Detecting Epileptic Seizures With Multimodal Non-EEG Data From Wearables

DISSERTATION

zur Erlangung des Grades eines
Doktors der Ingenieurwissenschaften

eingereicht bei der
Naturwissenschaftlich-Technischen Fakultät der Universität Siegen

von
Sebastian Böttcher

Siegen 2023

Detecting Epileptic Seizures With Multimodal Non-EEG Data From Wearables

DISSERTATION

for the degree of Doctor of Engineering Sciences

submitted to the School of Science and Technology
at the University of Siegen

by
Sebastian Böttcher

Siegen 2023

Created with LaTeX on 2023-07-05 22:12:23+02:00

This document is not digitally
signed and may have been altered

Betreuer und erster Gutachter
Supervisor and first reviewer
Prof. Dr. Kristof Van Laerhoven
Universität Siegen

Zweiter Gutachter
Second reviewer
Prof. Dr. Björn Eskofier
Friedrich-Alexander-Universität Erlangen-Nürnberg

Tag der mündlichen Prüfung
Day of the defense
19 May 2023

Zusammenfassung

Epilepsie ist eine der am weitesten verbreiteten chronischen neurologischen Erkrankungen, von der weltweit Millionen von Menschen aus allen Gesellschaftsschichten betroffen sind. Epilepsie äußert sich bei den Betroffenen durch wiederkehrende Anfälle mit einer Reihe unterschiedlicher Symptome, in wechselnden Abständen und mit variabler Ausprägung. Der Goldstandard zur Diagnose und Beobachtung von Epilepsie ist die Video-Elektroenzephalografie. Dabei werden Epilepsiepatienten einige Tage lang überwacht während Ärzte mit verschiedenen Mitteln Anfälle auslösen, in der Hoffnung, genügend Informationen für eine präzise Diagnose und Behandlung zu erhalten. Dieses Verfahren ist jedoch im täglichen Leben der Patienten und über längere Zeiträume hinweg nicht sinnvoll. Darüber hinaus haben sich die handschriftlichen Tagebücher, die manche Patienten führen, als unzuverlässig erwiesen, da sie die Zahl der Anfälle in der Regel stark unterbewerten. Es wird also eine Alternative für die Ultra-Langzeit-Überwachung benötigt, um die gängigen Behandlungen zu verbessern und die Entwicklung neuer Therapieoptionen zu erleichtern. Diese Dissertation untersucht das Potenzial multimodaler, nicht-elektroenzephalografischer, von tragbaren Geräten aufgezeichneter Daten zur Anfallserkennung für automatische Anfallstagebücher. Darüber hinaus wird eine mögliche Anwendung der Anfallserkennung als Teil eines automatischen Alarmsystems untersucht.

Im Rahmen dieser Arbeit wurde ein neuer Datensatz von Biosignaldaten erstellt und verwendet, aufgezeichnet in zwei europäischen Epilepsiezentren im Rahmen eines europäischen Forschungsprojekts. Mehr als 200 Epilepsiepatienten wurden in den beiden Einrichtungen rekrutiert, und mehr als 300 epileptische Anfälle unterschiedlicher Art wurden mit einem tragbaren Gerät aufgezeichnet. Dabei wurde das Empatica E4 verwendet, ein für die Forschung geeignetes, am Handgelenk getragenes Wearable das Biosignale der Bewegung, der elektrischen Hautleitfähigkeit und des Blutvolumenpulses erfasst. Diese Daten bildeten für mehrere Studien, die sich mit der Bewertung von Methoden zur Erkennung von epileptischen Anfällen befassten, die Grundlage zur Datenanalyse.

Die vorliegende Arbeit fasst mehrere Beiträge des Autors zur Erkennung epileptischer motorischer Anfälle mit multimodalen nicht-elektroenzephalografischen Daten von Wearables zusammen. Die enthaltenen Studien untersuchten insbesondere solche Anfälle mit Bewegungsmanifestationen in den Gliedmaßen und es wurde festgestellt, dass bestimmte Erkennungssysteme die auf maschinellem Lernen unter Verwendung physiologischer Biosignaldaten basieren gut geeignet sind.

Ein zentraler Teil dieser Arbeit befasst sich mit krampfartigen tonisch-klonischen Anfällen, also schweren und gefährlichen Anfällen die in beiden Gehirnhälften beginnen oder sich dorthin ausbreiten. Während dieser Anfälle ist das Bewusstsein des betroffenen Patienten beeinträchtigt und es kommt zu hochamplitudigen, hochfrequenten Zuckungen der Gliedmaßen und am gesamten Körper. In einer der hier vorgestellten Studien wurde eine automatische Erkennungsmethode auf Grundlage einer Kombination aus Beschleunigungsmessung und elektrodermalen Aktivitätssignalen bewertet. Ein maschinelles Lernmodell wurde auf Experten-markierten Daten trainiert und anhand einer dem Modell unbekannten Testgruppe bewertet. Die Leistung des Modells ist mindestens gleichwertig mit dem Stand der Technik, bei einer korrekten Erkennung von über 90 Prozent der Anfallsereignisse und einer Fehlalarmrate von weniger als 0,5 Prozent pro Tag. Die vorgeschlagene Methodik schneidet außerdem besser ab als das durchschnittliche monomodale Erkennungssystem aus vergleichbaren Studien. Auf konvulsive tonisch-klonische Anfälle folgt in der Regel eine Phase der Bewegungs- und Bewusstlosigkeit, wodurch das Risiko eines plötzlichen unerwarteten Todes bei Epilepsie erheblich erhöht ist. Eine weitere Studie untersucht die Brauchbarkeit von am Körper gemessenen Biosignaldaten, um diesen Zeitraum der Bewusstlosigkeit mithilfe einer heuristischen Erkennung basierend auf Beschleunigungssignalen zu erkennen und einzuschätzen. In Abhängigkeit von einer vorherigen automatischen Erkennung des Anfalls war die Methode in der Lage, alle Fälle von Unbeweglichkeit im Datensatz korrekt zu klassifizieren.

Ein weiterer wesentlicher Teil dieser Arbeit befasst sich mit der Erkennung fokaler Anfälle anhand von Daten aus Wearables. Fokale Anfälle haben, über verschiedene Patienten hinweg betrachtet, typischerweise sehr heterogene Symptome. Diese umfassen verschiedene Arten von Körperbewegungen, Reaktionen des autonomen Nervensystems und psychologische Veränderungen. In der hier einbezogenen Forschung wurden nur solche fokalen Anfälle mit spezifischen Bewegungen der Gliedmaßen analysiert. Eine erste explorative Studie untersuchte die Auswirkungen der hohen Varianz fokaler motorischer Anfälle auf relevante Biosignale und die Anfallserkennung mit diesen. In einer weiteren Studie wurden dann individualisierte und generische Modelle zur Erkennung fokaler motorischer Anfälle auf der Grundlage der vom Wearable aufgezeichneten Biosignale untersucht. Die Studie ergab, dass die optisch gemessenen Blutvolumenpulsdaten stark durch Bewegungsartefakte beeinträchtigt sind. Darüber hinaus schnitten generische Modelle deutlich schlechter ab als patientenspezifische Modelle und wiesen hohe Fehlalarmraten auf. Daher sind für die Erkennung fokaler Anfälle maßgeschneiderte Erkennungsmodelle für einzelne Patienten, und zwar speziell für eine Untergruppe von Epilepsiepatienten bei denen charakteristische Anfälle auftreten, wahrscheinlich die robusteste Methode.

Diese Arbeit kommt zu dem Schluss, dass generische Anfallserkennungsmodelle zwar für hoch-konvulsive Anfälle und unter Krankenhausbedingungen ausreichend sein können, dass sie aber derzeit nicht für die Erkennung fokaler Anfälle mit weniger oder gar keinen Bewegungen geeignet sind. Umgekehrt sind patientenspezifische Erkennungsmethoden für nicht-konvulsive motorische Anfälle vielversprechend. Erkennungsmodelle, die sich im Laufe der Zeit individualisieren, könnten letztendlich die beste Option für ambulante Anfallserkennung werden. In den hier einbezogenen Studien wurden insbesondere Erkennungssysteme untersucht, die auf maschinellem Lernen unter Verwendung physiologischer Biosignaldaten basieren. Die Ergebnisse zeigten, dass diese Systeme für die Erkennung von konvulsiven Anfällen und solchen mit leichteren Bewegungen der Gliedmaßen geeignet sind.

Abstract

Epilepsy is one of the most prevalent chronic neurological disorders, affecting millions worldwide throughout all societal groups. Epilepsy manifests in those affected as reoccurring seizures with a wide range of different symptoms at variable intervals and severity. The current gold standard to diagnose and monitor epilepsy is video-electroencephalography. Patients with epilepsy visit monitoring units for a few days, and clinicians provoke seizures through various means, hoping to get enough information for precise diagnosis and treatment. However, this procedure is not viable during the patients' daily lives and over more extended periods. Furthermore, the handwritten diaries that some patients keep have proven unreliable, typically severely under-counting the number of seizures occurring. An alternative for ultra-long-term monitoring is needed to improve current treatments and facilitate the development of new therapy options. This thesis investigates the potential of multimodal non-electroencephalography data recorded from wearable devices as a tool for seizure detection in the context of automated seizure diaries. It furthermore explores a potential application of seizure detection in the context of an automatic alarm system.

The work featured in this thesis produces and employs a new data set of wearable biosignal data, recorded at two European epilepsy centers in the context of a European collaborative research project. Over 200 patients with epilepsy were recruited at the two epilepsy monitoring units, and over 300 epileptic seizures of varying types were recorded with a wearable device. Here, the Empatica E4 is used, a research-grade wrist-worn wearable that captures the biosignal modalities of accelerometry (movement), electrodermal activity (electrical skin conductance), and blood volume pulse (optical pulse measurement via photoplethysmography). This data set was the basis for several data analysis studies concerning the evaluation of seizure detection methodologies.

This thesis compiles and provides a framework for several contributions of the author concerning the detection of epileptic motor seizures with multimodal non-electroencephalography data from wearables. Specifically, the included studies investigated those seizures with movement manifestations in the limbs and found detection systems based on supervised ensemble machine learning using physiological biosignal data to be viable.

One central part of this thesis is focused on convulsive tonic-clonic seizures, severe and dangerous seizures that start in or progress to both hemispheres of the brain. During these seizures, the awareness and consciousness of the affected patient are impaired, and high-amplitude, high-frequency jerks of the limbs and whole body occur. One of the studies presented

here assessed an automatic detection methodology based on a combination of accelerometry and electrodermal activity signals. A supervised ensemble machine learning model is trained on expert-labeled data and evaluated on an out-of-sample test set. It performs at least on par with state-of-the-art related work, correctly classifying more than 90 percent of seizure events with false alarm rates of less than 0.5 per day. The suggested methodology performs better than the average monomodal detection system in related work. Convulsive tonic-clonic seizures are typically followed by a period of unconsciousness and immobility, significantly increasing the risk of sudden unexpected death in epilepsy. A further study investigates the utility of wearable biosignal data to detect and gauge this period based on a heuristic detection using accelerometry signals. Contingent on a prior automatic detection of the seizure, the methodology was able to classify all instances of immobility in the data set correctly.

Another essential segment of this thesis highlights the detection of focal seizures with data from wearables. Focal seizures typically have very heterogeneous symptoms when regarded across patients. They include body movements of different kinds, responses of the autonomic nervous system, and psychological indications. The research included here analyzed only those focal seizures with specific movements of the limbs. An early exploratory study investigated the impact of the high variance of focal motor seizures on biosignals and the performance of seizure detection based on those signals. An additional study then considered individualized and generic models for detecting focal motor seizures based on the biosignals recorded by the wearable. The study found the optically measured blood volume pulse data to be highly impacted by noise from motion artifacts. Furthermore, generic models performed considerably worse than those specific to an individual patient, with high false alarm rates. Thus, for focal seizure detection, custom-made detection models for individual patients are likely to be the most robust methodology, and are specifically suitable for a subset of patients with epilepsy who experience characteristic seizures.

This thesis concludes that while generic seizure detection models may be sufficient for highly convulsive seizures and under in-hospital conditions, they are currently not feasible for detecting focal seizures with fewer or no movements. Conversely, patient-specific detection methodologies are promising for non-convulsive motor seizures. Detection models that individualize over time may eventually become the best option for ultra-long-term seizure detection. Specifically, the included studies investigated detection systems based on supervised ensemble machine learning using physiological biosignal data. Results showed them to be feasible for detecting convulsive and less-convulsive seizures with manifestations including movements of the limbs.

Contents

CHAPTER 1

Introduction

Epilepsy is one of the most prevalent chronic neurological disorders, with a worldwide incidence of up to 100 per 100,000 people per year, affecting over 70 million people. Several different types of epilepsy have been identified, manifesting in different symptoms, characteristically in seizures of varying severity. Seizures are typically involuntary movements of the whole body or parts of it, sometimes accompanied by a loss of consciousness. The current gold standard in epilepsy diagnosis and seizure monitoring is in-hospital video-electroencephalography. Patients with epilepsy are often diagnosed in epilepsy monitoring units over the course of several days. However, while being accurate and widely used for the diagnosis of epilepsy and the determination of treatment, this methodology is only feasible in relatively short-term applications. Thus, seizure monitoring at home in an ultra-long-term context, that is, over a period of multiple days or longer, is needed. Patients with epilepsy typically keep handwritten diaries of their seizure events. However, these diaries have proven unreliable and incomplete, especially when a loss of consciousness accompanies epileptic seizures. Wearable devices such as smartwatches and fitness trackers, which are readily available to a broad audience, may fill this gap in monitoring. They may automatically detect and log seizures of epilepsy patients in their day-to-day environment and serve as an automatic alarm to potentially alert caregivers in the event of a seizure.

This thesis aims to explore the feasibility of epileptic seizure detection using wearable devices and biosignal sensors in the context of automated diaries. Furthermore, it investigates a potential application of seizure detection in the context of an automatic alarm system. During a six-year collaborative research project, an exhaustive set of biosignal data was collected from patients with epilepsy. Further, several scientific research studies were conducted and published. These studies primarily focused on evaluating seizure detection methodologies on the collected data, including seizures with movement symptoms of varying severity. For this purpose, multimodal non-electroencephalography biosignal data were collected with a wrist-worn research-grade wearable device. The following chapter introduces the basics of epilepsy, wearable data recording, and seizure detection and gives this thesis's motivation and essential contributions.

1.1 Fundamentals of Epilepsy

> « The history of epilepsy can be summarised as 4000 years of ignorance, superstition, and stigma followed by 100 years of knowledge, superstition, and stigma. »
> — Rajendra Kale, 1997 [1]

Epilepsy and its symptoms have been part of medical texts since the beginning of recorded history [2, 3]. Some of the earliest accounts of apparent epileptic seizures date as far back as ca. 2000 B.C. [4], while epilepsy as a human condition is first described in some early Babylonian medical records, ca. 1000 B.C. [3, 4]. The condition was generally regarded as a divine, spiritual affliction until its first description as a medical disease in *On the Sacred Disease*, attributed to Hippocrates, ca. 400 B.C. [4]. Nevertheless, misunderstanding and stigmatization of the epileptic condition continued throughout the Middle Ages, and only modern scientific medicine started to acknowledge its neurological origins. The first anticonvulsive medications used in epilepsy were Bromide in 1857 and Phenobarbital in 1912 [5]. Since then, many different anti-epileptic drugs and treatments have been developed and improved [6]. However, even today, epilepsy remains a prevalent, albeit often treatable, neurological disorder, and patients with epilepsy (PWEs) are still stigmatized and discriminated against [7].

1.1.1 What is Epilepsy?

Epilepsy is clinically defined as a disease generally identified if two unprovoked seizures occur more than one day apart or if there is a reasonably heightened risk of subsequent seizures after an initial one [8]. The origin of an individual's epilepsy can be genetic, acquired, or unknown, and in many cases, is caused by a combination of these factors [9]. Various conditions may cause acquired epilepsy, including strokes, tumors, brain trauma, and nervous system infections [10].

