring in

Parkinson's disease and autism spectrum disorders. *Sensors*, 18(10): 3533, 2018.

[42] Andreas Bulling, Jamie A Ward, and Hans Gellersen. Multimodal recognition of reading activity in transit using body-worn sensors. *ACM Transactions on Applied Perception (TAP)*, 9(1):2, 2012.

[43] Catherine Lord, Edwin H Cook, Bennett L Leventhal, and David G Amaral. Autism spectrum disorders. *Neuron*, 28(2):355–363, 2000.

[44] Uta Frith and Francesca Happé. Autism spectrum disorder. *Current biology*, 15(19):R786–R790, 2005.

[45] A Brendel, G Dawson, L Elder, PF Gerhardt, K Margo, et al. 100-day kit for newly diagnosed families of school age children. *no. december*, 2014.

[46] Hadeel Faras, Nahed Al Ateeqi, and Lee Tidmarsh. Autism spectrum disorders. *Annals of Saudi medicine*, 30(4):295, 2010.

[47] Allison B Cunningham and Laura Schreibman. Stereotypy in autism: The importance of function. *Research in autism spectrum disorders*, 2(3):469–479, 2008.

[48] Marc J Lanovaz and Ingrid E Sladeczek. Vocal stereotypy in individuals with autism spectrum disorders: A review of behavioral interventions. *Behavior Modification*, 36(2):146–164, 2012.

[49] Mackenzie Katherine. Stereotypic movement disorders. In *Seminars in pediatric neurology*, volume 25, pages 19–24. Elsevier, 2018.

[50] Robert Didden, Peter Sturmey, Jeff Sigafoos, Russell Lang, Mark F OReilly, and Giulio E Lancioni. Nature, prevalence, and characteristics of challenging behavior. In *Functional assessment for challenging behaviors*, pages 25–44. Springer, 2012.

[51] Marc J Lanovaz, Kirsty M Robertson, Kara Soerono, and Nicholas Watkins. Effects of reducing stereotypy on other behaviors: A systematic review. *Research in Autism Spectrum Disorders*, 7(10):1234–1243, 2013.

[52] Sabine S Chebli, Valérie Martin, and Marc J Lanovaz. Prevalence of stereotypy in individuals with developmental disabilities: A systematic review. *Review Journal of Autism and Developmental Disorders*, 3(2):107–118, 2016.

[53] R Matthew Reese, David M Richman, John M Belmont, and Paige Morse. Functional characteristics of disruptive behavior in developmentally disabled children with and without autism. *Journal of Autism and Developmental Disorders*, 35(4):419–428, 2005.

[54] John G Nutt and G Frederick Wooten. Diagnosis and initial management of Parkinson's disease. *New England Journal of Medicine*, 353(10):1021–1027, 2005.

[55] Ryan J. Uitti, Yasuhiko Baba, Zbigniew K. Wszolek, and D. John Putzke. Defining the Parkinson's disease phenotype: initial symptoms and baseline characteristics in a clinical cohort. *Parkinsonism and Related Disorders*, 11(3):139 – 145, 2005. ISSN 1353-8020. doi: https://doi.org/10.1016/j.parkreldis.2004.10.007. URL `http://www.sciencedirect.com/science/article/pii/S1353802004001634`.

[56] Yasuyuki Okuma. Freezing of gait in Parkinson's disease. *Journal of neurology*, 253(7):vii27–vii32, 2006.

[57] Nir Giladi, Richard Kao, and Stanley Fahn. Freezing phenomenon in patients with parkinsonian syndromes. *Movement disorders: official journal of the Movement Disorder Society*, 12(3):302–305, 1997.

[58] Roongroj Bhidayasiri and Pablo Martinez-Martin. Clinical assessments in Parkinson's disease: scales and monitoring. In *International review of neurobiology*, volume 132, pages 129–182. Elsevier, 2017.

[59] Walter Maetzler, Josefa Domingos, Karin Srulijes, Joaquim J Ferreira, and Bastiaan R Bloem. Quantitative wearable sensors for objective assessment of Parkinson's disease. *Movement Disorders*, 28 (12):1628–1637, 2013.

[60] JM Shine, ST Moore, SJ Bolitho, TR Morris, V Dilda, SL Naismith, and SJG Lewis. Assessing the utility of freezing of gait questionnaires in Parkinson's disease. *Parkinsonism & related disorders*, 18(1):25–29, 2012.

[61] Valerie AJ Block, Erica Pitsch, Peggy Tahir, Bruce AC Cree, Diane D Allen, and Jeffrey M Gelfand. Remote physical activity monitoring in neurological disease: a systematic review. *PloS one*, 11(4):e0154335, 2016.

[62] Anindya Nag, Subhas Chandra Mukhopadhyay, and Jürgen Kosel. Wearable flexible sensors: A review. *IEEE Sensors Journal*, 17(13): 3949–3960, 2017.

[63] Donghee Son, Jongha Lee, Shutao Qiao, Roozbeh Ghaffari, Jaemin Kim, Ji Eun Lee, Changyeong Song, Seok Joo Kim, Dong Jun Lee, Samuel Woojoo Jun, et al. Multifunctional wearable devices for diagnosis and therapy of movement disorders. *Nature nanotechnology*, 9(5):397, 2014.

[64] Alejandro Rodríguez-Molinero, Albert Samà, Carlos Pérez-López, Daniel Rodríguez-Martín, Sheila Alcaine, Berta Mestre, Paola Quispe, Benedetta Giuliani, Gabriel Vainstein, Patrick Browne, et al.

Analysis of correlation between an accelerometer-based algorithm for detecting Parkinsonian gait and UPDRS subscales. *Frontiers in neurology*, 8:431, 2017.

[65] Merryn J Mathie, Adelle CF Coster, Nigel H Lovell, and Branko G Celler. Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement. *Physiological measurement*, 25(2):R1, 2004.

[66] Emil Chiauzzi, Carlos Rodarte, and Pronabesh DasMahapatra. Patient-centered activity monitoring in the self-management of chronic health conditions. *BMC medicine*, 13(1):77, 2015.