The neurological mechanism of an epileptic seizure is described as the abnormal and uncontrolled firing of neurons in the brain, disrupting the regular and organized brain activity. The source of this anomalous activity, that is, the seizure, is called the *focus*, and the *onset* and *offset* are the start and end times of the seizure. The focus of the seizure can be limited to one specific location of the brain, or span large areas of both hemispheres of the brain, describing *focal seizures* or *generalized seizures*, respectively (see also Section 1.1.2). The whole period of the seizure between onset and offset is called the *ictal*[1] period, with a typical duration of a few seconds to many minutes. The onset and offset of the seizure are thereby determined by clinicians reviewing the electroencephalography (EEG) signal. The periods immediately preceding and following the seizure are denoted *pre-ictal* and *post-ictal*, respectively. The pre-ictal phase can include symptoms such as behavioral or cognitive changes [11] and changes in heart rate (HR) [12]. In the post-ictal phase clinical symptoms of the seizure itself, such as cognitive impairments, often remain beyond the EEG-offset of the seizure.

The time from one post-ictal to the next pre-ictal phase is called the *inter-ictal* period. Pre- and post-ictal phases are not necessarily defined in length and can vary greatly, depending on the relevant context. Specifically, while the end of the pre-ictal phase and the beginning of the post-ictal phase are usually well-defined (as the beginning and end of the ictal phase), the transitions between inter-ictal and pre- or post-ictal phases are often ambiguous. For example,

[1] From Latin *ictus*: blow, stroke

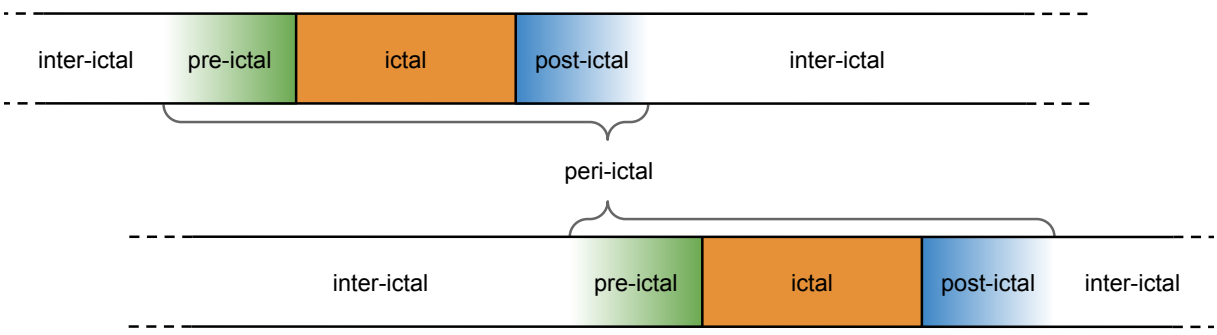


Figure 1.1: Schematic overview of the typical seizure phases, showing two distinct seizures (ictal), the periods before and after (pre- and post-ictal), and the periods in between seizures (inter-ictal).

after some seizures neither the clinicians nor the patients themselves can accurately pinpoint the end of all related symptoms. Nevertheless, the whole phase from the start of the pre-ictal to the end of the post-ictal period is sometimes called *peri-ictal* [13]. Figure 1.1 gives a simplified overview of the different seizure phases and their temporal relation to each other.

With a worldwide incidence of up to 100 new cases per 100,000 people per year and affecting over 70 million people worldwide, epilepsy is one of the most prevalent neurological disorders [14, 15]. At any given time, an estimated 4 to 10 per 1,000 people have epilepsy, that is, are at risk of continuing seizures and needing treatment [16]. Epilepsy is more common in men than in women, in children than in adults, and in low- and middle-income countries than in high-income countries [17]. Furthermore, concerning the three differentiated epilepsy origins, the estimated prevalence of epilepsy with unknown origin is highest, followed by acquired epilepsy and epilepsy due to genetic disposition [15, 17].

While the overall risk of death from epilepsy is generally considered low, *sudden unexpected death in epilepsy (SUDEP)* is still a serious and considerable outcome of epileptic seizures, especially of those that are severely convulsive [15]. SUDEP incidence varies in range by age and gender, with estimates at around 1.2 per 1000 PWEs per year [18, 19]. Furthermore, increased mortality rates are associated with risk factors such as alcohol misuse, missed anticonvulsant drug prescriptions, injuries, or depression [20]. Overall, compared to the general populace, the risk of premature death is up to three times higher for PWEs [16].

1.1.2 Epileptic Seizures

An epileptic seizure is an occurrence of symptoms due to abnormal neuronal activity in the brain [21]. Epileptic seizures can manifest in various symptoms [22], which can change over time in individual patients, especially children [23]. The expression of an epileptic seizure as a group of symptoms is also called *semiology*[1]. As the management and treatment of epilepsy can vary according to which type of symptoms occur, and some non-epileptic seizures exist, a classification of seizures into a specific set of categories can be of utmost importance. Over time there has been a miscellany of different seizure type classifications [24–26], with the latest being the 2017 *Operational classification of seizure types by the International League Against Epilepsy (ILAE)* [27]. However, terminology from older classifications is sometimes still in use.

[1]From Ancient Greek *sēmeîon*: sign; *lógos*: explanation

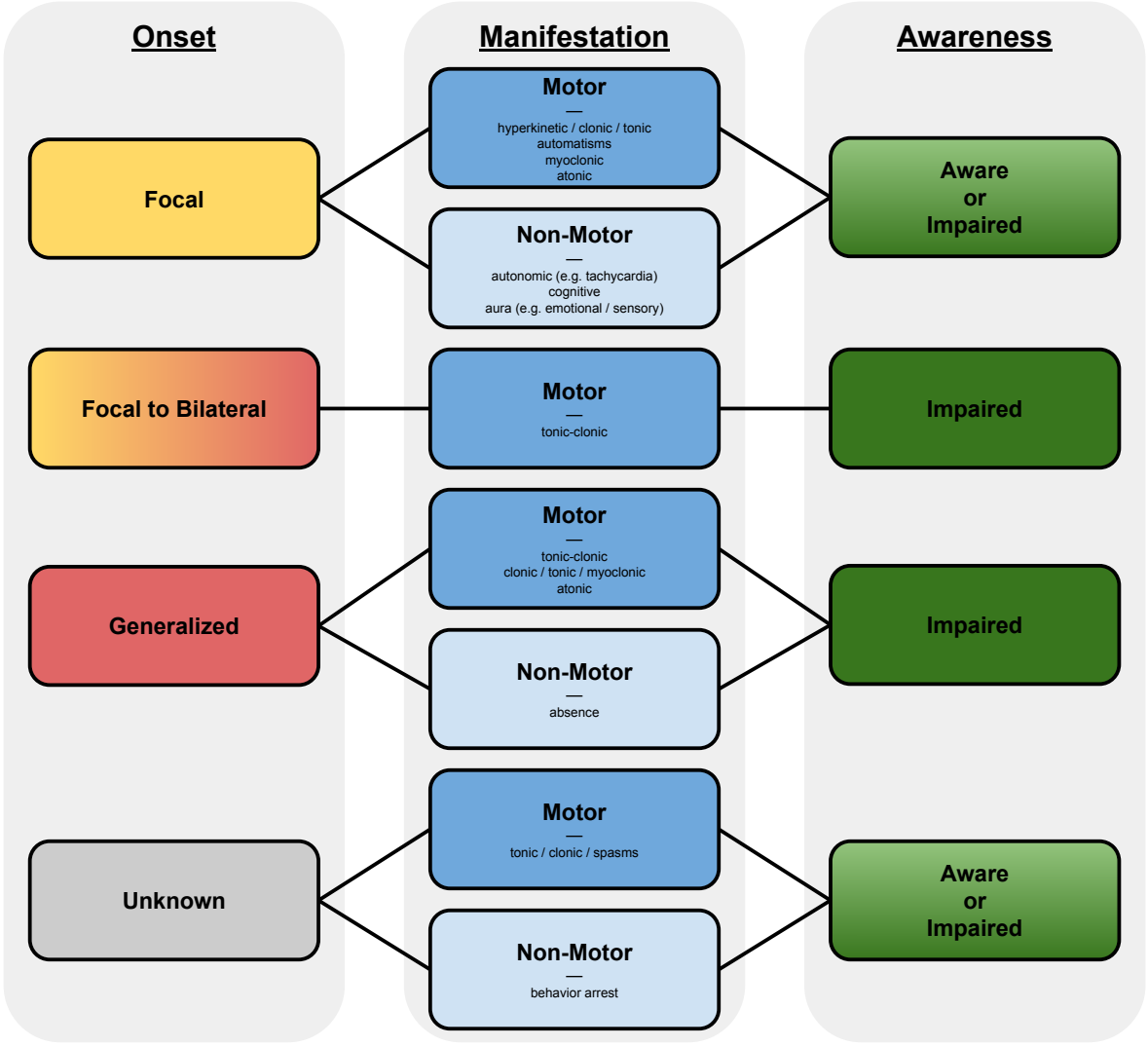


Figure 1.2: Schematic overview of different seizure types by onset, manifestation, and awareness, according to the *ILAE 2017 Classification of Seizure Types Expanded Version.* See also Fisher et al. [27], Figure 2.

Unless otherwise stated or referencing older publications, this thesis will use the 2017 ILAE seizure type classification, illustrated in Figure 1.2.

The most widely and colloquially known seizure type is that of the *generalized tonic-clonic seizure (GTCS)* [28]. As a generalized seizure, it onsets throughout the brain and progresses from the *tonic*[1] phase, a brief tensing of skeletal muscles in the limbs and throughout the body, to the *clonic*[2] phase with prolonged, violent, and rapid convulsions of the muscles [22]. GTCSs typically last no longer than 1 to 2 minutes, are accompanied by *tachycardia*[3] and *impaired awareness*, and are often followed by a period of unconsciousness [22, 29]. Directly related to GTCSs are *focal to bilateral tonic-clonic seizures (FBTCSs)*, which have a focal onset and typically progress to generalization during the tonic phase but are often indistinguishable

[1]From Ancient Greek *tónos*: tension
[2]From Ancient Greek *klónos*: confused motion, trembling
[3]Heart rate exceeding the normal resting baseline

from GTCSs [30]. As such, GTCSs and FBTCSs are major risk factors concerning SUDEP [19]. While there are other types of generalized seizures like generalized absences, they are much less frequent in the data sets included for this thesis and are not further detailed here.

Contrary to GTCSs and FBTCSs, *focal seizures (FSs)* have a localized onset in the brain and do not progress over both hemispheres. FSs are less well-known to the public than GTCSs, and their often less severe manifestations are also not as prevalent in the general media coverage concerning epilepsy [28]. Nevertheless, they make up a considerable portion of seizures [15]. They are much more varied in their manifestations than GTCSs, as symptoms can depend on the localization of the focus in the brain [31]. Seizure classifications generally subdivide FSs by the awareness status of the patient and the occurrence of motor symptoms during the seizure [27]. Older seizure classifications designated focal seizures without impairment of awareness as simple partial seizures, and those where the consciousness is impacted as complex partial seizures [25]. However, those notations are not used anymore in current literature in favor of more verbose descriptions of the seizures [32]. Any combination of awareness, motor, and non-motor manifestation can occur during the same seizure.

Among possible variants of FSs, those with tonic or clonic motor features, or both, are closest to GTCSs concerning manifestation and thus also wearable monitoring. While the patterns of movement during the ictal phase mimic those during a GTCS, these seizures are much less violently convulsive than their generalized counterparts. Moreover, they are usually not followed by a period of unconsciousness. Overall, possible motor manifestations in FSs can include the following features, roughly ordered by perceived severity of movement:

Hyperkinetic: Intense, irregular, asymmetric, projecting, and violent movements of one or more limbs. Vocalization and fear are common non-motor symptoms accompanying hypermotor seizures [33].

Clonic: Regular and repeated alternation between tensing and relaxing muscles in one or more limbs. In focal clonic seizures, the patient usually maintains consciousness at seizure onset [22].

Tonic: Tensing of the muscles of one or more limbs. Tonic features are often a prelude to more intense clonic movements [22].

Automatism: Regular or complex movements of a specific body part, which can include the limbs or parts of the limbs, the head, the mouth, or the eyelids [22, 34].

Myoclonic: Very short single twitch of a body part, often repeated several times in quick succession within just a few seconds. They are most frequently associated with juvenile myoclonic epilepsy, a genetic syndrome emerging in adolescents [35].

Atonic: Complete, involuntary relaxation of muscles in a part of or the whole body. The sudden loss of muscle control in atonic seizures substantially increases the risk of injury [36].

A similar array of possible symptoms exists regarding non-motor manifestations of FSs. The most common autonomic manifestation of epileptic seizures is tachycardia [29], which is often defined as an increase in HR above a certain absolute threshold or relative to a baseline. Various research has linked ictal and pre-ictal tachycardia to focal and generalized seizures [12,

29, 37], often coinciding with motor symptoms [38, 39]. Overall, three main classes of non-motor manifestations exist:

Autonomic: Changes in the autonomic nervous system. Cardiac modulations like tachycardia and bradycardia are among the most investigated [40, 41], and others like electrodermal activity changes are also prevalent [42, 43].

Cognitive: Impaired awareness or other cognitive dysfunctions of the patient. These symptoms may not be tied to a single seizure but could be long-lasting or permanent impairments of the patient's cognitive abilities [44, 45].

Aura: The patient experiences certain feelings, not necessarily directly measurable by biosignals. Among them are sensory auras (e.g., epigastric, olfactory, gustatory, auditory), emotional auras (e.g., pleasure, excitement, fear, anxiety), or psychic auras (e.g., disordered thoughts, déjà vu, hallucinations) [46, 47].

Short-term impaired awareness is a common symptom of any type of seizure, independent of other motor or non-motor manifestations. It can display as short-term memory loss, loss of speech, or other impairment of cognitive abilities and understanding of simple concepts [48]. These symptoms may also last for a short while after electrographic seizure offset [49]. Many seizures are left unclassified, especially in outpatient settings where the seizure was not witnessed by another person or recorded by video. Moreover, even under monitored conditions, some seizures cannot be associated with a defined seizure type category [27].

1.1.3 Diagnosis, Monitoring, Management, and Treatment

Diagnosis and Monitoring

Epilepsy can be diagnosed in a clinical ambulatory setting by anamnesis, whereby a patient visits a neurologist and describes potential past seizures, genetic disposition, or other relevant circumstances that may lead to a diagnosis of epilepsy. Another common way to diagnose epilepsy and epileptic seizures, sometimes reinforcing findings concluded from anamnesis, is through differential diagnosis in an epilepsy monitoring unit (EMU), usually part of a hospital or medical center for neurological healthcare. The clinical gold standard for epilepsy monitoring in an EMU is video-electroencephalography (vEEG). The stay in an EMU typically ranges between two days and two weeks in duration, depending on the purpose of the visit and individual needs, although research has shown that the diagnostic yield in EMUs drops significantly after five days [50]. In a vEEG monitoring unit, clinical technicians attach a standard array of EEG electrodes to the patient's scalp, most often in a 10-20 pattern [51, 52], and a video camera is pointed towards the patient's hospital bed, continuously recording during the monitoring phase. Depending on local practices and individual needs, the patients may be allowed to leave the hospital room for a short duration, removing themselves from the vEEG system. The EEG and video data are monitored around the clock by technicians or clinicians, and the patients have access to an alarm bell that can notify staff of an impending, ongoing, or passed seizure.

Nevertheless, epileptic seizures themselves are very infrequent and take up relatively little time of an epilepsy patient's lifetime. For example, one estimate from an ultra-long-term study puts the time spent having a seizure at only $0.05\,\%$ [53, 54]. Long-term monitoring of epilepsy

and (semi-)automatic detection of epileptic activity with EEG is possible, however, it often requires the patient to wear high-density EEG electrode arrays [55, 56], or have electrodes implanted [57, 58]. An additional practice of monitoring seizures, especially in an outpatient context, is keeping a seizure diary of manually recorded occurrences of seizures in the patient's daily life. Systematic and long-term documentation of seizures is essential for therapeutic decision-making and clinical and pharmacological studies [59, 60]. As the primary aim of epilepsy treatment is for the patient to become seizure-free, correctly gauging the number of seizure occurrences is paramount to assessing the efficiency of a given treatment.

Management and Treatment

In modern medicine, there are several options to manage and treat epilepsy. Anti-epileptic drugs (AEDs) can potentially suppress seizures in two out of three PWEs, but they rarely change the long-term prognosis, which is most affected by epilepsy surgery [14]. In this process, vEEG monitoring units are essential for effectively treating PWEs. Differential diagnosis and pre-surgical evaluation are the most common reasons to visit an EMU. Patients selected for a surgical intervention based on their diagnosis undergo pre-surgical evaluation. To better understand their seizures in general and gain knowledge about the localization of seizure onsets to discover potential targets for resection, clinicians provoke seizures in a controlled environment through different means, such as reduction of AEDs, sleep deprivation, or stimulation with light pulses [61]. During seizures, the patients are cared for by specialized staff who monitor the seizure manifestation and progression and test for symptoms like muscle tensing or cognitive impairments. This information may help to identify the seizure type and onset localization.

Resective epilepsy surgery is only relevant for patients who are AED-refractory, that is, who are not responding to AEDs, and for whom clinicians were able to define a focal seizure onset zone. The aim of the surgery is to potentially reduce or eliminate the recurrence of seizures [62]. During surgery, brain tissue is removed at the location of typical seizure onset while carefully weighing a positive outcome for the patient against minimizing the risk of further neurological damage. On average, between $25\,\%$ and $50\,\%$ of patients who underwent epilepsy surgery become free of seizures without needing AEDs, but this result can be delayed up to 5 years after surgery [62, 63].

A different method of managing and treating epilepsy and epileptic seizures is neurostimulation [64]. Clinicians usually administer this therapy to patients who are AED-refractory and for whom surgery is not an option. Generally, there are three types of neurostimulation frequently employed in epilepsy patients.

Vagus nerve stimulation (VNS) is the oldest type of stimulation and uses an extra-cranial methodology to stimulate the vagus nerve on a pre-programmed regular schedule. While VNS can significantly reduce seizure frequency in up to half of the treated population, it has common side effects like hoarseness and speech modulation, especially during stimulation phases [64, 65].

Conversely, deep brain stimulation (DBS) is an intracranial stimulation technique that targets a specific brain region to apply regular pulses of electricity, suppressing the propagation or even onset of epileptic seizures. DBS is a promising intervention method for drug-resistant epilepsy patients, but further research and studies must investigate if it significantly improves outcomes [64, 66, 67].

Responsive neurostimulation (RNS) does not stimulate on a set schedule but uses seizure detection methods to decide when to send electrical pulses to the brain [68]. As such, efficient and precise implementations of both detection and stimulation are necessary for RNS therapy to be successful, but the benefit of significantly reduced periods of stimulation compared to other regularly stimulating systems advocates for this neurostimulation method [69].

1.2 Fundamentals of Wearable Biosignal Recording

> « Miniaturization of components has enabled systems that are wearable and nearly invisible, so that individuals can move about and interact freely, supported by their personal information domain. »
>
> — Steve Mann, 1997 [70]

A wearable device, or *wearable*, is a self-sufficient electronic device that can be worn on the human body in some form. In this thesis wearables are battery-powered devices which can record at least one sensor modality that represents one of the wearer's biosignals. While wearable technology in a broader sense has existed for some hundred years, wearable electronic computing devices only emerged in the 1970s and 1980s [71]. Since then, wearables have made giant leaps in the reduction of size and the increase of functionality. While the first systems included backpacks containing sizeable electrical equipment, the invention and rise of mobile phones and smartphones have considerably boosted the development of small, consumer-grade wearables like smartwatches. Nowadays, wearables can record many different kinds of biosignals from the human body by way of a varied array of different kinds of recording principles and sensors. Signals may, for example, represent brain, muscle, cardiac, electrodermal, or thermal activity, as well as the motion of different body parts (Figure 1.3).

Electroencephalography (EEG) captures changes in the brain's electrical activity by electrodes placed on the scalp. Likewise, surface-electromyography (EMG) records muscle activity via electrodes placed above the muscle to capture the electric potentials induced by its activity. A common technique to record cardiac activity is electrocardiography (ECG) which, like EEG and EMG, also requires electrodes to be placed on the skin. Although some notable exceptions exist [72–74], the need for electrodes to measure these electrical signals is an important factor in why wearable systems rarely measure them, especially in a long-term context. Electrodes are challenging to apply correctly by oneself, they can be very obtrusive and stigmatizing, and typically only last for a few days until they need to be reapplied, producing additional cost during long-term and ultra-long-term recordings. Furthermore, they are often recorded at high sample rates of at least $250\,\mathrm{Hz}$, producing large amounts of data that need to be stored and transferred. While some studies presented in this thesis also included devices with electrodes and wires to record biosignals, the research presented here excludes them for the above reasons.

Consumer-grade wearables and similar devices, however, most often record signals without the use of electrodes and wires. Thereby, sensors frequently target different aspects of the autonomic nervous system (ANS), with biosignal modalities such as electrodermal activity (EDA), photoplethysmography (PPG)[1], and skin temperature (TMP). Moreover, modalities such as accelerometry (ACC) and gyroscopy measure the motion of body parts. Section 2.2 goes into further detail regarding these biosignal modalities. A notable difference between electrographic and wearable biosignal modalities is the variety of the sample rates at which the devices typically record the signals. While brain activity changes occur in milliseconds, changes in the autonomic nervous system happen over seconds to minutes. Thus, sample rates of wearable biosignal recordings, specifically those made with consumer-grade non-electrographic wearables like smartwatches, usually do not exceed frequencies of $100\,\mathrm{Hz}$.

[1] From Ancient Greek *pháos*: light; *plēthusmós*: increasing, enlargement; *gráphō*: to write

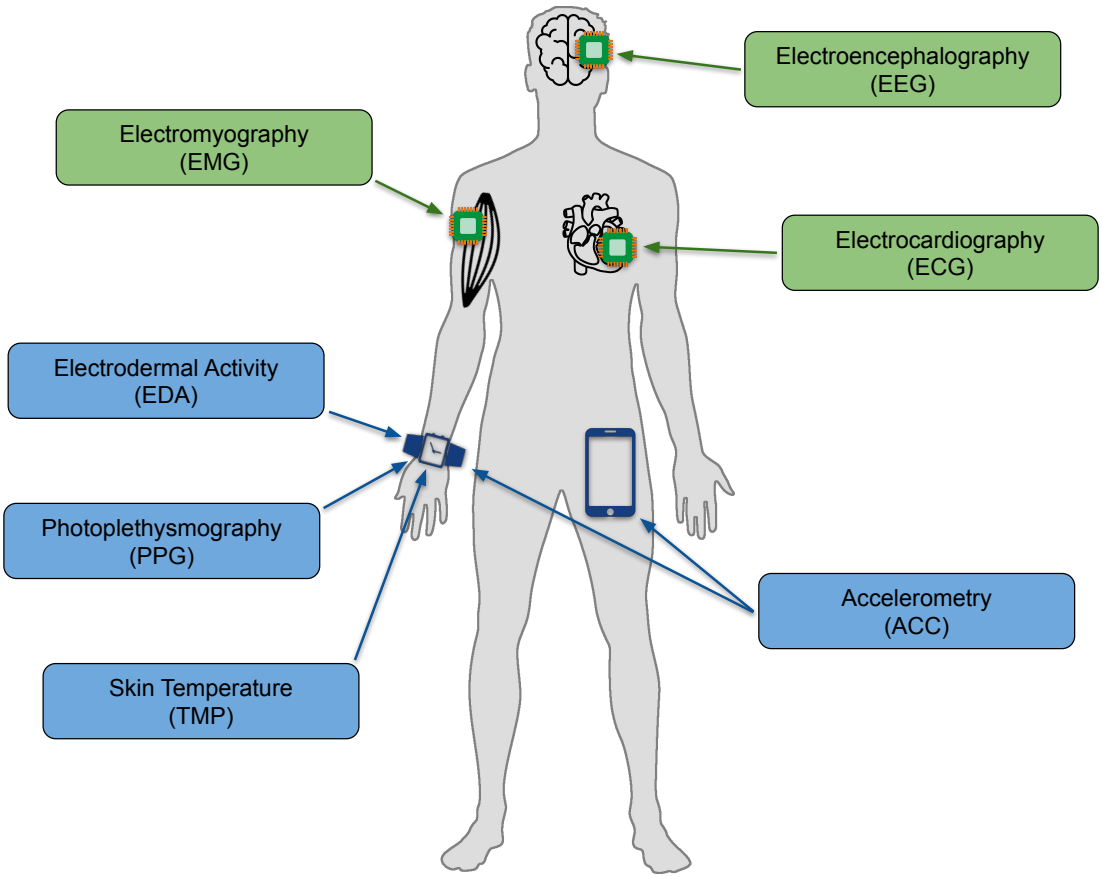


Figure 1.3: Overview of biosignal modalities typically recorded with different types of wearables. Devices with electrodes and wires record EEG, ECG, and EMG (green). Wearables such as a wristwatch record ACC, EDA, PPG, and TMP (blue). Smartphones commonly also record ACC, among other modalities.

In this thesis, the data analyses concerning the design and evaluation of seizure detection methodologies make use of the biosignal modalities ACC, EDA, and blood volume pulse (BVP) as a product of PPG. These modalities were recorded by the wrist-worn wearable (*Empatica E4*, see Section 3.2.1) employed in the data collection studies relevant to this thesis. While the device also recorded the TMP modality, these data were not included in the analyses presented here, as the effects of thermoregulation in epilepsy are not yet well understood and the quality of the sensor data is questionable (see Section 2.2.4). Furthermore, a different device worn by a small subset of study participants also recorded the electrographic modalities, which were not included here for the above-mentioned disadvantages devices measuring these modalities have.

ACC captures any movement the wearer of the device makes with the body part the wearable is attached to, usually the left or right arm in the case of this wrist-worn device, although it can be worn on the ankles as well. The output of the sensor is a three-dimensional vector comprising the accelerations in any direction of a Cartesian coordinate system, including earth's gravitational acceleration. Although ACC sensors are most often also included in modern smartphones, these data were not used in the analyses presented here as the study partici-

pants were unlikely to carry their phones with them while visiting the epilepsy monitoring unit and being constrained to a hospital bed.

EDA data represent changes in the electrical conductance of the skin at the sensor location. While sweat gland activity is the primary origin of these changes, EDA is a general surrogate of ANS activity, and symptoms like piloerection[1] may also have an effect on the EDA signal. As such, changes in the signal are typically slow and occur below a frequency of 1 Hz. Nevertheless, the sensor relies on constant contact to the skin and will only record zero-values as soon as contact is lost.

BVP is a signal produced by filtering and further processing the raw output of a PPG sensor. It is representative of the volume of arterial blood in the tissue where the sensor is located. A robust estimation of the heart rate from this signal is possible for periods of good data quality. However, the BVP signal is highly susceptible to artifacts from motion and external light sources. Thus, in this thesis, the signal was only considered for the analysis of less-severe focal seizures as the signal quality was insufficient during convulsive seizures.

For a more detailed description of these three biosignal modalities captured by the *Empatica E4* wearable and their application in further research, see Section 2.1.

[1] Colloquialism: “goose bumps”

1.3 Fundamentals of Seizure Detection

> « Until recently, recorded seizures have been inspected only if the time of their occurrence was reported [...]. Any unreported seizures were not examined because of the long time required to search the many hours of tape for only a few minutes of relevant data. In order to detect [...] seizures which might go unreported, an electronic circuit was designed to detect and report all electrographic seizure patterns. »
>
> — Thomas L. Babb et al., 1974 [75]

The automatic detection of seizures without the need for manual human labor has been a relevant problem since the beginnings of electrographic epilepsy examinations. Early research focused on analog electrical circuits and systems to detect seizures directly from electroencephalography signals [75], or simple digital computers implementing basic detection algorithms [76]. At the same time, machine learning methodologies developed and advanced, making use of the steadily rising computational capabilities of digital computers. Nowadays, modern seizure detection methodologies can draw on a vast library of different machine learning techniques.

1.3.1 Types of Machine Learning

Machine learning is a field of research concerned with designing methodologies and algorithms capable of inferring specific information from unspecific data. The constructs implementing this inference are called models, and the methodologies that govern these models are generally described by three different principles:

Supervised Learning: The input to the machine learning model is labeled data, that is, each data sample is given a specific description in relation to the problem at hand. For example, in seizure detection the simplest categorization would be a binary labeling of whether data samples are concurrent with a seizure or not. This data labeling is often gained from a gold standard knowledge source, for example, marked by epilepsy clinicians reviewing the video-electroencephalography data recorded at the epilepsy monitoring unit. The labeled data are used to "*train*" the machine learning model. The general goal is to be able to use this trained model to automatically derive labels for new, unlabeled data [77].

Unsupervised Learning: Data are not labeled by some external mechanism, such that machine learning methodologies are tasked with automatically finding structures or patterns in the data. A common type of unsupervised modeling are clustering algorithms, which find data samples that are similar to each other by some metric. Optimally, these clusters or patterns can then be linked to certain aspects of the problem at hand, and thereby classify the data. Semi-supervised learning is a hybrid type wherein only a small subset of data is labeled by hand, increasing the performance of the classification at the expense of some human overhead. With respect to seizure detection, an unsupervised detection model might detect anomalies where some data periods are substantially different from others, corresponding to times when a seizure occurred [78–80].

Reinforcement Learning: The machine learning algorithm gets direct feedback from the space it operates in with regard to its performance, being either rewarded or penalized. The algorithm can adjust its functionality in accordance, with the goal of maximizing rewards over time. Feedback can be given automatically by some outcome metric, or provided manually by a user. This often requires many cycles of the classification-reward mechanic and may be constrained by time and space requirements in a real-world application. A possible application of reinforcement learning in the field of epileptic seizure detection may implement a pre-trained supervised detection model which is further capable of receiving feedback from the patient whenever a seizure was detected, saying it was either a true seizure or a false detection, thus improving its performance over time [81, 82].

In the data analysis included here, only supervised machine learning is considered. While other methodologies would certainly be viable as well, the goal of the work that formed the basis of this thesis was to use simple supervised models that are easily reproducible, can achieve good performance on small data sets, and can be applied to data in real time.

1.3.2 Supervised Machine Learning Models

In supervised learning, algorithms trained with labeled data must generalize to previously unseen data. This poses a trade-off between bias and variance concerning the ability of a model to predict the correct output for some arbitrary input data. A model with a high bias, for example, would not properly represent the structure of the data because it makes certain incorrect assumptions about the complexity of the data, or lack thereof. It would be underfitted and thus miss some events, regarding them as unimportant because they do not fit into the structure of the model. Conversely, a high variance would cause a model to map to training data too well and regard noise as relevant information. It would be over-fitted and drastically change its output even with only minor changes in the training data. One of the goals of supervised machine learning models is to find a balance between these two aspects. Other factors are the amount, imbalance, dimensionality, and heterogeneity of training data [77, 83].

There are a multitude of different types of supervised machine learning models, and each type is further divided into different subtypes and implementations. Furthermore, there are some meta-methodologies concerned with improving predictive performance by creating ensembles of individual models. A detailed description of each of the supervised machine learning methodologies is beyond the scope of this introduction, but the following will give a brief overview of some major types of algorithms. See relevant introductory literature for more in-depth descriptions of these machine learning techniques [77, 84–86].

Decision Trees: A decision tree is a machine learning model that consists of several chained binary decisions that can be arranged in a tree-like structure. Each decision node splits according to a logic applied to an input variable, and each leaf node represents a possible outcome. In its absolute simplest form, a decision tree has only one node with two leafs, and the output is determined by the value of a single input variable [87].

Support Vector Machines: This type of methodology maps input data into a higher-dimensional space such that the distances between data points of differing output classes are maximized. The trained mapping is then applied to the unlabeled test data and the

prediction is made corresponding to the side of the decision gap the mapped sample falls on [88].

Nearest Neighbor: This algorithm compares data samples by some distance metric and selects a class according to the training samples nearest to the input sample. Like most supervised classifiers, it relies on sensible feature engineering and parameter tuning to achieve good performances. Dimensionality reduction is another preprocessing step that is often necessary for this type of learning, as high-dimensional data tend to require exponentially more training data to produce a robust classifier [89].

Naive Bayes: While the previous algorithms are deterministic in nature, Bayesian networks are probabilistic learners. Naive Bayes classifiers, a simple form of Bayesian networks, assume that the input variables are strongly independent and uses maximum likelihood estimation to infer the probabilities of a data sample to be of a certain class. They are theoretically not applicable if the independence assumption does not hold, although they have been used effectively regardless of that limitation [90].