[67] Lotta Hamari, Tiina Kullberg, Jukka Ruohonen, Olli J Heinonen, Natalia Díaz-Rodríguez, Johan Lilius, Anni Pakarinen, Annukka Myllymäki, Ville Leppänen, and Sanna Salanterä. Physical activity among children: Objective measurements using fitbit one® and actigraph. *BMC research notes*, 10(1):161, 2017.

[68] Navid Yazdi, Farrokh Ayazi, and Khalil Najafi. Micromachined inertial sensors. *Proceedings of the IEEE*, 86(8):1640–1659, 1998.

[69] J Včelák, P Ripka, J Kubik, A Platil, and P Kašpar. AMR navigation systems and methods of their calibration. *Sensors and Actuators A: Physical*, 123:122–128, 2005.

[70] Louise Foley and Ralph Maddison. Use of active video games to increase physical activity in children: a (virtual) reality? *Pediatric exercise science*, 22(1):7–20, 2010.

[71] Michael B Del Rosario, Stephen J Redmond, and Nigel H Lovell. Tracking the evolution of smartphone sensing for monitoring human movement. *Sensors*, 15(8):18901–18933, 2015.

[72] Alain Béliveau, Guy T Spencer, Keith A Thomas, and Scott L Roberson. Evaluation of MEMS capacitive accelerometers. *IEEE Design & Test of Computers*, 16(4):48–56, 1999.

[73] Steven M LaValle, Anna Yershova, Max Katsev, and Michael Antonov. Head tracking for the oculus rift. In *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pages 187–194. IEEE, 2014.

[74] Davide Figo, Pedro C Diniz, Diogo R Ferreira, and João M Cardoso. Preprocessing techniques for context recognition from accelerometer data. *Personal and Ubiquitous Computing*, 14(7):645–662, 2010.

[75] Dunzhu Xia, Cheng Yu, and Lun Kong. The development of micromachined gyroscope structure and circuitry technology. *Sensors*, 14 (1):1394–1473, 2014.

[76] James Lenz and S Edelstein. Magnetic sensors and their applications. *IEEE Sensors journal*, 6(3):631–649, 2006.

[77] Shyamal Patel, Hyung Park, Paolo Bonato, Leighton Chan, and Mary Rodgers. A review of wearable sensors and systems with application in rehabilitation. *Journal of neuroengineering and rehabilitation*, 9 (1):21, 2012.

[78] Huiyu Zhou and Huosheng Hu. Human motion tracking for rehabilitationa survey. *Biomedical Signal Processing and Control*, 3(1):1–18, 2008.

[79] JHM Bergmann and AH McGregor. Body-worn sensor design: what do patients and clinicians want? *Annals of biomedical engineering*, 39(9):2299–2312, 2011.

[80] Oresti Banos, Mate Toth, Miguel Damas, Hector Pomares, and Ignacio Rojas. Dealing with the effects of sensor displacement in wearable activity recognition. *Sensors*, 14(6):9995–10023, 2014.

[81] Kilian Förster, Pascal Brem, Daniel Roggen, and Gerhard Tröster. Evolving discriminative features robust to sensor displacement for activity recognition in body area sensor networks. In *2009 International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, pages 43–48. IEEE, 2009.

[82] Ulf Großekathöfer, Nikolay V Manyakov, Vojkan Mihajlović, Gahan Pandina, Andrew Skalkin, Seth Ness, Abigail Bangerter, and Matthew S Goodwin. Automated detection of stereotypical motor movements in autism spectrum disorder using recurrence quantification analysis. *Frontiers in neuroinformatics*, 11:9, 2017.

[83] Ulf Großekathöfer, Nikolay V. Manyakov, Vojkan Mihajlovi, Gahan Pandina, Andrew Skalkin, Seth Ness, Abigail Bangerter, and Matthew S. Goodwin. Automated detection of stereotypical motor movements in autism spectrum disorder using recurrence quantification analysis. *Frontiers in Neuroinformatics*, 11:9, 2017. ISSN 1662-5196. doi: 10.3389/fninf.2017.00009. URL `http://journal.frontiersin.org/article/10.3389/fninf.2017.00009`.

[84] Cheol-Hong Min, Ahmed H Tewfik, Youngchun Kim, and Rigel Menard. Optimal sensor location for body sensor network to detect self-stimulatory behaviors of children with autism spectrum disorder. In *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, pages 3489–3492. IEEE, 2009.

[85] Oresti Banos, Juan-Manuel Galvez, Miguel Damas, Hector Pomares,

and Ignacio Rojas. Window size impact in human activity recognition. *Sensors*, 14(4):6474–6499, 2014.

[86] Benish Fida, Ivan Bernabucci, Daniele Bibbo, Silvia Conforto, and Maurizio Schmid. Varying behavior of different window sizes on the classification of static and dynamic physical activities from a single accelerometer. *Medical engineering & physics*, 37(7):705–711, 2015.

[87] Maja Stikic, Tâm Huynh, Kristof Van Laerhoven, and Bernt Schiele. ADL recognition based on the combination of RFID and accelerometer sensing. In *Pervasive Computing Technologies for Healthcare, 2008. PervasiveHealth 2008. Second International Conference on*, pages 258–263. IEEE, 2008.

[88] Hector P Martinez, Yoshua Bengio, and Georgios N Yannakakis. Learning deep physiological models of affect. *Computational Intelligence Magazine, IEEE*, 8(2):20–33, 2013.

[89] Francisco Javier Ordóñez and Daniel Roggen. Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. *Sensors*, 16(1):115, 2016.

[90] Nils Y Hammerla, Shane Halloran, and Thomas Ploetz. Deep, convolutional, and recurrent models for human activity recognition using wearables. *arXiv preprint arXiv:1604.08880*, 2016.

[91] Julia Camps, Albert Sama, Mario Martin, Daniel Rodriguez-Martin, Carlos Perez-Lopez, Joan M Moreno Arostegui, Joan Cabestany, Andreu Catala, Sheila Alcaine, Berta Mestre, et al. Deep learning for freezing of gait detection in Parkinson's disease patients in their homes using a waist-worn inertial measurement unit. *Knowledge-Based Systems*, 139:119–131, 2018.