Artificial Neural Networks: Based in principle on biological neural networks, these models are a collection of artificial neurons and connections, or more abstractly nodes and edges. This network propagates a high-dimensional input to a low-dimensional output, through a series of weighted functions. They typically implement back-propagation to gradually adjust weights over multiple cycles, representing the learning ability of the network. These models require a large amount of raw input data and are difficult to interpret in detail [91, 92].

In addition to these individual types of supervised machine learning, techniques to compile and employ multiple instances of these classifiers in a collection have been developed. While there are a few different methodologies of creating these ensembles of learners, the two most prominent and widely used are briefly described here:

Bagging: Bootstrap aggregation, also called bagging, is a method of independently training multiple individual classifiers with randomly sampled data from the training data set in parallel. During testing, new data are then given to all of these sub-learners and a majority vote decides on the output classification of the whole ensemble. This methodology may mitigate over-fitting that would occur if only a single instance of the individual learner was used, and thus will typically outperform them. It does however not help with under-fitting as a high bias in each individual model still results in a high output bias. The method of aggregating specifically decision tree models in this way is called *random forest* [93, 94].

Gradient Boosting: Contrary to bagging, gradient boosting trains multiple specifically weak learners that are only slightly better than chance, and does so iteratively and not in parallel. The output of the whole ensemble is described as a weighted sum of the individual learners' outputs. For each new learner that gets added to the ensemble the training samples are also weighted proportional to their error in the previous iteration by some loss function, such that new learners target training samples with worse performance. In each iteration multiple potential new learners are trained, and the one with the least amount of total weighted error is chosen to be added to the ensemble [95, 96].

In the main analysis work compiled in this thesis gradient boosted decision trees (GBT) models are employed as the classification model to implement the binary detection of epileptic seizures.

1.3.3 Seizure Detection Pipeline

Aside from the classification model itself, supervised seizure detection requires a number of different data analysis steps, which can be integrated into a complete seizure detection pipeline from raw input data to classification labels [84, 97, 98]. Figure 1.4a gives an overview of how this pipeline was designed and implemented in the work presented here. Raw data are first prepared through various different preprocessing steps before features are extracted in the next step. The feature vectors are then used to optimize and train the model, and other, previously unseen test data are then processed by the same pipeline to be classified by the created model. Lastly, the model performance is evaluated by comparing the classifications to the true labels. The following describes the sequence of analysis steps in more detail:

Data Preparation: Raw sensor data typically require some form of preprocessing before it can be used for further analysis. The amount and type of processing steps involved depend on the format and quality of the raw data and the purpose of the analysis. They can range from trivial operations like de-duplication or sorting, over data quality checks and synchronization of time stamps, to modifying processes like signal filtering.

Feature Extraction: Preprocessed sensor data can potentially be used directly, as is for instance done in artificial neural network models. In typical supervised learning applications, however, it is usually further processed into a number of different *features*, that is, calculated values representing different aspects and properties of the raw data. These features are most often calculated from sections of data instead of single samples, and thus data are first segmented before features are extracted. These data "windows" have a certain defined length and interval, both counted in number of samples, which can potentially result in overlap. The types of features themselves are highly dependent on the modality of the data and the purpose of the data analysis. Features are typically grouped as time domain and frequency domain features, according to the nature of the calculation with respect to the time series data.

Selection of Training and Test Data: Training data for seizure detection models are typically selected from the peri-ictal data of a number of seizures of a specific type, either from a single study participant or multiple different ones. In order to properly evaluate a trained detection model, the test data given to the final model for classification and evaluation need to be completely separate from the data the model was trained with. Accordingly, the data set is split into training and test data sets before creating the model, such that the trained model has never seen any of the data in the test set. In seizure detection, this is optimally done by selecting the test data from a different patient cohort than the training data, or at least selecting it from other patients in the same cohort, in the case of generic model evaluation. For individualized models, the test data would comprise the complete rest of the data for a particular participant, that was not selected for the training process, whereby the training data would comprise a specific set of seizures' peri-ictal data.

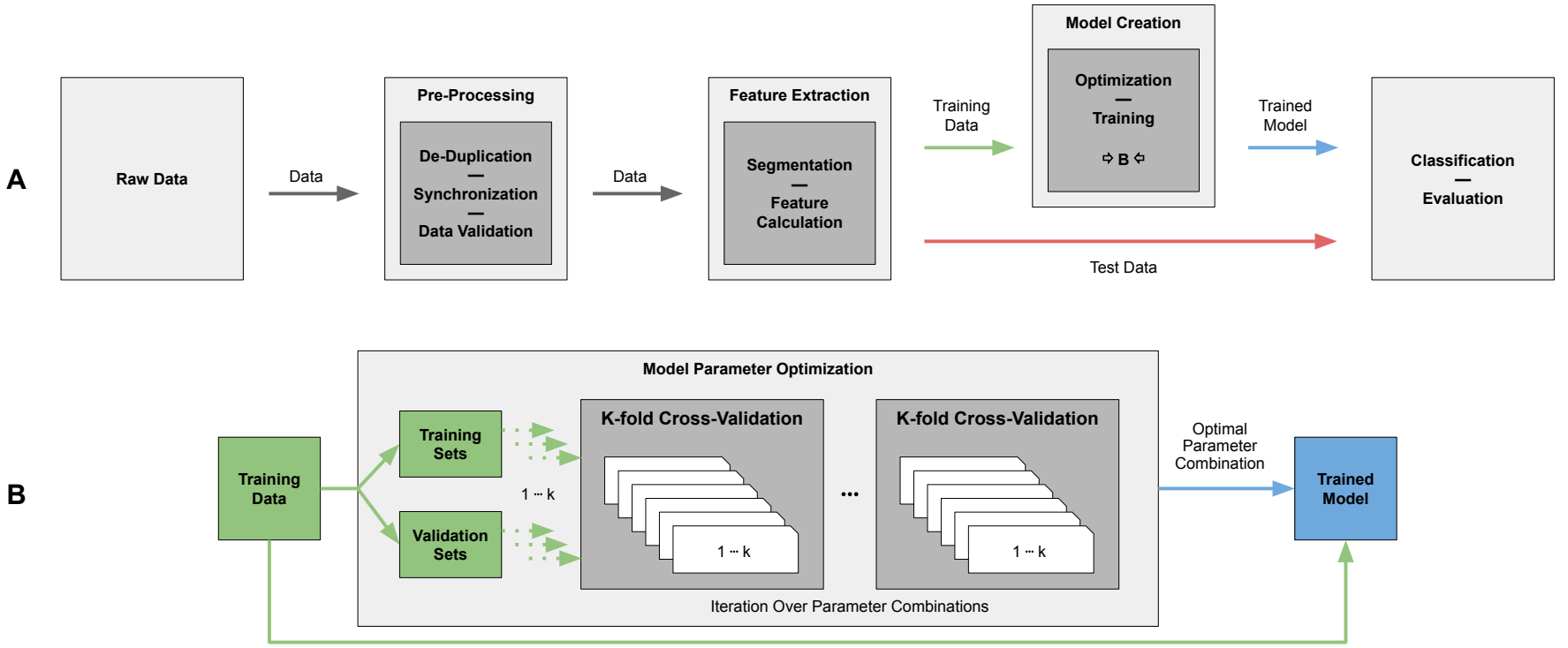


Figure 1.4: (a) Overview of the seizure detection pipeline as implemented in the data analyses presented here. The raw data pass through the same preprocessing and feature extraction steps, and is then split into training and test data. The training data are used to create the classification model, which is then applied to the test data to discover events (that is, detect seizures). **(b)** Further detail of the model creation process using parameter optimization by cross-validations. The incoming training data are further split into a number of training and validation set pairs. These are then used in a k-fold cross-validation, gauging the performance of a model with a specific parameter combination. This step is repeated with each possible parameter set, finding the optimal parameters to use in the final model.

Model Creation and Optimization: The process of creating a detection model with the GBT algorithm includes a parameter optimization step before training the final model with data. Figure 1.4b illustrates a more detailed overview of the parameter optimization process as it was implemented in the seizure detection pipeline employed here. Parameter optimization is carried out as follows:

1. For each model parameter that is to be optimized, a range of potential values is selected;
2. For each unique combination of parameter values, a k-fold cross-validation is performed with the input training data, split into training and validation sets and with the folds corresponding to, for example, the number of unique participants (leave-one-participant-out) or the number of seizures (leave-one-seizure-out);
3. The parameter combination that achieves the best mean performance with respect to the cross-validation is chosen for the final model;
4. The final model is trained with the whole training set.

Classification and Performance Evaluation: The test data are processed by the model resulting in a series of binary classifications for each of the feature vectors in the test set, either detecting an event or not. To evaluate these results, they are compared to the corresponding ground truth labels, and certain scoring metrics are calculated indicating the performance of the model on this out-of-sample test data. In seizure detection this is typically done by deriving ground truth and detection events from the sample-wise labels and checking for overlaps between those events, such that an overlap between a ground truth and prediction event would constitute a true positive, a predicted event without a corresponding event in the ground truth would count as a false positive, and a false negative in the reverse case. From these direct confusion matrix scores, other metrics like the sensitivity and false alarm rate are derived. These metrics can then be used to compare the performances of different methodologies, even across studies, although discrepancies between studies in terms of study cohorts, the size of the data sets, and data pre- and postprocessing need to be kept in mind.

1.4 Motivation

> « Why do we need wearable devices in epilepsy? Seizure diaries derived from seizures reported by patients and caregivers are unreliable, yet they constitute the input for therapeutic decisions in clinical practice and for the outcomes in drug trials. An objective quantification of seizure burden could improve clinical decision-making and the quality of the drug trials. »
>
> — Sándor Beniczky et al., 2020 [99]

Broadly, the goals and use cases of automatic seizure detection are twofold: Providing an alarm system or generating a seizure diary. In this thesis, the feasibility of automatically generating a seizure diary by way of seizure detection with wearable data is investigated. Nevertheless, patients who suffer from seizures that could be dangerous or life-threatening, that is, being at risk of sudden unexpected death in epilepsy, need seizure alarms. Especially during the night or if the patient is alone, an alarm could automatically notify caregivers or emergency services to provide help. As such, the performance requirements for seizure alarm systems are high, necessitating short latencies from seizure onset to alarm. Some research on automated seizure alarm systems has used wearable systems, and some products on the market already offer this functionality, albeit only under specific conditions and for convulsive seizures.

Furthermore, while several studies have employed at-home video-electroencephalography systems similar to those used in epilepsy monitoring units [56, 100, 101], several fundamental challenges currently limit their feasibility and thus their adoption in general clinical practice. They are limited to a few days of recording time, are impractical to set up, present considerable overhead concerning integration with in-hospital services, and pose regulatory challenges to overcome [100].

This thesis investigates seizure detection in automated seizure diaries for ultra-long-term epilepsy monitoring with wearables. The primary methodology for monitoring epilepsy and seizures in an ambulatory setting in clinical use is record-keeping in manual seizure diaries (Section 1.1.3). Manual seizure diaries, however, are highly unreliable [54, 102] (Figure 1.5). Estimates show that patients typically self-report less than $50\,\%$ of seizures, especially concerning seizures occurring during sleep [59, 102]. Wearable devices could fill this gap in ultra-long-term epilepsy monitoring, providing relevant biosignals that enable the automated detection of epileptic seizures.

To explore the feasibility of automated seizure diaries, biosignal data from wearable devices were collected from patients with epilepsy and several evaluations of methodologies were done in the context of this thesis. The relevant data collection studies and investigations were conducted within the scope of the European Union collaborative research project *Remote Assessment of Disease and Relapse - Central Nervous System*[1]. Within this major undertaking, numerous research groups developed new ways of monitoring major depressive disorder, multiple sclerosis, and epilepsy with the help of wearable devices.

[1] radar-cns.org

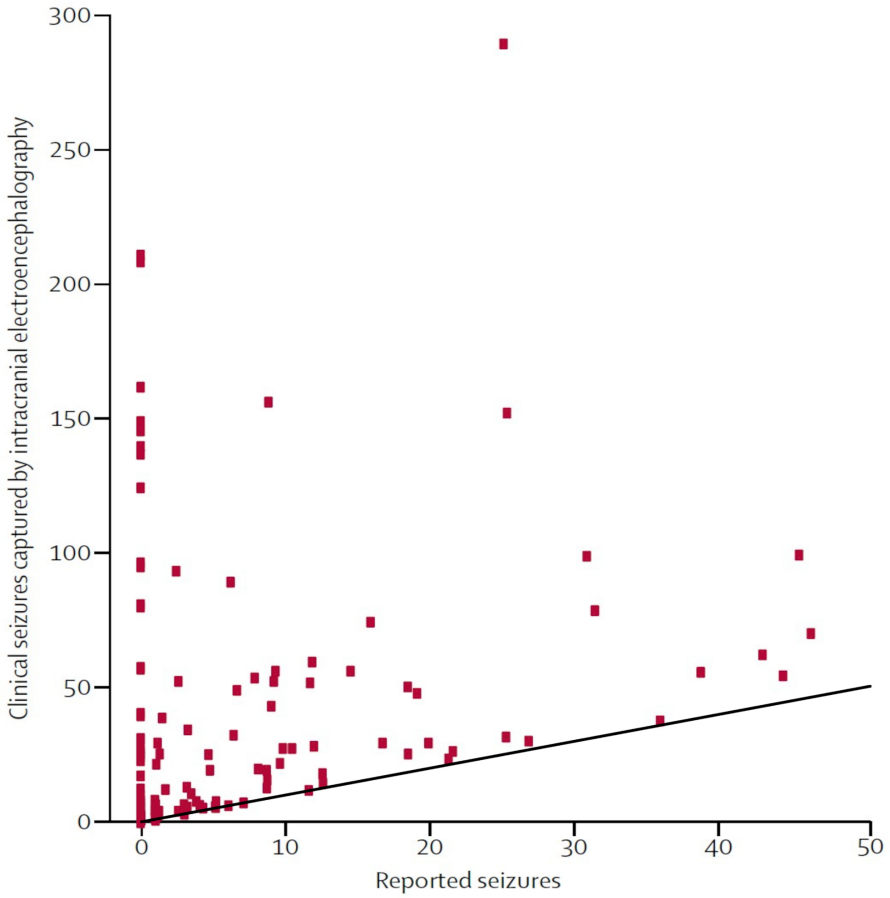


Figure 1.5: Ambulatory seizure diaries significantly under-report events compared to seizures discovered from ambulatory intracranial electroencephalography. From Cook et al. [54], Figure 4, identity line added to illustrate the optimal case.

1.5 Key Contributions and Outline

This thesis provides a framework for several published articles on seizure detection with wearable devices. It focuses on detecting motor-onset seizures using multimodal data from a wrist-worn, smartwatch-like wearable. After this introduction, the thesis comprises four parts. The second chapter explores relevant research in epilepsy monitoring with wearables. The subsequent three chapters highlight five publications with substantial contributions to this subject:

Methods and Study Design: This chapter describes the data collection study procedures. *Bruno and Böttcher et al. (2021) [103]* compiled multiple studies with wearable devices in epilepsy monitoring conducted by an international group of epilepsy researchers. In particular, this part emphasizes the procedures of the in-patient study at the epilepsy monitoring units of the University Medical Center Freiburg and the King's College Hospital London, which provided the data set used in most of the other work presented in this thesis.

Detection of Major Convulsive Seizures: The detection of convulsive seizures, such as generalized tonic-clonic seizures and focal to bilateral tonic-clonic seizures, is the first step to a functioning seizure detection system, as they are generally considered the most straightforward to detect among the different seizure types, particularly with data from non-EEG wearable devices. *Böttcher et al. (2021) [104]* describe, implement, and evaluate a seizure detection pipeline for the supervised classification of these convulsive seizures. *Bruno and Böttcher et al. (2020) [105]* outline a possible application of tonic-clonic seizure detection, which explores the feasibility of detecting post-ictal immobility and unconsciousness, a major sudden unexpected death in epilepsy risk factor, by wearable biosignal data.

Detection of Focal Onset Seizures: While detecting major convulsive seizures was possible with robust performances, focal seizures with less convulsive motor onsets often manifest in more moderate or arbitrary ways. *Böttcher et al. (2019) [106]* explore the heterogeneous manifestations of focal motor seizures. Furthermore, *Böttcher et al. (2022) [107]* modify the existing seizure detection pipeline and evaluate it on focal motor seizures, highlighting differences between individually trained models and inter-participant detectors.