[92] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. springer, New York, 2006.

[93] Yutaka Sasaki et al. The truth of the F-measure. *Teach Tutor mater*, 1(5):1–5, 2007.

[94] Tom Fawcett. An introduction to ROC analysis. *Pattern recognition letters*, 27(8):861–874, 2006.

[95] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, 323 (6088):533, 1986.

[96] Ossama Abdel-Hamid, Abdel-rahman Mohamed, Hui Jiang, and Gerald Penn. Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition. In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, pages 4277–4280. IEEE, 2012.

[97] Yann LeCun and Yoshua Bengio. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10), 1995.

[98] Jose Bernal, Kaisar Kushibar, Daniel S Asfaw, Sergi Valverde, Arnau Oliver, Robert Martí, and Xavier Lladó. Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: a review. *Artificial intelligence in medicine*, 2018.

[99] Sergey Levine, Peter Pastor, Alex Krizhevsky, Julian Ibarz, and Deirdre Quillen. Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection. *The International Journal of Robotics Research*, 37(4-5):421–436, 2018.

[100] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[101] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[102] Zachary C Lipton, David C Kale, Charles Elkan, and Randall Wetzel. Learning to diagnose with LSTM recurrent neural networks. *arXiv preprint arXiv:1511.03677*, 2015.

[103] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pages 3104–3112, 2014.

[104] Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton. Speech recognition with deep recurrent neural networks. In *Acoustics, speech and signal processing (icassp), 2013 ieee international conference on*, pages 6645–6649. IEEE, 2013.

[105] Alex Graves and Navdeep Jaitly. Towards end-to-end speech recognition with recurrent neural networks. In *International Conference on Machine Learning*, pages 1764–1772, 2014.

[106] Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. *science*, 313(5786):504–507, 2006.

[107] Ruslan Salakhutdinov and Geoffrey Hinton. Semantic hashing. *International Journal of Approximate Reasoning*, 50(7):969–978, 2009.

[108] SoHyun Lee, Samuel L Odom, and Rachel Loftin. Social engagement with peers and stereotypic behavior of children with autism. *Journal of Positive Behavior Interventions*, 9(2):67–79, 2007.

[109] Rachel L Loftin, Samuel L Odom, and Johanna F Lantz. Social interaction and repetitive motor behaviors. *Journal of Autism and Developmental Disorders*, 38(6):1124–1135, 2008.

[110] Nuri Firat Ince, Cheol-Hong Min, Ahmed Tewfik, and David Vanderpool. Detection of early morning daily activities with static home and wearable wireless sensors. *EURASIP Journal on Advances in Signal Processing*, 2008:31, 2008.

[111] Robert Glenn Stockwell, Lalu Mansinha, and RP Lowe. Localization of the complex spectrum: the S transform. *Signal Processing, IEEE Transactions on*, 44(4):998–1001, 1996.

[112] Kunihiko Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, 36(4):193–202, 1980.

[113] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[114] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[115] Piotr W Mirowski, Yann LeCun, Deepak Madhavan, and Ruben Kuzniecky. Comparing SVM and convolutional networks for epileptic seizure prediction from intracranial EEG. In *Machine Learning for Signal Processing, 2008. MLSP 2008. IEEE Workshop on*, pages 244–249. IEEE, 2008.

[116] Jian Bo Yang, Minh Nhut Nguyen, Phyo Phyo San, Xiao Li Li, and Shonali Krishnaswamy. Deep convolutional neural networks on mul-

tichannel time series for human activity recognition. In *Proceedings of the 24th International Conference on Artificial Intelligence*, pages 3995–4001. AAAI Press, 2015.

[117] Ming Zeng, Le T Nguyen, Bo Yu, Ole J Mengshoel, Jiang Zhu, Pang Wu, and Juyong Zhang. Convolutional neural networks for human activity recognition using mobile sensors. In *Mobile Computing, Applications and Services (MobiCASE), 2014 6th International Conference on*, pages 197–205. IEEE, 2014.

[118] T. Zebin, P. J. Scully, and K. B. Ozanyan. Human activity recognition with inertial sensors using a deep learning approach. In *2016 IEEE SENSORS*, pages 1–3, Oct 2016. doi: 10.1109/ICSENS.2016.7808590.

[119] Francisco Javier Ordóñez Morales and Daniel Roggen. Deep convolutional feature transfer across mobile activity recognition domains, sensor modalities and locations. In *Proceedings of the 2016 ACM International Symposium on Wearable Computers*, pages 92–99. ACM, 2016.

[120] Nitish Srivastava, Geoffrey E Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1):1929–1958, 2014.

[121] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *Knowledge and Data Engineering, IEEE Transactions on*, 22(10): 1345–1359, 2010.

[122] Emmanuel Munguia Tapia, Stephen S Intille, Louis Lopez, and Kent Larson. The design of a portable kit of wireless sensors for naturalistic

data collection. In *International Conference on Pervasive Computing*, pages 117–134. Springer, 2006.

[123] Fahd Albinali, Matthew S Goodwin, and Stephen S Intille. Recognizing stereotypical motor movements in the laboratory and classroom: a case study with children on the autism spectrum. In *Proceedings of the 11th international conference on Ubiquitous computing*, pages 71–80. ACM, 2009.

[124] François Chollet et al. Keras. `https://keras.io`, 2015.

[125] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(Nov):2579–2605, 2008.

[126] Richard O Duda, Peter E Hart, and David G Stork. *Pattern classification*. John Wiley & Sons, 2012.

[127] Oscar D Lara and Miguel A Labrador. A survey on human activity recognition using wearable sensors. *IEEE Communications Surveys and Tutorials*, 15(3):1192–1209, 2013.

[128] Miguel A Labrador and Oscar D Lara Yejas. *Human Activity Recognition: Using Wearable Sensors and Smartphones*. CRC Press, 2013.

[129] Jeffrey W Lockhart and Gary M Weiss. Limitations with activity recognition methodology & data sets. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, pages 747–756. ACM, 2014.

[130] Martin Berchtold, Matthias Budde, Hedda R Schmidtke, and Michael Beigl. An extensible modular recognition concept that makes activity recognition practical. In *Annual Conference on Artificial Intelligence*, pages 400–409. Springer, 2010.

[131] Oscar D Lara, Alfredo J Pérez, Miguel A Labrador, and José D Posada. Centinela: A human activity recognition system based on acceleration and vital sign data. *Pervasive and mobile computing*, 8 (5):717–729, 2012. doi: dx.doi.org/10.1016/j.pmcj.2011.06.004.

[132] Florian Eyben, Martin Wöllmer, Björn Schuller, and Alex Graves. From speech to letters-using a novel neural network architecture for grapheme based asr. In *Automatic Speech Recognition & Understanding, 2009. ASRU 2009. IEEE Workshop on*, pages 376–380. IEEE, 2009.

[133] Patrick Doetsch, Michal Kozielski, and Hermann Ney. Fast and robust training of recurrent neural networks for offline handwriting recognition. In *Frontiers in Handwriting Recognition (ICFHR), 2014 14th International Conference on*, pages 279–284. IEEE, 2014.

[134] Wojciech Zaremba, Ilya Sutskever, and Oriol Vinyals. Recurrent neural network regularization. *arXiv preprint arXiv:1409.2329*, 2014.

[135] Thomas G Dietterich. Ensemble learning. In M. Arbib, editor, *The Handbook of Brain Theory and Neural Networks*, volume 2, pages 110–125. MIT Press: Cambridge, MA, 2002.

[136] Ludmila I. Kuncheva. *Combining Pattern Classifiers: Methods and Algorithms*. Wiley-Interscience, 1 edition, July 2004. ISBN 0471210781. URL `http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0471210781`.

[137] Li Deng and John Platt. Ensemble deep learning for speech recognition. 2014.

[138] Lin-peng Jin and Jun Dong. Ensemble deep learning for biomedi-

cal time series classification. *Computational intelligence and neuroscience*, 2016, 2016. doi: 10.1155/2016/6212684.

[139] Bartosz Krawczyk, Leandro L. Minku, Joo Gama, Jerzy Stefanowski, and Micha Woniak. Ensemble learning for data stream analysis: A survey. *Information Fusion*, 37:132 – 156, 2017. ISSN 1566-2535. doi: http://dx.doi.org/10.1016/j.inffus.2017.02.004. URL `http://www.sciencedirect.com/science/article/pii/S1566253516302329`.

[140] Heitor Murilo Gomes, Jean Paul Barddal, Fabrício Enembreck, and Albert Bifet. A survey on ensemble learning for data stream classification. *ACM Comput. Surv.*, 50(2):23:1–23:36, March 2017. ISSN 0360-0300. doi: 10.1145/3054925. URL `http://doi.acm.org/10.1145/3054925`.

[141] Tijmen Tieleman and Geoffrey Hinton. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural Networks for Machine Learning*, 4(2), 2012.

[142] Matthew S Goodwin, Stephen S Intille, Fahd Albinali, and Wayne F Velicer. Automated detection of stereotypical motor movements. *Journal of autism and developmental disorders*, 41(6):770–782, 2011.

[143] Dorra Trabelsi, Samer Mohammed, Faicel Chamroukhi, Latifa Oukhellou, and Yacine Amirat. An unsupervised approach for automatic activity recognition based on hidden Markov model regression. *IEEE Transactions on Automation Science and Engineering*, 10(3): 829–835, 2013.

[144] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):15, 2009.

[145] Clayton Scott and Gilles Blanchard. Novelty detection: Unlabeled data definitely help. In *Artificial Intelligence and Statistics*, pages 464–471, 2009.

[146] Nastaran Mohammadian Rad, Seyed Mostafa Kia, Calogero Zarbo, Giuseppe Jurman, Paola Venuti, and Cesare Furlanello. Stereotypical motor movement detection in dynamic feature space. In *Data Mining Workshops (ICDMW), 2016 IEEE 16th International Conference on*, pages 487–494. IEEE, 2016.

[147] Sinziana Mazilu, Ulf Blanke, Alberto Calatroni, Eran Gazit, Jeffrey M. Hausdorff, and Gerhard Trster. The role of wrist-mounted inertial sensors in detecting gait freeze episodes in Parkinson's disease. *Pervasive and Mobile Computing*, 33:1 – 16, 2016. ISSN 1574-1192. doi: https://doi.org/10.1016/j.pmcj.2015.12.007. URL `http://www.sciencedirect.com/science/article/pii/S157411921600002X`.

[148] Andre F Marquand, Thomas Wolfers, Maarten Mennes, Jan Buitelaar, and Christian F Beckmann. Beyond lumping and splitting: a review of computational approaches for stratifying psychiatric disorders. *Biological psychiatry: cognitive neuroscience and neuroimaging*, 1(5):433–447, 2016.

[149] Marco AF Pimentel, David A Clifton, Lei Clifton, and Lionel Tarassenko. A review of novelty detection. *Signal Processing*, 99: 215–249, 2014.

[150] VVRPV Jyothsna, VV Rama Prasad, and K Munivara Prasad. A review of anomaly based intrusion detection systems. *International Journal of Computer Applications*, 28(7):26–35, 2011.

[151] Shijoe Jose, D Malathi, Bharath Reddy, and Dorathi Jayaseeli. A survey on anomaly based host intrusion detection system. In *Jour-*

nal of Physics: Conference Series, volume 1000, page 012049. IOP Publishing, 2018.

[152] Lionel Tarassenko, Paul Hayton, Nicholas Cerneaz, and Michael Brady. Novelty detection for the identification of masses in mammograms. 1995.

[153] John A Quinn and Christopher KI Williams. Known unknowns: Novelty detection in condition monitoring. In *Iberian Conference on Pattern Recognition and Image Analysis*, pages 1–6. Springer, 2007.

[154] Seyed Mostafa Kia, Christian F. Backmann, and Andre F. Marquand. Scalable multi-task gaussian process tensor regression for normative modeling of structured variation in neuroimaging data. *arXiv preprint arXiv:1808.00036*, 2018.

[155] Yu Zhang, Chris Bingham, Miguel Martínez-García, and Darren Cox. Detection of emerging faults on industrial gas turbines using extended Gaussian mixture models. *International Journal of Rotating Machinery*, 2017, 2017.