Finally, the last chapter concludes with a brief discussion of the presented research and an outlook of possible paths to further advance the field, both in general seizure detection research in controlled in-patient conditions and concerning ambulatory contexts. For each of the five major publications included in this thesis, a more detailed listing of own contributions is given at the start of the corresponding section.

CHAPTER 2

Related Work

Seizure detection with wearables has been a research topic for over 10 years now, and as wearable technology develops with its increased use in society and popular culture, it is becoming more and more relevant for epilepsy as well. After the previous chapter introduced the general topic of this thesis and important fundamentals, this chapter dives deeper into the state-of-the-art of some related fields of research.

Monitoring epilepsy patients is a necessary clinical tool not only to diagnose and treat individuals, but also to develop new avenues of treatment (Section 2.1). While some possibilities exist to monitor patients at home with video or video-electroencephalography systems, they are typically short-term, cumbersome, and intrusive. Wearables on the other hand are much less burdensome, but still lack in seizure detection performance. Further, patients' and caregivers' needs should be taken into account when developing wearable seizure monitoring technologies.

The cornerstone of wearable monitoring are biosignal sensors, technology that has existed and has been developed for much longer than wearable epilepsy monitoring itself (Section 2.2). A multitude of different biosignal modalities are available in modern sensor systems, each with their own opportunities and pitfalls. Here, relevant sensor technology in the context of epilepsy monitoring is introduced and discussed.

Seizure detection with wearables, while still a relatively new field of research, has already undergone meaningful development since its inception (Section 2.3). Especially with regard to convulsive motor seizures, detection is already at a stage allowing for first prospective ambulatory clinical trials, a major step towards widespread adoption of the technology in clinical routine. However, the detection of other less severe or non-motor seizures is still in its infancy, and new promising methodologies are being developed every year [108–110].

Seizure prediction and forecasting is a related yet different field of research, aiming to predict seizures before they occur, or at least give estimates of seizure likelihood for some timeframe (Section 2.4). Even further removed, but in some key ways still comparable, is the monitoring of other conditions similar to epilepsy. Among these are neurological disorders such as multiple sclerosis and Parkinson's disease, but also other types of seizures that are not epileptic in nature (Section 2.5).

2.1 Epilepsy Monitoring with Wearables

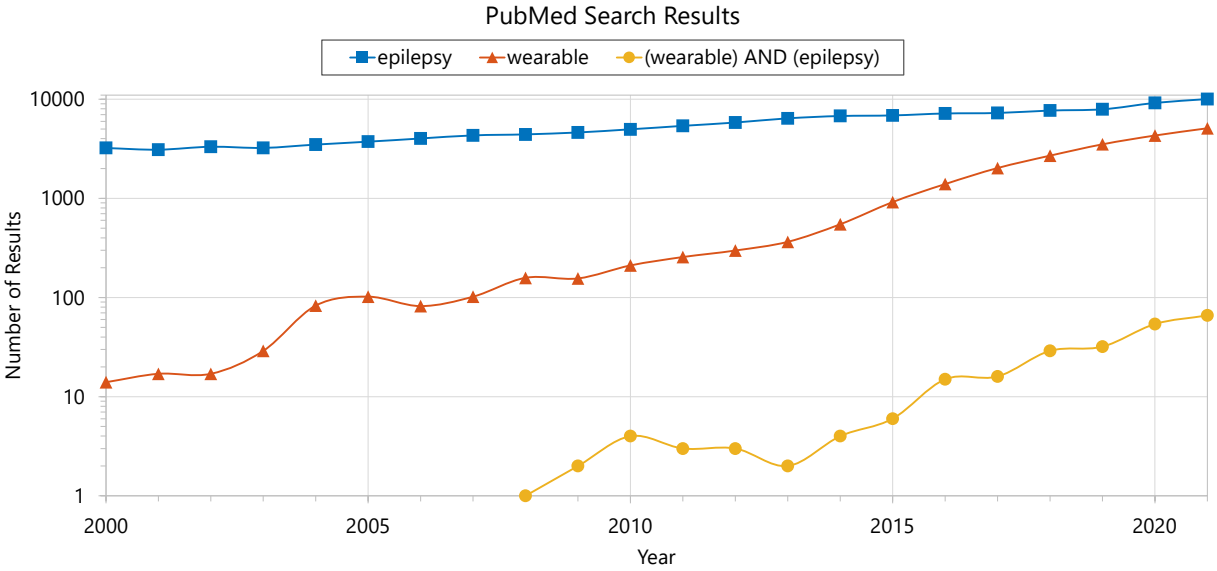


Figure 2.1: Illustration of yearly publication frequency in the years 2000 to 2021 of related work including the search terms “epilepsy”, “wearable”, and both.

In this section the advancements and state-of-the-art in epilepsy monitoring are highlighted. Wearable devices and their applications have gone through extensive development and research in the past, especially in the last 10-20 years. Not only has their size decreased and computing power increased substantially, but the sensor technology integrated into the devices has improved as well [111, 112]. In the same way, usage of wearable devices in epilepsy research has picked up substantially in the past 10 years. Figure 2.1 shows this development over the past 20 years on the basis of publication frequency, as indexed on PubMed. Wearables in healthcare monitoring have the potential to provide real-time, continuous, and ultra-long-term biosignal data on a multitude of aspects of the human body [111]. A wide-spread adoption of wearable devices in public healthcare systems, however, is still hindered by factors such as security and privacy [112], integration into electronic health records [113], and engineering challenges like power supply and communication abilities [111, 114]. Yet, wearables are featured in a large variety of state-of-the-art healthcare research, from classic use cases like cardiology [115], over sophisticated microfluidic sensing platforms [111], to applications in psychophysiology [116].

2.1.1 Monitoring of Patients with Epilepsy

In the diagnosis and treatment of epilepsy, short- and long-term monitoring in the in-patient as well as the out-patient setting are essential [102]. Not only is monitoring necessary to separate epilepsy from non-epileptic syndromes like psychogenic non-epileptic seizures [117], but also to improve the management of already diagnosed epilepsy [118]. Another important indication for admission to a monitoring unit is presurgical evaluation, that is, determining if a patient is a candidate for a resective brain surgery [61]. However, the informational yield of in-hospital video-electroencephalography (vEEG) monitoring is reduced significantly after

only a few days [50], such that monitoring outside the hospital environment in an ambulatory setting is necessary.

VEEG monitoring at home has been used in several studies to varying but generally positive results [56, 100]. While hospital conditions ensure good quality recordings and the highest chances of capturing seizure events, at-home monitoring is overall more accepted by patients and incurs less cost [56], making it more suitable as a first step in managing an individual's epilepsy. Nevertheless, ambulatory vEEG monitoring is limited to the patient's home, requiring wearable devices for continuous monitoring, for example, while travelling [100].

Implantable sub-scalp electroencephalography (EEG) devices are in development, potentially providing robust recordings of ultra-long-term ambulatory EEG [119]. Adoption of these devices may substantially change the landscape of ambulatory epilepsy monitoring. Furthermore, smartphones can serve as an ad hoc solution to provide video monitoring of seizures outside of dedicated video monitoring units [120]. Data from ambulatory epilepsy monitoring with wearable devices may also contribute to advances in seizure forecasting and prediction [60]. Multiple studies have investigated the feasibility of wearable non-EEG devices for epilepsy monitoring and seizure detection out of the hospital, but assessment without an established gold standard is still problematic [110, 121–123].

2.1.2 Experiences and Needs of Patients

While wearable devices are much less intrusive and burdensome than vEEG systems, there are still a variety of aspects to consider when it comes to their acceptability and adoption among patients with epilepsy [124]. A need to balance costs and benefits arises for patients with respect to wearing devices, with potential costs being higher stigma and increased anxiety, and perceived benefits being improved safety and feedback on their condition [125]. Additionally, patients are now in charge of facilitating effective data recordings, and devices and systems need to account for the large heterogeneity in technology management skills of patients [126]. Other important factors to consider are visibility, wear comfort, removability, technical support, and knowledge or at least perception of the effectiveness of the wearable [72]. In one study the perceived performance of the wearable device, that is, the number of missed seizures or false alarms, was even the main reason for participants to stop using the device, with the least important factor being the design of the device [127]. However, wearables that utilize electrodes, cables, patches, and similar parts can substantially change the self-awareness of patients with epilepsy and can lead to higher perceived stigma, with the devices further highlighting their condition [128]. The monetary cost of a wearable device is also a potential barrier and should not be disregarded when considering the impact on patients [129].

These experiences and views of patients result in a number of requirements for wearable devices. Fundamentally, the clinician's need for a robust seizure detection device translates to patients' and caregivers' needs for improved safety, independence, and management of the epilepsy [130]. They must be able to trust a given device, that is, be able to depend on it performing without fail, a need which is amplified when considering pediatric patients and their parental caregivers [131, 132].

2.2 Biosignal Modalities

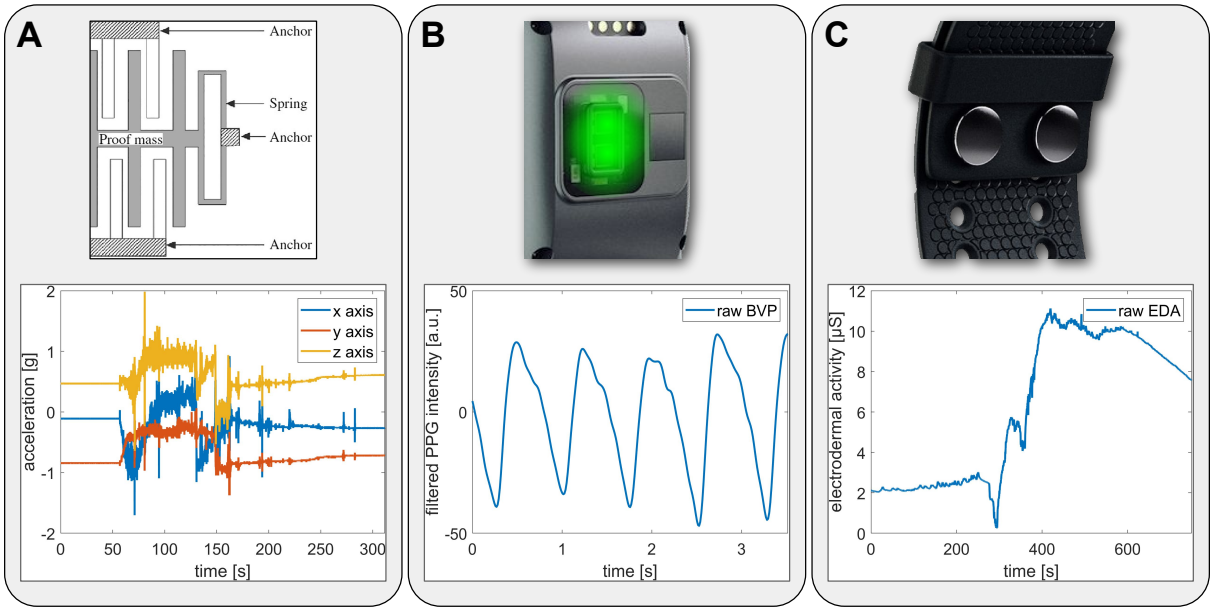


Figure 2.2: Biosignal modalities and associated sensors. **(a)** Schematic overview of a capacitive microelectromechanical accelerometry (ACC) sensor, from Dwivedi et al. [133], Figure 3, modified; Example of raw ACC data. **(b)** Photoplethysmography (PPG) sensor at the back of the Empatica E4 device; Example of raw blood volume pulse (BVP) data. **(c)** Electrodermal activity (EDA) sensor dry electrodes built into the wrist strap of the Empatica E4 device; Example of raw EDA data.

This section highlights some of the biosignal modalities that can be used for monitoring epilepsy and seizures, and outlines the working principles of the respective sensors. Specifically, information on the four main biosignal modalities captured by the Empatica E4 device is compiled (see also Section 3.2.1). Figure 2.2 gives a brief overview of some of these modalities. Movement is registered by ACC sensors and is a major source of information regarding epileptic seizures with motor components [134]. EDA is an important surrogate for autonomic nervous system activity [135]. Cardiovascular activity during seizures can be measured by PPG sensors [136]. Finally, while not specifically used in the analyses presented here, other relevant biosignals include the skin temperature (TMP) [137] and electrographic modalities like electromyography (EMG).

2.2.1 Movement

ACC sensors in the form of microelectromechanical systems have been some of the most ubiquitously used sensors in smart technology and nowadays are a regular part of wearables, smartphones, and other such devices [138]. The accelerometer captures activity induced by the sensor's motion or displacement, typically by variable capacitive measurement (although piezoresistive ACC sensors also exist), whereby a freely movable comb-like proof mass swings between fixed counterparts [139]. This structure effectively combines into capacitors with varying distances between plates, and thus the capacitance is measurable in an analog elec-

trical circuit. If the sensor begins to move in the direction of its free axis, inertia compels the proof mass to lag behind the fixed part of the sensor, which is measurable as a change of capacitance. To achieve three-axis accelerometers, either three such sensors are combined in one package, one in every direction in three-dimensional space, or in more complex chips a single proof mass is used with a series of springs providing the necessary degrees of freedom [140]. An analog-to-digital converter then translates the analog signals from each sensor to digital outputs, and a controller interprets them as amplitudes in the unit of g, or $\mathrm{m/s^2}$. Important to note is that, as the name suggests, these devices only capture the acceleration of a movement. Thus, an ACC sensor moving at a constant speed will show a constant signal. This signal includes the earth's gravitational acceleration of $\sim 9.81\,\mathrm{m/s^2}$ or $1\,\mathrm{g}$, which a filter may later isolate from the linear movements in the sensor output, producing an estimate of orientation [141]. Furthermore, ACC sensors for human activity recognition typically employ sampling frequencies of $25\,\mathrm{Hz}$ and above [97], capturing typical human motions and epileptic movement patterns, both of which generally don't exceed frequencies of $10\,\mathrm{Hz}$ [142, 143].

Accelerometers are generally robust against measurement artifacts. Fundamentally, any movement large enough to cause a change in the analog capacitance could be measured, and there are no external, physical, or other reasons for the sensor to produce a relevant change in the signal if there was no movement, provided that it is correctly placed on the body. The most important constraint of the ACC sensor system is the analog-to-digital conversion, which places a minimum and maximum range and sensitivity on the output signal [144]. Controller instructions can sometimes change these parameters to be more or less sensitive towards smaller movements. For example, a sensor range of $\pm 2\,\mathrm{g}$ means that the output signal can be at most two times the earth's gravitational acceleration. For movements of more considerable acceleration, as would be expected in clonic movements of generalized tonic-clonic seizures, the output signal will be saturated, that is, capped at the value of $\pm 2\,\mathrm{g}$.

Generally, with regard to epileptic seizures, ACC signals can show very heterogeneous patterns, depending on the seizure type and specific semiologies during the seizure. Figure 2.3 shows sample ACC recordings of three different seizures with movement components. Convulsive tonic-clonic seizures produce high-amplitude and high-frequency variations in the ACC signal, that can easily saturate the sensor range. Tonic movements of focal motor seizures display as changes in the gravitational component of the signal with overlaid low-amplitude tremors visible in the linear component. Myoclonic jerks can sometimes be visible in the ACC trace as small signal spikes, but only if no other movements are occurring at the same time, and they are usually indistinguishable from other random short movements.

Besides accelerometers, specialized in capturing linear translation, gyroscopy (GYR) sensors can measure rotation more accurately. In many cases, these two sensors are used in conjunction, applying sensor fusion to better estimate the device's position and angle [145]. An inertial measurement unit combines both ACC and GYR sensors, sometimes also including a magnetometer [146].

2.2.2 Electrodermal Activity

With increasing body or skin temperature, the sweat gland activity increases, and sweat production changes the skin's electric properties like its conductance, which EDA recordings can capture [148]. Besides thermoregulatory processes, other sympathetic nervous system responses like piloerection or psychophysiological arousal may also induce changes in the

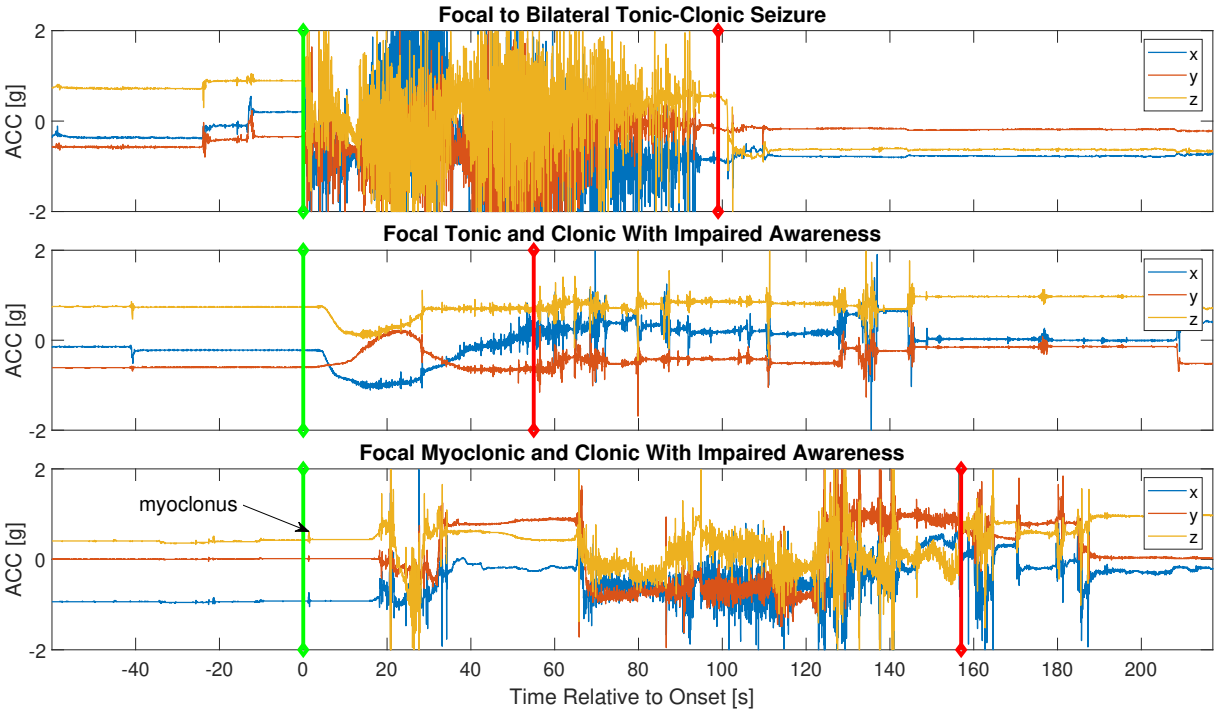


Figure 2.3: Three-axis ACC signals from three epileptic seizures of different types, recorded by a sensor with a range of ± 2 g. The green and red vertical lines mark seizure onset and offset, respectively. **(top)** Convulsive focal to bilateral tonic-clonic seizure. **(middle)** Focal seizure (FS) with impaired awareness, starting with tonic and evolving to clonic arm movements. **(bottom)** FS with impaired awareness, with marked myoclonus at seizure onset, and clonic movements throughout. See also Section 1.1.2. Remaster of Figure 1 in Schulze-Bonhage et al. [147], own work.

EDA signal [149, 150]. The literature describes two components that contribute to measurable changes in EDA: The fast galvanic skin responses, also called phasic responses, occur in an order of 0.5 to 5 seconds, while the tonic component develops as changes in the EDA level over multiple minutes [135, 151]. These changes in EDA have also been linked to epileptic seizures, predominantly showing as tonic increases in the post-ictal phase [42, 43, 152, 153].

Sensors typically record EDA using dry electrodes, which do not necessarily need to be adhesive, as long as they have continuous contact with the skin [154]. For instance, some devices, integrate electrodes into a bracelet or the device housing itself. The sensor can record the biosignal in two different methods: The passive endosomatic method, measuring the electrical properties of the body without an external electrical source, relying only on changes in electrical energy originating from the skin; And the active exosomatic method, using an external electrical current source to measure the electrophysical properties of the skin directly [155]. While sensors rarely use the former method, the latter is the most prevalent and is further divided into sensor systems with direct current (DC) or alternating current (AC) sources [155]. Depending on the internal circuitry used in the sensor, either the skin conductance or resistance is measured in a constant voltage or a constant current system, respectively, if supplied by a DC source. Likewise, the skin admittance or impedance are measured with the use of an AC source. In all cases, the resulting current through the skin is minuscule [156]. The devices in the studies presented here only employ sensors using the exosomatic AC method.

EDA signals are commonly recorded at sample rates of above 1 Hz, while the relevant frequency range of EDA information in the signal is below 1 Hz for both the phasic and tonic activities [157]. Low-pass or band-pass filters can filter typical sensor noise on the signal in the range of >1 Hz. Connection loss of the electrodes to the skin due to, for example, body motion often causes artifacts in the EDA signal [158]. In these cases, the output signal will show a sharp decrease or increase within just a few samples or fall to a minimum/zero-line value altogether. Simple thresholding or an analysis of the rate of amplitude change can detect these kinds of artifacts [159].

2.2.3 Photoplethysmography

PPG is an optical measurement method to determine changes in volume in the blood flow of a specific body part [160]. The recorded PPG signal is thus sometimes also called BVP. There are two types of PPG sensors: Transmissive mode PPG measures the amount of light arriving at the other side of the body part when shining light through the body, and reflective mode PPG which measures the amount of light reflected when shining it through the skin [136]. Transmissive mode is thereby constrained to body parts thin enough for light to shine through them, and clinical settings commonly use them in the form of pulse oximeters applied to a person's finger, not only measuring the change in blood volume but also the oxygen saturation [161]. Reflective mode is more frequently used in ambulatory and wearable contexts, as it can be applied anywhere on the body, such as on the wrist or upper arm.

In reflective PPG measurements, a light-emitting diode (LED) shines light toward the skin, and a photoelectric sensor records the amount of reflected light, modulated by the different light absorption rates of the tissue at the recording location [162] and the blood during a typical pulse wave cycle [163]. While green light has a lesser optical penetration depth than red and infrared light [164], shorter light wavelengths also have a higher absorption coefficient in blood [165], resulting in a higher signal-to-noise ratio with respect to monitoring blood volume in tissue. Thus, green LEDs are generally considered a better alternative to red ones for PPG sensors, which has been confirmed in comparative studies [166, 167]. Additionally, PPG sensors can measure changes in blood oxygen saturation by combining green, red, and infrared light [160, 168, 169].

The shape of a clean PPG signal is closely related to the blood pulse wave of the human body and can vary depending on the specific measurement site [170]. Generally, a heartbeat is visible through the signal's rising edge, culminating in the systolic peak, followed by a decline towards the dicrotic notch and diastolic peak, and ending in a further decline towards a baseline before the next pulse starts [171]. Figure 2.4 shows the BVP signal in detail. The heart rate (HR) is thereby directly and accurately derivable from the peak to peak intervals between single systolic peaks. Depending on the circumstances and properties of the sensor, a lower-frequency component attributed to the breathing frequency of the subject may also modulate the raw signal [172]. Some sensors and devices, however, may immediately filter out this signal component and not include it in the output data. The PPG signal may also sometimes be referred to as BVP in literature, and in this thesis the term BVP is used for the filtered output signal of a wearable device, whereas PPG is the raw sensor data.

Because the raw output signal of a PPG sensor is a function of the amount of light falling into the photoelectric sensor, external light contaminating the reflected light from the sensor's LED can heavily skew this signal [136, 160, 173]. This signal corruption is especially critical

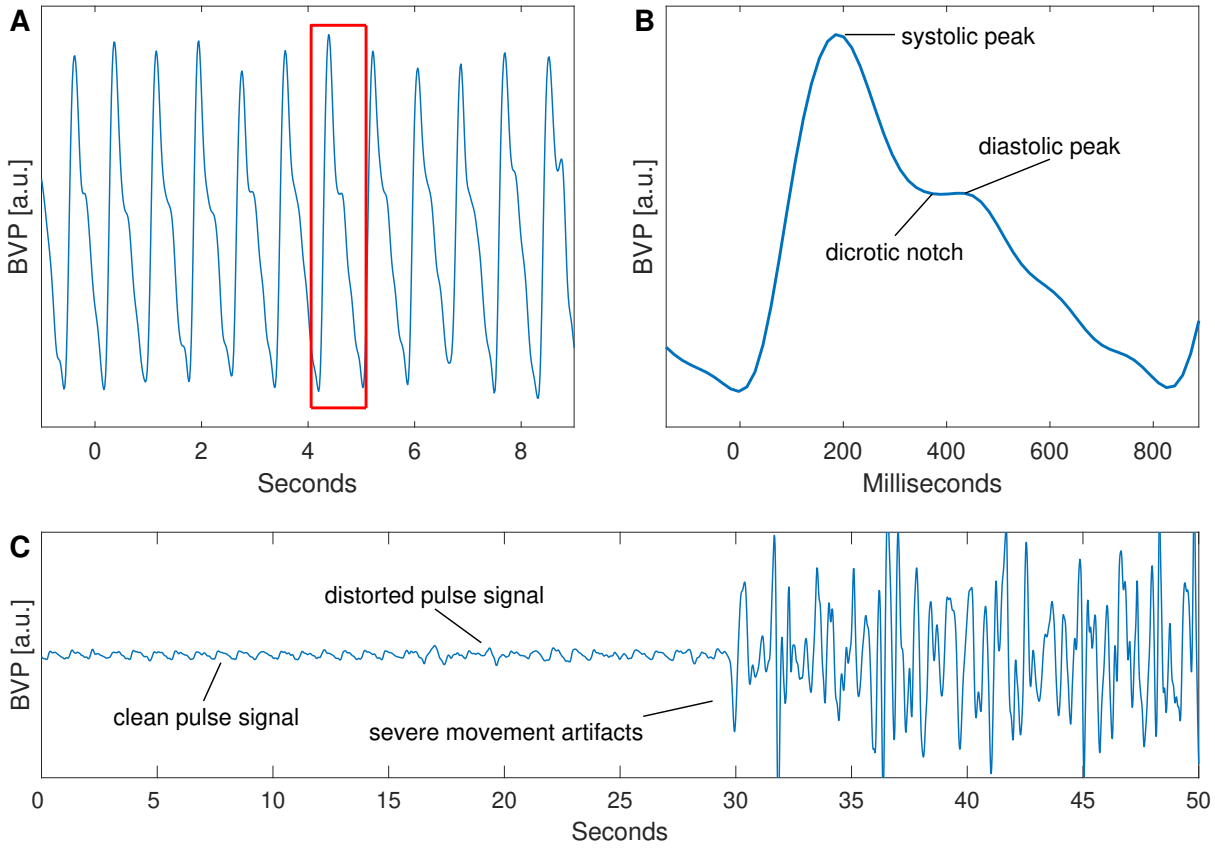


Figure 2.4: Detailed look at the BVP raw signal as recorded by the Empatica E4 wearable device. **(a)** 10 s span of clean BVP data at excellent signal quality. The red square marks the data shown in (b). **(b)** Pulse wave of a single heart beat. **(c)** Example of bad BVP signal quality, including a short period of slightly distorted pulses and severe artifacts caused by continued motion of the device.

in wearable devices since the sensor might only be loosely attached to the body to keep it comfortable to wear. Any movement of the device or the body part it is attached to may result in the sensor lifting from the skin, which will invariably result in some motion artifact in the signal [174]. While some research presents methods to clean PPG signals from motion artifacts [175–177], including some using a simultaneously recorded ACC signal [178–180], they are only effective in removing minor short artifacts or those resulting from repetitive motions like running [181].

2.2.4 Temperature

Some wearables can record thermal activity by measuring the skin temperature (TMP) of the user's body. Note that this does not capture the core temperature, so measured values are often lower than those commonly known for human core body temperature, for example, when determining fever, and are more prone to environmental influences [182]. Thus, TMP reflects a combination of the ambient and skin temperature at the recording location [183]. TMP signals are prone to sudden modulation as the environment changes, for example, due to the subject going outside in winter or putting on clothes that cover the body part to which

the wearable is attached. Furthermore, the baseline of the signal may change substantially in a seasonal context, as average temperatures would be higher during the summer than during the winter. Like other responses of the sympathetic nervous system, changes in peripheral body temperature can take multiple seconds or minutes and might occur delayed [184]. Therefore, artifacts in the temperature signal can register as sudden spikes or dips, for example, as the subject removes the wearable from the body for a short time. These could be detected by monitoring and thresholding the rate of amplitude change, similar to the artifacts seen in the EDA signal [159].

Temperature is commonly measured with wearables by one of two principles: Direct measurement with the sensor unit in immediate contact with the skin or indirect by infrared radiation measurements. The former methodology is most commonly achieved using a thermistor component, a resistor highly dependent on temperature [137]. The latter methodology can be implemented by thermopiles, components that connect different kinds of metal in series and thus exploit the thermoelectric effect, or pyroelectric sensors, measuring changes in the charge over the surface of pyroelectric crystals [185]. Concerning epilepsy, some research has investigated the role of thermoregulation [186, 187], however its causes, effects, and interactions are still poorly understood. Overall, there is no conclusive evidence as to the relevance of TMP measurements for epilepsy monitoring and the analyses presented in this thesis do not include this signal.

2.2.5 Other Sensors and Biosignals

Electrodes attached to stationary systems in a hospital or dedicated monitoring unit usually capture electrographic signals like electroencephalography (EEG), electrocardiography (ECG), and EMG. Wearables recording these biosignal modalities do exist, although they frequently include cumbersome wires and adhesive electrode patches, making them less comfortable and more noticeable to wear, the opposite of how users prefer wearable devices to be [124, 188]. Furthermore, other biosignal recording systems that are not wearable but relevant to the topic are mattress sensors and at-home video and audio recording systems.

EEG signals measure electric activity on the scalp as a surrogate marker of brain activity [189]. Wet electrodes attached to the skin by conductive gel or dry electrodes using adhesive patches or caps typically capture EEG. The internationally used 10-20 system provides the standard electrode placement and naming convention [51, 52]. In epilepsy, EEG systems are most notably used in epilepsy monitoring units (EMUs) in combination with video monitoring, as outlined in Section 1.1.3. Wearable EEG devices for ambulatory recordings have existed for some time, traditionally as full-fledged electrode arrays in the form of a dry electrode cap worn over the whole head [190]. More recently, low- or single-channel EEG wearables are developing, using dry electrodes [73, 191] or even subcutaneously implanted electrodes [192] to measure the electrical activity of specific brain areas.

Clinical ECG measures the heart's and its muscles' electrical activity by electrodes attached to the skin [193]. Multiple electrodes distributed over the thorax region and limbs facilitate proper ECG measurements [194]. Recorded data provide features such as HR, heart rate variability, and further derived characteristics. Section 2.2.3 describes that PPG sensors are often used in wearable contexts to estimate these features. Nevertheless, some wearable ECG devices exist. For example, some devices use electrodes with wires similar to clinical systems [74], others have integrated the electrodes into chest bands or wearable fabric [193, 195], implantable

ECG sensors are sometimes used for ambulatory measurements [196], and even smartwatches measuring ECG exist [197].

Like ECG measurements, EMG records the electrical activity generated by muscle tissue when activated but is specific to skeletal muscles [198]. Wearable EMG systems are often designed to fit around the limbs, measuring muscle activity during arm movements [199], for example, and have been used in a multitude of research [200–203].

In epilepsy biosignal monitoring, video and audio signals are paramount to discovering and classifying seizures [50, 100, 118]. However, the use of these sensors is often limited to EMUs (Section 1.1.3). Nevertheless, at-home video monitoring finds application in some cases where longer-term recordings, which are not feasible at the hospital, are necessary [204], and dedicated home-video recording systems are in development [205–207]. Similarly, pressure sensors under a mattress can indicate epileptic activity in seizures with convulsive motor manifestations [208–210], but these sensors have some noteworthy disadvantages [102, 211], making them niche devices unattractive for regular clinical use. Section 2.5 compiles further related work regarding applications of wearables outside the context of epilepsy.