[156] Marek Słoński. Gaussian mixture model for time series-based structural damage detection. *Computer Assisted Methods in Engineering and Science*, 19(4):331–338, 2017.

[157] Ling-Jyh Chen, Yao-Hua Ho, Hsin-Hung Hsieh, Shih-Ting Huang, Hu-Cheng Lee, and Sachit Mahajan. ADF: an anomaly detection framework for large-scale PM2. 5 sensing systems. *IEEE Internet of Things Journal*, 5(2):559–570, 2018.

[158] Raihan Ul Islam, Mohammad Shahadat Hossain, and Karl Andersson. A novel anomaly detection algorithm for sensor data under uncertainty. *Soft Computing*, 22(5):1623–1639, 2018.

[159] Mohammad Sabokrou, Mohsen Fayyaz, Mahmood Fathy, Zahra Moayed, and Reinhard Klette. Deep-anomaly: Fully convolutional neural network for fast anomaly detection in crowded scenes. *Computer Vision and Image Understanding*, 2018.

[160] Waqas Sultani, Chen Chen, and Mubarak Shah. Real-world anomaly detection in surveillance videos. *Center for Research in Computer Vision (CRCV), University of Central Florida (UCF)*, 2018.

[161] Wael Khreich, Babak Khosravifar, Abdelwahab Hamou-Lhadj, and Chamseddine Talhi. An anomaly detection system based on variable N-gram features and one-class SVM. *Information and Software Technology*, 91:186–197, 2017.

[162] Risto Vaarandi, Bernhards Blumbergs, and Markus Kont. An unsupervised framework for detecting anomalous messages from syslog log files. In *NOMS 2018-2018 IEEE/IFIP Network Operations and Management Symposium*. IEEE, 2018.

[163] Andre F Marquand, Iead Rezek, Jan Buitelaar, and Christian F Beckmann. Understanding heterogeneity in clinical cohorts using normative models: beyond case-control studies. *Biological psychiatry*, 80 (7):552–561, 2016.

[164] Yarin Gal and Zoubin Ghahramani. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059, 2016.

[165] Evanthia E Tripoliti, Alexandros T Tzallas, Markos G Tsipouras, George Rigas, Panagiota Bougia, Michael Leontiou, Spiros Konitsiotis, Maria Chondrogiorgi, Sofia Tsouli, and Dimitrios I Fotiadis. Automatic detection of freezing of gait events in patients with Parkin-

son's disease. *Computer methods and programs in biomedicine*, 110 (1):12–26, 2013.

[166] Sinziana Mazilu, Michael Hardegger, Zack Zhu, Daniel Roggen, Gerhard Troster, Meir Plotnik, and Jeffrey M Hausdorff. Online detection of freezing of gait with smartphones and machine learning techniques. In *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2012 6th International Conference on*, pages 123–130. IEEE, 2012.

[167] Daniel Rodríguez-Martín, Albert Samà, Carlos Pérez-López, Andreu Català, Joan M Moreno Arostegui, Joan Cabestany, Àngels Bayés, Sheila Alcaine, Berta Mestre, Anna Prats, et al. Home detection of freezing of gait using support vector machines through a single waist-worn triaxial accelerometer. *PloS one*, 12(2):e0171764, 2017.

[168] Guglielmo Cola, Marco Avvenuti, Alessio Vecchio, Guang-Zhong Yang, Benny Lo, et al. An on-node processing approach for anomaly detection in gait. *IEEE Sensors Journal*, 15(11):6640–6649, 2015.

[169] Trong-Nguyen Nguyen, Huu-Hung Huynh, and Jean Meunier. Skeleton-based abnormal gait detection. *Sensors*, 16(11):1792, 2016.

[170] Sarah M Erfani, Sutharshan Rajasegarar, Shanika Karunasekera, and Christopher Leckie. High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning. *Pattern Recognition*, 58:121–134, 2016.

[171] Aleksei Vasilev, Vladimir Golkov, Ilona Lipp, Eleonora Sgarlata, Valentina Tomassini, Derek K Jones, and Daniel Cremers. q-space novelty detection with variational autoencoders. *arXiv preprint arXiv:1806.02997*, 2018.

[172] Xiaoran Chen, Nick Pawlowski, Martin Rajchl, Ben Glocker, and Ender Konukoglu. Deep generative models in the real-world: An open challenge from medical imaging. *arXiv preprint arXiv:1806.05452*, 2018.

[173] Shehroz S Khan and Babak Taati. Detecting unseen falls from wearable devices using channel-wise ensemble of autoencoders. *Expert Systems with Applications*, 87:280–290, 2017.

[174] Gabriel Ziegler, Gerard R Ridgway, Robert Dahnke, Christian Gaser, Alzheimer's Disease Neuroimaging Initiative, et al. Individualized Gaussian process-based prediction and detection of local and global gray matter abnormalities in elderly subjects. *NeuroImage*, 97:333–348, 2014.

[175] Stuart Coles, Joanna Bawa, Lesley Trenner, and Pat Dorazio. *An introduction to statistical modeling of extreme values*, volume 208. Springer, 2001.

[176] Anthony C Davison and Raphaël Huser. Statistics of extremes. *Annual Review of Statistics and its Application*, 2:203–235, 2015.

[177] Stephen J Roberts. Extreme value statistics for novelty detection in biomedical data processing. *IEE Proceedings-Science, Measurement and Technology*, 147(6):363–367, 2000.

[178] Bernhard Schölkopf, John C. Platt, John C. Shawe-Taylor, Alex J. Smola, and Robert C. Williamson. Estimating the support of a high-dimensional distribution. *Neural Comput.*, 13(7):1443–1471, 2001. ISSN 0899-7667. doi: 10.1162/089976601750264965. URL `https://doi.org/10.1162/089976601750264965`.

[179] Marc Bachlin, Daniel Roggen, Gerhard Troster, Meir Plotnik, Noit Inbar, Inbal Meidan, Talia Herman, Marina Brozgol, Eliya Shaviv, Nir Giladi, et al. Potentials of enhanced context awareness in wearable assistants for Parkinson's disease patients with the freezing of gait syndrome. In *Wearable Computers, 2009. ISWC'09. International Symposium on*, pages 123–130. IEEE, 2009.