2.3 Seizure Detection with Wearables

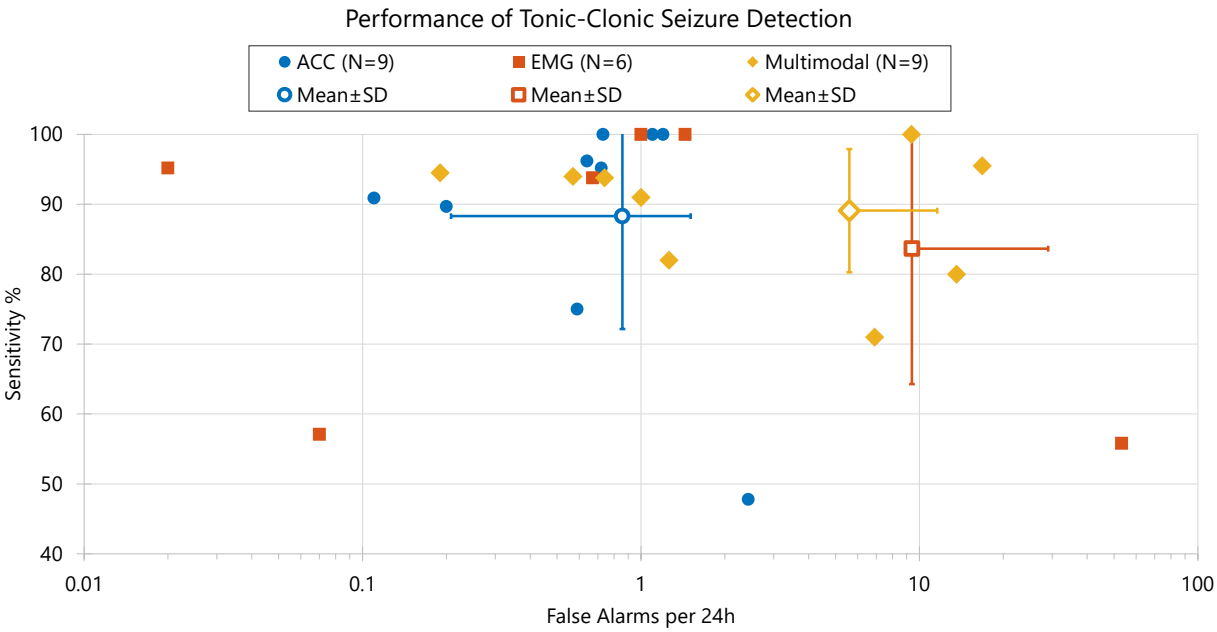


Figure 2.5: Performance of tonic-clonic seizure (TCS) detection in related work, grouped by biosignal modality used in the detection model. "Multimodal" represents studies using multiple different combinations of the biosignals accelerometry (ACC), electrodermal activity (EDA), heart rate (HR), and electromyography (EMG) for their seizure detection. Data taken from Naganur et al. [108].
SD: standard deviation.

In this section, the state-of-the-art literature concerning seizure detection with wearables is presented. In recent years a variety of review articles have been published from different groups, assessing wearable devices and their performances in seizure detection methodologies. A clinical practice guideline and standards for testing have been developed, introducing best practices for evaluating and validating seizure detection devices and methodologies [109, 212]. An exhaustive review of commercially available seizure detection devices has been published, compiling their capabilities and evidence of performance, if available [122]. Furthermore, recent reviews of studies predominantly evaluating convulsive TCSs have shown that mobile health devices are a promising method for epilepsy monitoring especially in out-of-hospital contexts (Figure 2.5) [108, 109, 121, 213]. However, while there are a number of studies evaluating the performance of focal seizure (FS) detection, there still is a lack of evidence for a robust detection even in controlled conditions, much less in an ambulatory setting [110].

2.3.1 Major Convulsive Seizures

Primarily convulsive TCSs, that is, generalized tonic-clonic seizures or focal to bilateral tonic-clonic seizures (FBTCSs), are the most dangerous and severe seizure type due to their violent pathology and high risk factor in the occurrence of sudden unexpected death in epilepsy (SUDEP) [18, 19]. Thus, a robust detection of these seizures is of utmost importance not only in the development of alarm systems, but also the improvement of seizure diaries for

the advancement of therapies. Furthermore, their symptoms inherently make TCSs the most straightforward seizure type to detect with wearables. Accordingly, there already are numerous published studies investigating these seizures with a variety of devices and methodologies, some few even rated as phase 4 studies designed prospectively and with a predefined and fixed methodology [109, 212]. The current best state-of-the-art performances reported in some studies include sensitivities of >90 % and false alarm rates (FARs) of as low as 0.2 per 24 h.

Monomodal detection of TCSs has been considered with almost every kind of relevant biosignal modality that can be measured from wearables. ACC is the most obvious modality to use for these convulsive motor seizures [134]. Johansson et al. [214], for example, use a wrist-worn inertial sensor to detect TCSs in adult patients visiting a video-electroencephalography monitoring unit. They evaluate three different supervised learning algorithms, with the best sensitivity resulting from a nearest neighbor model, detecting all the 37 TCSs, and the best FAR coming from a random forest model with 0.24 per day. Thus, they report performances comparable to some current multimodal systems. Kusmakar et al. [215] also use a wrist-worn ACC device to find TCSs with an outlier detection algorithm. They report a sensitivity of 95 % and a FAR of 0.72 per day.

Like most epileptic seizures [216], TCSs have a strong cardiac component in their ictal and post-ictal phase, and cardiac abnormalities are a major risk factor concerning SUDEP [217]. Thus, the detection of TCSs with HR and heart rate variability (HRV) data from wearables has been investigated by several studies. Mohammadpour Touserkani et al. [218] record photoplethysmography (PPG) signals from a wrist- or ankle-worn wearable sensor and extract several features from it. They do not specifically investigate a detection algorithm, but describe in detail the pre- and post-seizure periods of TCSs compared to a baseline in terms of these extracted features. The authors do this to avoid the severe motion artifacts PPG signals would have during the ictal phase of TCSs, which further highlights the lack of evidence of studies using PPG in monomodal TCS detection. There are some studies using electrocardiography (ECG) signals instead to detect convulsive seizures, but these are usually recorded from hospital monitoring systems [219, 220].

Beniczky et al. [203] and Zibrandtsen et al. [221] both use wearable EMG devices to detect TCSs in epilepsy patients, the former with a surface patch EMG device and the latter with an in-ear device. Both report sensitivities of above 90 % and less than one false positive per day, but neither comments on the comfort and acceptance of the device from the patients' point of view. Furthermore, there have also been numerous studies in the past few years investigating the value of wearable electroencephalography (EEG) recordings [121]. Lastly, some efforts have also been made to detect convulsive seizures during the night with the help of at-home video systems, using video motion analysis to isolate periods of high activity [204, 222].

Multimodal detection of TCSs, that is, using two or more different biosignal modalities, with data from wearable devices, is the next logical advancements from the monomodal detection that the previously mentioned studies use. Onorati et al. [223, 224] use ACC and EDA data from a wrist-worn wearable device to detect primarily convulsive TCSs, although they also included a few FSs with tonic and clonic motor components in some of their evaluations. They train a support vector machine (SVM) model with feature sets from both modalities, and in their most recent work describing a prospective clinical validation study [223], they report sensitivities of up to 98 % and FARs below one per day. However, their detailed methodology and algorithms are proprietary, as the research group is directly linked to the manufacturer

of the wearable device they used, the *Empatica E4* also used in the studies included here (see Section 3.1.3 and Section 3.2.1).

Multimodal detection has also been investigated in other groups. Tang et al. [225] work with a wrist-worn wearable device (the same as Onorati et al. [223, 224]) to record ACC, EDA, and blood volume pulse (BVP) data, using different combinations of features from these signals to train convolutional neural network models. Their data set includes TCSs as well as different types of FSs recorded from a predominantly pediatric cohort. For TCSs, they report a sensitivity of $80\,\%$ and a FAR of more than 13 per day when using the combination of ACC and EDA features. Interestingly, the combination using features from all three modalities did not perform best for any of the seizure types included in this study. Arends et al. [226] present a study with an upper-arm-worn wearable device recording ACC and PPG data exclusively during the night. Their methodology includes thresholds for features calculated from both signals, and additionally checks if the ACC signal indicates that the subject is lying down before issuing an alarm. They report a sensitivity of $86\,\%$ and a FAR of 0.25 per 8-hour night. Cogan et al. [227] recorded HR, blood oxygen saturation, and EDA data with a wearable device for the evaluation of a multi-stage seizure detection methodology. They included three convulsive seizures and 23 complex partial seizures (CPSs) in their data set, although they do not further specify whether the CPSs included motion manifestations. They report better FAR performances when extending their base methodology with a personalization step.

Figure 2.5 shows the performances of some studies investigating TCS detection using both monomodal and multimodal methods [108]. A direct quantitative comparison of the results in this way is subject to uncertainties like differences in modalities, timespans of as much as 10 years between studies, or fundamental differences in methodologies, and thus unadvisable. Nevertheless, it gives an idea of rough average performances of related work, which helps to contextualize the results of this thesis, especially those of the TCS detection presented in Section 4.1.

2.3.2 Focal Motor Seizures

FSs are the most commonly occurring type of epileptic seizure in the general population [15, 228, 229]. However, due to their relatively lower footprint and higher heterogeneity in terms of possible symptoms, they are substantially harder to detect from typical data recorded with wearable devices. Here, FSs do not include FBTCSs, even though they technically have a focal onset. In a later part of the thesis, Table 5.7 also compiles a list of related work that includes FS in the evaluation of various seizure detection methodologies.

Vandecasteele et al. [74] use a SVM model to detect FSs with impaired awareness from HRV features, extracted from both ECG and PPG signals. They compare the performances of the models and conclude that the data from a wrist-worn PPG sensor are insufficient for seizure detection (sensitivity $32\,\%$), whereas the wearable ECG data show promise (sensitivity $70\,\%$). They specifically acknowledge motion artifacts as the main reason for the poor performance of the model based on PPG data.

Munch Nielsen et al. [230] propose a multimodal FS detection and record ECG, ACC, and behind-the-ear EEG from wearable devices. In their analysis the authors included one patient with predominantly motor FSs with stereotypical tonic movement symptoms. Employing a SVM classifier, the authors report a sensitivity of up to $91\,\%$ and as low as 8 false alarms per day, depending on the decision threshold.

Baumgartner et al. [231] extract features from surface EMG signals and use them in an amplitude analysis threshold detection methodology, detecting TCSs as well as a variety of different motor FSs. Among the FS types, they could correctly detect $50\,\%$ of seizures with tonic and/or clonic features, and $76\,\%$ of seizures with other symptoms like automatisms or hyperkinetic movements. They do not further specify a FAR, but show that automated detection performance is similar to expert rater classification from the same data.

Tang et al. [225] also include FSs in their data set (see Section 2.3.1), but do not further specify performances of their methodology for these seizure types in a comparable manner. However, they report that ACC is the best modality for focal tonic seizures, EDA together with BVP for automatisms, and BVP for behavior arrests. They also conclude that individualized models may improve detection performance.

Vandecasteele et al. [191] use recordings from behind-the-ear EEG electrodes in a semi-automated FS detection. First, they had an expert rater visually annotate seizures in the data, and then used these labels to train a SVM detection model. In a patient-specific evaluation they could report a sensitivity of $69\,\%$ and 0.5 false alarms per day. The authors conclude that while these results seem low, they are still better than patient self-reports for these types of seizures.

Due to their typically short duration of just a few seconds and low movement profile of slight jerks of a body part, myoclonic FSs are some of the hardest epileptic seizure types to detect from wearable device data. Thus, research investigating this problem is sporadic at best. Hyppönen et al. [232], for example, develop a video-based detection system for the evaluation of myoclonic FSs. Specifically, they aim to automatically quantify myoclonic jerks by amplitude and frequency, estimated from optical flow of the video. They report a high level of agreement of their methodology with clinical evaluation, however their data set included a limited number of subjects and video was recorded only during explicit and supervised tasks.

2.3.3 Focal Non-Motor and Other Seizures

To date there is only sporadic research specifically investigating the detection of non-motor FSs with wearables. Some studies include these seizure types as a subset, sometimes more as an aside to convulsive seizures than their own seizure type. Not only is the detection of non-motor seizures an inherently more difficult task, but there is also less immediate need in general for a robust detection system for these seizures, as opposed to convulsive seizures that are much more intrusive and dangerous to the patient. Studies considering the detection of non-motor FSs, naturally, are often limited to non-movement-related biosignal modalities. Note also that for the purposes of this thesis, seizure semiologies such as oral automatisms or other small movements of the head are regarded as non-motor features, as they do not register on typical wearable movement sensors placed on the wrist or similar body locations.

Tachycardia, that is, a substantial rise in HR over a short amount of time or a consistently high HR over longer periods, is a major symptom of epileptic seizures in general, and also non-motor seizures specifically [12, 29, 37]. Thus, many studies about the detection of non-motor seizures with wearables focus on modalities like ECG or PPG, and HR or HRV signals, to detect events [41]. However, the restriction to only HR and HRV features has shown to be problematic when applied to general population cohorts. Ictal autonomic changes are limited to “responders”, that is, only some individual patients exhibit these symptoms during seizures [219, 233]. Furthermore, the current evidence of FS detection with monomodal sig-

nals shows performances of below $80\,\%$ sensitivity with often multiple false detections per hour [234, 235].

Munch Nielsen et al. [230] propose a multimodal FS detection and record ECG, ACC, and behind-the-ear EEG from wearable devices. In their analysis the authors included two patients with predominantly non-motor FSs with symptoms like auras and behavioral arrests. Employing a SVM classifier, the authors report that all non-motor seizures could be found, with FARs of 5 and 13 per day for the two participants.

Vandecasteele et al. [236] evaluate a multimodal detection algorithm to find predominantly FSs with impaired awareness using ECG and behind-the-ear EEG. While the authors give no further information on the seizure pathology or potential ictal movements, the application of wearable EEG suggests that most seizures would not have had prevalent motor features due to the motion artifacts they would have caused in the signal. They report an average sensitivity of $89\,\%$ with FARs of less than two per hour over three different cohort data sets.

Of note concerning the general body of related work on FS detection is that many studies do not further specify if movements during or in the peri-ictal proximity of the seizures were included in their analysis [74, 227, 237, 238]. Studies will usually categorize seizures by one of the clinical classifications (e.g., Fisher et al. [27]), but if there is no specific focus on motor seizures in their evaluation and modalities like ACC are not used, information about movements during seizures is often omitted. While the inclusion of this information may not seem immediately relevant for this type of evaluation, it is important to include it regardless. Artifacts from movement are a major disruptive factor for all physiological modalities recorded from wearables. Thus, including at least a binary movement or no movement categorization during seizures is fundamentally relevant for the interpretation of wearable seizure detection results.

2.3.4 Relation to Own Work

The research presented in this thesis specifically targets the detection of epileptic seizures with motor manifestations of the limbs, both with convulsive TCSs (Section 2.3.1) and FSs (Section 2.3.2) with tonic or clonic semiology. For the detection convulsive TCS, both monomodal and multimodal methodologies have been shown to be successful. Figure 2.5 actually seems to indicate a preference for monomodal detection with ACC signals which, on average, trends towards lower FARs than multimodal detection. However, this does not account for differences in wearable sensors, study protocols, or recent developments in the state-of-the-art. Current studies using multimodal detection with both the ACC and EDA modalities, including the analysis included in this thesis (Section 4.1.3), have reported better results than the average monomodal detection using just ACC.

Related work concerning the detection of FSs, with or without movement components, is even more heterogeneous. At the same time, there are also less published studies investigating the wearable non-EEG detection of exclusively FSs, as compared to studies including TCSs. Furthermore, current work generally does not highlight the need for, and utility of, individualized detection models. Thus, the analyses included in this thesis, specifically the work included in Section 5.2, represents a new and further step towards robust FS detection. While the work presented here focuses on seizures with epileptic movements, it could also be used similarly for non-motor FSs, possibly with a different combination of biosignals that omits ACC in favor of other signals representing the autonomic nervous system.

2.4 Seizure Prediction and Forecasting

Seizure prediction, as opposed to detection, specifically refers to the intention of predicting seizures before they happen, with the goal of potential intervention. Seizure forecasting is sometimes regarded as a subfield of study that usually doesn't aim to predict a specific future seizure but estimates the likelihood of seizures happening at some point in the future [239, 240]. Both prediction and forecasting have been shown to be possible for some patients with epilepsy in controlled retrospective studies, but application in the real world is still lacking [240, 241]. While seizure prediction research predominantly started by using in-hospital electroencephalography (EEG) data, other biomarkers have since been introduced, in particular also with the application of wearable devices, looking towards ultra-long-term ambulatory forecasting [53, 60, 239].