[180] Sutharshan Rajasegarar, Christopher Leckie, James C Bezdek, and Marimuthu Palaniswami. Centered hyperspherical and hyperellipsoidal one-class support vector machines for anomaly detection in sensor networks. *IEEE Transactions on Information Forensics and Security*, 5(3):518–533, 2010.

[181] Lei A Clifton, Hujun Yin, and Yang Zhang. Support vector machine in novelty detection for multi-channel combustion data. In *International Symposium on Neural Networks*, pages 836–843. Springer, 2006.

[182] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[183] Jesse Davis and Mark Goadrich. The relationship between precision-recall and ROC curves. In *Proceedings of the 23rd international conference on Machine learning*, pages 233–240. ACM, 2006.

[184] Edwin V Bonilla, Kian M Chai, and Christopher Williams. Multi-task Gaussian process prediction. In *Advances in neural information processing systems*, pages 153–160, 2008.

[185] Mauricio A Álvarez and Neil D Lawrence. Computationally efficient convolved multiple output Gaussian processes. *Journal of Machine Learning Research*, 12:1459–1500, 2011.

[186] Barbara Rakitsch, Christoph Lippert, Karsten Borgwardt, and Oliver Stegle. It is all in the noise: Efficient multi-task Gaussian process inference with structured residuals. In *Advances in neural information processing systems*, pages 1466–1474, 2013.

[187] Seyed Mostafa Kia and Andre Marquand. Normative modeling of neuroimaging data using scalable multi-task Gaussian processes. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018.

[188] Mary M Moya, Mark W Koch, and Larry D Hostetler. One-class classifier networks for target recognition applications. *NASA STI/Recon Technical Report N*, 93, 1993.

[189] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[190] Alireza Makhzani, Jonathon Shlens, Navdeep Jaitly, Ian Goodfellow, and Brendan Frey. Adversarial autoencoders. *arXiv preprint arXiv:1511.05644*, 2015.

[191] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

Publications

- **Nastaran Mohammadian Rad**, Twan van Laarhoven, Cesare Furlanello, and Elena Marchiori. "Novelty detection using deep normative modeling for IMU-based abnormal movement monitoring in parkinsons disease and autism spectrum disorders." Sensors, 18(10): 3533, 2018.

- **Nastaran Mohammadian Rad**, Seyed Mostafa Kia, Calogero Zarbo, Twan van Laarhoven, Giuseppe Jurman, Paola Venuti, Elena Marchiori, and Cesare Furlanello. "Deep learning for automatic stereotypical motor movement detection using wearable sensors in autism spectrum disorders." Signal Processing, 144:180-191, 2018.

- **Nastaran Mohammadian Rad**, Seyed Mostafa Kia, Calogero Zarbo, Giuseppe Jurman, Paola Venuti, and Cesare Furlanello. "Stereotypical motor movement detection in dynamic feature space." In Data Mining Workshops (ICDMW), 2016 IEEE 16th International Conference on, pages 487-494. IEEE, 2016.

- **Nastaran Mohammadian Rad** and Cesare Furlanello. "Applying deep learning to stereotypical motor movement detection in autism spectrum disorders." In Data Mining Workshops (ICDMW), 2016 IEEE 16th International Conference on, pages 1235-1242. IEEE, 2016.

- **Nastaran Mohammadian Rad**, Andrea Bizzego, Seyed Mostafa Kia, Giuseppe Jurman, Paola Venuti, and Cesare Furlanello. "Convolutional neural network for stereotypical motor movement detection in autism." arXiv preprint arXiv:1511.01865, 2015.

Conclusioni

Lo sviluppo di un'applicazione mobile in tempo reale per il rilevamento di movimenti anomali nei pazienti con PD e ASD pu essere considerato come un obiettivo finale nel contesto del rilevamento automatico del movimento anomalo tramite sensori indossabili. Sviluppare un'applicazione di questo tipo che pu essere adattata in modo incrementale ai dati dei nuovi pazienti o allo stesso paziente considerato che la cinematica del paziente cambia nelle impostazioni della vita reale, potrebbe sia promuovere studi di ricerca e fornire nuovi strumenti di intervento per aiutare ricercatori, medici e operatori sanitari a monitorare continuamente, comprendere meglio e far fronte a comportamenti anormali nei pazienti con ASD e PD. Il raggiungimento di questo obiettivo tuttavia impegnativo a causa dell'elevata variabilit - di tipo inter e intra-soggetto - dei dati IMU acquisiti e della mancanza di informazione semantica sul tipo di anomalia dei dati raccolti nelle impostazioni reali. L'obiettivo principale della ricerca presentata in questa tesi quello di affrontare queste importanti problematiche nel rilevamento automatico del movimento anormale.

In questa tesi, il problema della variabilit inter e intra-soggetto viene affrontato utilizzando funzionalit di apprendimento funzionale e di trasferimento delle reti neurali conosciute come 'deep neural networks'. In particolare, viene proposta una nuova applicazione delle reti neurali con architettura nota come 'convolutional neural networks' (CNN) per apprendere una rappresentazione intermedia dei segnali IMU in individui con ASD.

L'architettura proposta viene inoltre utilizzata per trasferire le conoscenze, apprese da un set di dati SMM a un altro set di dati tramite l'inizializzazione della rete in uno scenario longitudinale. I risultati sperimentali su un insieme di dati simulati e due reali indicano che i) sostituire le caratteristiche artigianali con l'apprendimento automatico delle funzioni utilizzando CNN produce caratteristiche piu' discriminanti e robuste; e ii) l'apprendimento del trasferimento dei parametri migliora l'adattabilit e, di conseguenza, migliora le prestazioni del sistema di rilevamento SMM quando applicato a set di dati in un'impostazione longitudinale. Sebbene la proposta di rete neurale profonda fornisca un accurato modello di rilevamento SMM, lo schema completamente supervisionato proposto per addestrare il modello di rilevamento SMM problematico nell'adattamento in tempo reale poich l'adattamento a nuovi dati dovrebbe essere eseguito solo in base ai dati in ingresso senza etichetta. Questa limitazione motiva le future ricerche sull'adattamento online del sistema in modo non supervisionato. L'apprendimento del trasferimento transduttivo con la possibilit di un addestramento incrementale pu eventualmente facilitare l'adattamento online di un rilevatore di movimento anormale automatico in scenari in tempo reale.

Per beneficiare maggiormente delle fluttuazioni temporali dei segnali IMU, viene proposta una combinazione di LSTM con un'architettura CNN per apprendere le caratteristiche dinamiche su una sequenza di segnali IMU. I risultati sperimentali dimostrano che i) utilizzando le informazioni temporali si migliora la separabilit dei campioni SMM e no-SMM; ii) mentre l'asimmetria dei campioni influisce negativamente sulle prestazioni del rivelatore SMM nello spazio delle caratteristiche statiche, lo sfruttamento dei pattern temporali dei segnali IMU multi-sensore recupera le sue prestazioni. Pertanto, l'apprendimento delle informazioni temporali dei segnali IMU facilita l'applicazione del rilevamento SMM in ambienti pi realistici in cui i

campioni in ingresso sono fortemente distorti. Inoltre, al fine di migliorare l'affidabilit e la stabilit del rivelatore SMM, viene proposto un algoritmo di apprendimento basato sui migliori studenti LSTM. I risultati sperimentali mostrano un leggero miglioramento delle prestazioni medie rispetto ai singoli classificatori LSTM. Inoltre, l'insieme di LSTM fornisce un rilevatore SMM pi affidabile rispetto a ogni singolo classificatore LSTM riducendo la variabilit dei risultati.

Il problema della mancanza di dati etichettati nelle applicazioni cliniche dei sensori indossabili viene affrontato ridefinendo il problema di rilevamento dei movimenti anomali in un paradigma di rilevamento della novit utilizzando il metodo di modellazione normativa profonda. A tal fine, i movimenti normali eterogenei nei pazienti con PD e ASD sono modellati utilizzando un autoencoder di denoising probabilistico. Quindi il modello normativo risultante viene utilizzato per rilevare FOG e SMM in pazienti con PD e ASD. I risultati sperimentali su tre set di dati di riferimento illustrano l'efficacia del metodo proposto, che sovraperforma il popolare one-class SVM e gli approcci di rilevamento della novit basati sulla ricostruzione. Inoltre il metodo di rilevamento delle novit completamente non supervisionato proposto ha raggiunto una performance relativamente vicina alle sue controparti supervisionate.

Le possibili direzioni future da questa tesi si possono riassumere come segue:

1. **Valutare i nostri metodi di apprendimento profondi su un ampio set di dati raccolti in condizioni realistiche**: in questa tesi, i metodi di apprendimento approfondito sono addestrati e valutati su diversi set di dati IMU benchmark raccolti da pazienti con ASD e PD in condizioni cliniche controllate. Per garantire la generalizzazione attraverso i set di dati del mondo reale, necessario valutare i metodi proposti sui grandi set di dati IMU raccolti da pazienti in

ambienti meno controllati.

2. **Un'implementazione in tempo reale del sistema di rilevamento dei movimenti anomali**: il metodo di modellazione normativa profonda proposto offre la possibilit di modellare i normali dati del movimento umano durante le attivit della vita quotidiana senza la necessit di avere accesso ai campioni etichettati. Inoltre, questo metodo ha la bassa complessit computazionale che lo rende un metodo appropriato da implementare nel rilevamento dei movimenti di rilevamento anormale in tempo reale nei pazienti neuro-evolutivi e neuro-degenerativi.

3. **Incorporando i dati fisiologici per prevedere i movimenti anormali**: Sebbene i sensori IMU siano adatti per il rilevamento anormale del movimento nei pazienti con ASD e PD, i dati raccolti utilizzando questi sensori non forniscono sempre informazioni adeguate per ottenere una valutazione affidabile dei movimenti anomali di impostazioni cliniche. Ad esempio, a volte difficile rilevare da soli i dati IMU se una lentezza del movimento dovuta alla bradicinesia o a causa di affaticamento, ansia o altri fattori ambientali. Poich i cambiamenti nei comportamenti fisiologici possono verificarsi durante e prima dell'inizio di movimenti anormali, quindi il monitoraggio e la rilevazione di questi cambiamenti possono aiutare ad avere prestazioni di rilevamento dei movimenti anormali pi elevate. L'incorporazione dei dati fisiologici acquisiti mediante elettrocardiografia (ECG) e sensori elettromiografici (EMG) nel sistema di rilevamento fornisce la possibilit di prevedere l'insorgenza di movimenti anormali piuttosto che rilevarli.

Conclusies

Het ontwikkelen van een real-time mobiele applicatie voor het detecteren van abnormale bewegingen bij PD- en ASS-patinten kan worden beschouwd als een ultiem doel in het kader van automatische abnormale bewegingsdetectie met behulp van draagbare sensoren. Het ontwikkelen van een dergelijke toepassing die stapsgewijs kan worden aangepast aan de gegevens van nieuwe patinten of aan dezelfde patint als zijn / haar kinematica in de praktijk veranderen, kan zowel onderzoeksstudies vooruit helpen als nieuwe interventietools bieden die onderzoekers, clinici en zorgverleners helpen. voortdurend monitoren, beter begrijpen en omgaan met abnormaal gedrag bij ASS- en PD-patinten. Het bereiken van dit doel is echter een uitdaging vanwege de grote inter- en intra-subject variabiliteit in de verkregen IMU-gegevens en die geen toegang hebben tot de labels van verzamelde gegevens in real-life settings. De belangrijkste focus van het onderzoek gepresenteerd in dit proefschrift was het aanpakken van deze specifieke uitdagingen in automatische abnormale bewegingsdetectie.

In dit proefschrift wordt het probleem van inter- en intra-subject variabiliteit aangetast door het gebruik van feature learning en transfer learning mogelijkheden van diepe neurale netwerken. In het bijzonder wordt een nieuwe toepassing van CNN voorgesteld om een middenrepresentatie van IMU-signalen bij personen met ASS te leren. De voorgestelde architectuur wordt verder gebruikt om de kennis over te dragen, geleerd van de ene SMM-dataset naar een andere dataset via de netwerkinitial-

isatie in een longitudinaal scenario. De experimentele resultaten op een gesimuleerde en twee echte datasets geven aan dat i) het vervangen van handgemaakte functies door automatische functieleren met behulp van CNN resulteert in meer onderscheidende en robuuste functies in scenario's voor verschillende onderwerps SMM-detectie; en ii) leren van parameteroverdracht verbetert de aanpasbaarheid en verbetert als gevolg daarvan de prestaties van het SMM-detectiesysteem wanneer toegepast op nieuw geziene gegevensverzameling in een longitudinale instelling. Hoewel het voorgestelde diepe neurale netwerk een nauwkeurig SMM-detectiemodel biedt, is het voorgestelde volledig gecontroleerde schema voor het trainen van het SMM-detectiemodel problematisch in de aanpassing in de werkelijke tijd aangezien de aanpassing aan nieuwe ongeziene gegevens alleen op basis van de niet-gelabelde gegevens moet worden uitgevoerd. Deze beperking motiveert toekomstige onderzoeken naar de online aanpassing van het systeem op een onbewaakte manier. Transcendent transfer learning met de mogelijkheid van incrementele training kan mogelijk de online aanpassing van een automatische abnormale bewegingsdetector in real-time scenario's vergemakkelijken.

Om meer te profiteren van de temporele fluctuaties van IMU-signalen, wordt een combinatie van LSTM's met een CNN-architectuur voorgesteld om de dynamische kenmerken van een reeks IMU-signalen te leren. De experimentele resultaten demonstreren i) het gebruik van de temporele informatie verbetert de scheidbaarheid van SMM- en niet-SMM-monsters; ii) terwijl de scheefheid van de monsters een negatieve invloed heeft op de prestaties van de SMM-detector in de ruimte met statische elementen, herstelt het gebruik van de temporele patronen van IMU-signalen met meerdere sensoren de prestaties. Het leren van temporele informatie van IMU-signalen vergemakkelijkt dus de toepassing van SMM-detectie in meer realistische omgevingen waarin de binnenkomende monsters sterk scheef

staan. Bovendien wordt, om de betrouwbaarheid en stabiliteit van de SMM-detector te verbeteren, een algoritme voor het leren van een ensemble op basis van de beste LSTM-leerders voorgesteld. De experimentele resultaten laten een lichte verbetering zien in de gemiddelde prestaties vergeleken met enkele LSTM classifiers. Bovendien biedt het ensemble van LSTM's een betrouwbaardere SMM-detector in vergelijking met elke afzonderlijke LSTM-classificatie door de variabiliteit van de resultaten te verminderen.

Het probleem van het ontbreken van gelabelde gegevens in klinische toepassingen van draagbare sensoren wordt aangepakt door het abnormale bewegingsdetectietekort opnieuw te definiren in een nieuwheidsdetectieparadigma met behulp van de diepe normatieve modelleringsmethode. Hiertoe worden heterogene normale bewegingen bij PD- en ASS-patinten gemodelleerd met behulp van een probabilistische denoising autoencoder. Vervolgens wordt het resulterende normatieve model gebruikt om FOG's en SMM's te detecteren bij respectievelijk PD- en ASS-patinten. De experimentele resultaten op drie benchmarkgegevensverzamelingen illustreren de effectiviteit van de voorgestelde methode, die beter presteert dan SVM van n klasse en de op reconstructie gebaseerde nieuwe detectiewerkwijzen. De voorgestelde volledig unsupervised nieuwheidsdetectiewerkwijze bereikte een vrij dichte prestaties aan zijn gecontroleerde tegenhangers.

Mogelijke toekomstige richting uit dit proefschrift kan als volgt worden samengevat:

1. **Onze diepe leermethoden evalueren op een grote dataset verzameld in de realistische omstandigheden**: In dit proefschrift worden de deep learning-methoden getraind en gevalueerd op verschillende benchmark-IMU-gegevensverzamelingen verzameld door ASS- en PD-patinten onder gecontroleerde klinische omstandigheden. Om de generalisatie tussen real-world datasets te verzekeren, is het

noodzakelijk om de voorgestelde methoden te evalueren op de grote IMU-datasets die zijn verzameld van patinten in ongecontroleerde en wilde, real-life omgevingen.

2. **Een real-time implementatie van een abnormaal bewegingsdetectiesysteem**: de voorgestelde methode voor diepe normatieve modellering biedt de mogelijkheid om de normale menselijke bewegingsgegevens te modelleren tijdens activiteiten in het dagelijks leven zonder de noodzaak van toegang tot gelabelde monsters. Bovendien heeft deze methode de lage rekenkundige complexiteit waardoor het een geschikte methode is om te worden gemplementeerd in de real-time abnormale detectie bewegingsdetectie bij neuro-ontwikkelings- en neurodegeneratieve patinten.

3. **Fysiologische gegevens opnemen om abnormale bewegingen te voorspellen**: Hoewel IMU-sensoren zeer geschikt zijn voor de abnormale bewegingsdetectie bij ASS- en PD-patinten, bieden de verzamelde gegevens met behulp van deze sensoren niet altijd voldoende informatie om een betrouwbare beoordeling te krijgen van de abnormale bewegingen van medische instellingen. Het is bijvoorbeeld soms moeilijk om alleen uit de IMU-gegevens te detecteren of een traagheid van beweging te wijten is aan bradykinesie of als een resultaat van vermoeidheid, angst of andere omgevingsfactoren. Aangezien veranderingen in fysiologisch gedrag kunnen optreden tijdens en vr het begin van abnormale bewegingen, kan het monitoren en detecteren van deze veranderingen bijdragen tot een hogere abnormale bewegingsdetectiewerking. Het opnemen van fysiologische gegevens vastgelegd door elektrocardiografie (ECG) en elektromyografische (EMG) sensoren in het detectiesysteem biedt de mogelijkheid om het begin van abnormale bewegingen te voorspellen in plaats van ze alleen te de-

tecteren.