Generally speaking, seizure prediction methodologies found in literature can be separated into two broad categories: those that explicitly leverage cyclic structures in the re-occurrence of seizures, and those that do not. At first glance, epilepsy seems to appear as random occurrences of seizures. However, rhythmic manifestation in circadian[1], multidien[2], or even circannual[3] cycles can be found in a considerable portion of patients [242, 243]. Concerning circadian rhythms, differences between daytime and nighttime periods are the driving factor, affecting the autonomic nervous system [244], and thus possibly the occurrence of some seizure types [42, 245]. These rhythms can be captured by wearable devices recording data like heart rate (HR) or electrodermal activity [246, 247]. Multidien HR cycles could also be linked to seizure occurrence, with the data being recorded by a wearable device with a photoplethysmography sensor [248].

Data from wearable devices have also been used in some studies investigating seizure forecasting without expressly looking for rhythmic patterns. Notably, deep learning approaches have emerged as successful in direct seizure prediction [249, 250]. In a seminal study, ultra-long-term intracranial EEG data recorded with electrodes implanted on the surface of the brain were successfully used to estimate seizure likelihood in a small cohort of patients [54]. In another study using intracranial EEG recorded in the hospital environment, early data until up to 10 seconds after seizure onset could infer the seizure type of the ongoing seizure [251]. While not strictly seizure prediction, this result may still provide value towards early warning and intervention systems, specifically in those seizures where clinical symptom onset is delayed after EEG onset. Some research has also investigated the feasibility of seizure predictors from metadata such as manual seizure diaries kept by the patients [252].

There has notably also been some progress made towards crowdsourcing the problem of seizure prediction [253, 254]. Potential solutions are dependent on evaluation with ultra-long-term data recordings that include multiple seizures over periods of sometimes months, and this kind of data are generally not readily available to all researchers. Disseminating data sets in a crowdsourcing manner may help substantially in improving prediction and forecasting algorithms.

[1] 24-hour
[2] multi-day
[3] yearly

2.5 Non-epileptic Seizures and Similar Conditions

In the wider field of general healthcare there are a variety of conditions and adverse events that can manifest with symptoms comparable to those of epileptic seizures. The most immediate relative are psychogenic non-epileptic seizures (PNESs), which are also often explicitly considered in epilepsy research. Further, there are a number of diseases that are symptomatically related to epilepsy, like multiple sclerosis (MS) and Parkinson's disease (PD), both of which include motor spasms or tremors in their list of possible symptoms. In addition, there are some more distantly related problems in general healthcare-related activity tracking.

A PNES is a seizure event initiated by psychological mechanisms, as opposed to specific epileptic activity in the brain. One of the main outcomes of in-hospital monitoring units is the differential diagnosis between epileptic seizures and PNES, and there are a variety of different diagnostic tools to aid in this effort [255]. Some research has investigated distinction between the two from physiological non-video-electroencephalography data. For example, muscle activation patterns, captured by surface electromyography signals, have been found to indicate the type of seizure event [201, 256]. Similarly, movement patterns captured by accelerometry (ACC) were used to distinguish between convulsive epileptic seizures, PNES, and normal movements [142, 257].

MS is a disease of the central nervous system characterized by a wide range of symptoms, including motor dysfunctions which can manifest as walking impairments. Therefore, gait analysis has been a target of research, employing ACC sensors as a source of movement data [258, 259]. In this context, wearable devices have similar advantages as in epilepsy research: they provide continuous data in a long-term ambulatory setting while being unobtrusive and easy to use [260–262].

PD is a neurodegenerative condition that manifests in different symptoms, with involuntary tremors being a fundamental indication that includes a motor component. PD is also a highly progressive disease, highlighting the need for continuous monitoring and management [263]. Therefore, ACC-based monitoring has been investigated to both classify and quantify tremors using, for example, smartphones [264] or wrist-worn wearables [265, 266].

Automatically detecting falls would give meaningful opportunity for timely intervention and assistance, and research on this topic has seen an upsurge in recent years [267]. Not only is it highly relevant for the care of elderly people, but there is a meaningful connection to epilepsy as well, with some seizure types like atonic motor seizures often resulting in the patient suddenly falling to the ground [268]. Fall detection has been attempted with a variety of different sensors, among them audio/video systems [269] and ACC sensors [270].

Lastly, there are a multitude of other applications of wearable sensors in healthcare [271], like monitoring physical activity [272], sleep quality assessment [273], observation of obsessive compulsive disorder [274], affect recognition [116], gait analysis [275], cardiac monitoring [276], diabetes treatment [277], or smoking cessation [278].

2.6 Summary

This chapter compiled the state-of-the-art literature in epilepsy monitoring and seizure detection with wearable devices. The usage of wearables in epilepsy research has only begun in the last 10-15 years, but since then the number of publications on this topic grew exponentially. Robust monitoring of epilepsy not only in the hospital but also at home is necessary to enable effective diagnosis, intervention, management, and research. While the gold standard of video-electroencephalography is used in monitoring units, it is not feasible for ambulatory assessment, and wearables could fill this gap. However, patients also want effective, affordable, and inconspicuous devices.

Wearables can record a variety of different biosignals, some of which are relevant to epilepsy. Accelerometry captures movements of the body, data which can be used to train detection models for epileptic seizures with movement manifestations. Electrodermal activity sensors record electrical properties of the skin that can change during seizures, but are dependent on dry electrodes and skin contact. Photoplethysmography signals can be used to estimate heart rate (HR) and other cardiac features, but the sensors are very susceptible to motion artifacts. Other biosignal modalities exist and are recorded by a number of different sensors, but these are not relevant to the content of this thesis.

Seizure detection can be monomodal or multimodal, that is, using one or more biosignal modality, but a general preference is not obvious from current literature and depends on context. The detection of tonic-clonic seizures has the highest-performing evidence in related work, and phase 4 prospective ambulatory studies are beginning to yield results. Motor focal seizures (FSs) can be difficult to detect, but some evidence shows that there is good potential for robust wearable seizure detection systems, especially if model personalization is implemented. Non-motor FSs are possibly the hardest to detect with wearables, relying on physiological modalities like HR, and there is little evidence in literature so far that attempts detection with non-electroencephalography (EEG) data.

Both the prediction of seizures and the detection of non-epileptic seizure-like events are not directly involved in the research included in this thesis. However, they are closely related in that general concepts of machine learning can be applied similarly to these topics. Seizure prediction aims to predict seizures before they happen, or in the case of forecasting, to estimate a probability of seizures occurring in a specific time frame. Generally, this is done either by leveraging predictable cyclic structures in the occurrences of seizures for some patients, or by deep learning modeling with wearable data. As these tasks require a lot of data over months or years, data sharing and crowdsourcing has a high priority. Non-epileptic seizures are seizure-like events that do not show typical epileptic EEG patterns, and the closest relative in this regard are psychogenic non-epileptic seizures. To correctly diagnose these, epilepsy monitoring units perform differential diagnosis. Other diseases with partially similar symptoms to epilepsy and where wearables are used in research are multiple sclerosis or Parkinson's disease. The analysis methodologies presented in this thesis could be modified to enable the detection of events like these, likely involving the use of different types of biosignals or features thereof. From a broader perspective, the use of wearables in epilepsy is only one medical application, with others including fields like fall detection, sleep quality assessment, or cardiological applications.

This chapter established the context of current related work in the field of seizure detection with wearable non-EEG sensors, both in terms of relevant methodologies and qualitative

performances, facilitating a placement of the results included in the following two chapters within the state-of-the-art. Concluding from current developments in the field, the detection of epileptic seizures with low-profile, low-cost, non-EEG wearables is a very promising approach, and the research presented in this thesis takes a meaningful step towards robust detection systems.

CHAPTER 3

Methods and Study Design

This chapter contextualizes the thesis within the scope of the data collection studies conducted to gather wearable biosignal data from patients with epilepsy. *Remote Assessment of Disease and Relapse - Central Nervous System* was a collaborative European research project that included, among other studies for other neurological disorders, several epilepsy-centered data collection studies. Specifically, patients with epilepsy were recruited at the epilepsy monitoring units of the University Medical Center Freiburg (UKF) and the King's College Hospital London (KCL). Both in-hospital and ambulatory data collection was implemented, however, this thesis only includes data from the former. Data collection in the patients' everyday life, while an important experience, was not done with the evaluation of seizure detection methodology in mind. Rather, it was primarily focused on assessing the feasibility of such an ambulatory study with wearables in general.

The first section of this chapter (Section 3.1) was previously published as a collaborative journal article [103] commenting on the wearable study experiences of four different international epilepsy centers, including those that this thesis' data set was recorded at. Therefore, it includes information on data collection procedures, wearable device choices, considerations regarding the reporting of results for seizure detection studies, and other such discussion. As this thesis' author's main contributions in this article were the data collection studies at the aforementioned UKF and KCL sites, only information concerning these were included in this chapter.

Extending on those notes, Section 3.2 gives further details on the selected wearable device, the technical implementation of the data collection, and basic biosignal data preprocessing done for all the analyses included in the remainder of the thesis. These individual data analysis studies will in turn include more information on methodologies specific to the respective evaluation, like distinct feature set computations and seizure detection models.

3.1 Study Procedures

[103] ⇒ Bruno, Elisa and **Böttcher, Sebastian**, et al.
Wearable devices for seizure detection: Practical experiences and recommendations from the Wearables for Epilepsy And Research (WEAR) International Study Group
2021, Epilepsia, DOI:10.1111/epi.17044

Parts of this publication were removed or edited to fit into the composition of this complete thesis. No substantial changes altering the results were made.

Own Contributions:

- Contribution to description of clinical study design (3.1.1)
- Description of data collection and technical infrastructure (3.1.2)
- Description of devices (3.1.3)
- Description of reporting results (3.1.4)
- Description of data sharing (3.1.5)

3.1.1 Clinical Study Design

Identification of Study Aims

The definition of study aims and related methods determine the patient selection, device choice, data annotations, curation, and data analysis. Seizure detection may serve many different purposes, from closed-loop treatment of acute seizures and impending status epilepticus [279, 280], to retrospective assessment of clinical seizure burden and assessment of the risk of sudden unexpected death in epilepsy (SUDEP), as well as evaluation of clinical devices or medication trials [281]. Given the unreliability of self-reported seizure diaries [54, 102, 282], an accurate seizure detection device could be used to optimize medical treatment, avoiding undertreatment due to unreported seizures, and minimizing unnecessary side effects due to seizure over-reporting. Offline detection could contribute to the diagnosis of non-epileptic paroxysmal events, from psychogenic seizures [142, 201, 257, 283] to cardiogenic events. Seizure detection devices may also be studied for their potential to measure disease severity, for example, pathologies like ictal autonomic changes [153], ictal surface electromyography patterns [284], post-ictal immobility [105], and post-ictal central apnea [285] are all potentially measurable by wearable devices and are associated with post-ictal generalized electroencephalography suppression, a risk factor for SUDEP.

The particular seizure semiology types targeted in a study may affect the study design, device choice, and data annotation protocols. Generalized tonic-clonic seizures and focal motor seizures with limb involvement may require movement or electromyography sensor devices and may prompt placement of devices on the body segment with greatest ictal movements, whereas non-motor seizure types like focal impaired awareness seizures may require devices

that sense autonomic biomarkers such as electrodermal activity, heart rate (HR), or skin temperature, or a combination of these. Detection of daytime seizures requires wearable devices to be mobile and to be robust to patient movement, whereas devices for detection of nocturnal seizures may be stationary and attached to the patient [226, 286] or the bed [208, 287], or a camera may be pointed at the patient from a fixed location [204]. Device acceptability and adherence by patients are essential in seizure detection, and device studies should include assessment of acceptability in the overall study aims [130].

The primary aims of the studies included in this thesis were to study the detection feasibility and optimal combinations of bio-signals obtained from non-electroencephalography (EEG) wearable sensors, for the detection of convulsive tonic-clonic seizures and focal seizures with motor and/or autonomic features, as compared with video-electroencephalography (vEEG) ground truth. Further aims were the assessment of patients' acceptance towards wearing the devices and the technical feasibility of conducting studies with remote measurement technology in cohorts of epilepsy patients.

Policies and Agreements

The process to obtain ethical approval from an institutional review board or ethics committee (EC) may be time-consuming and requires careful planning. Essential steps include delineating a clear research plan and developing the study protocol, but also seeking agreements with device manufacturers, interacting with hospital authorities, and arranging monitoring plans.

The process of obtaining informed consent is regulated by principles embodied in the current biomedical research on human subjects [288], which also considers the needs of vulnerable populations (e.g., children, cognitively impaired or unconscious patients) [289]. Comprehensive information must be provided to enable people to voluntarily decide whether to participate in a research study and is essential for valid informed consent as defined by the Guidelines for Good Clinical Practice [288]. Despite the low invasiveness of wearable devices, studies involving wearable devices are subject to these regulations, and in particular the transfer and sharing of anonymized data with other groups (see Section 3.1.5) requires approval. In particular, sharing anonymized data internationally can be heavily regulated and may require specific consent by the research subject. Opt-in and opt-out policies, also called nudges, have the tendency to promote one choice in favor of the other, while still keeping the choice easy to avoid [290]. Of course, this will also need to conform to local data protection regulations.

Each study center must be guided by its country's local policies and regulations, and additional approvals may be needed when testing devices without existing Conformité Européenne (CE) or US Food and Drug Administration approvals. Such studies may be considered clinical trials or performance evaluation studies, requiring additional documentation and in some cases authorization by government regulatory bodies. The rules vary in different countries, and this generates disparities in how devices can be tested and scientific data acquired. In this thesis, the device used for data collection already had a CE mark as a medical device, and thus the studies were not considered as clinical trials.

Another important consideration is the security rules governing the computer network infrastructure in the hospital environment. Hospitals regulate and limit access to internal networks to protect sensitive data, and specific approvals are often required to use existing wireless connections or create new networks. Data safety and protection are important considerations, especially with the European Union (EU) General Data Protection Regulation (GDPR) governing data collection and transfer inside the EU and with international collaborators. All

clinical institutions based within the EU must follow these rules, whether collecting data or receiving data from partners outside the EU.

The studies presented in this thesis followed these principles and considerations, acquiring approval from the local EC and data protection office. The studies were designated as "miscellaneous", that is, neither medical drug nor medical device trials. A detailed participant information sheet was given to each patient included in the study, and each provided written informed consent.

Study Population

Selecting the study population to appropriately address the research question is crucial in study design. In particular, it is important to match the subject characteristics, epilepsy type, or seizure semiology in the study cohort to the goals of the study. In these prospective cohort studies patients with a diagnosis of epilepsy were recruited when they presented for epilepsy care at the hospital epilepsy monitoring unit (EMU), and were asked to wear one or more wearable devices. While in the EMU, patients were monitored for seizures via vEEG, which was recorded along with the sensor data from the wearables.

As physiological responses and signal alterations during epileptic seizures may vary across age groups, the inclusion of participants of different ages needs to be taken into account. In the studies included here, the lower and upper limit of age for study participants ranged between 7 years and 80 years, respectively. Moreover, at the stage of protocol development, it is important to identify those comorbidities that may interfere with study adherence or with data collection and quality. Patients with conditions impeding the ability to participate (cognitive, psychiatric, acutely ill), to wear the device (skin conditions), or with frequent vigorous involuntary movements (e.g., chorea, athetosis) were excluded from the studies.

Baseline characteristics of the included participants allow the population under study to be better characterized, the results obtained to be understood and contextualized, and for generalizability of the data to be discussed. For all the study participants, data collected during the study period included basic demographic characteristics including age and gender, clinical information, and seizure characteristics including etiology, localization, type, onset, and frequency of seizures and medications.

During the studies presented in this thesis, patients with epilepsy were recruited at the EMUs of the King's College Hospital London (KCL) and the University Medical Center Freiburg (UKF). Table 3.1 gives an overview of some of the relevant information and key demographics regarding these data collection procedures [174].

3.1.2 Data Collection and Technical Infrastructure

Video EEG recordings and seizure annotation

Recording data continuously over days with the support of vEEG is essential to capture an adequate number of events and to reliably identify and characterize seizures through a gold standard.

In the studies included here, as part of the clinical workup, patients were admitted to the EMU and connected via scalp electrodes to an EEG monitoring system within view of a video camera. The length of stay in the EMU varied based on the patient's clinical care. The majority

Table 3.1: Information on data collection parameters and demographics of enrolled participants with recorded Empatica E4 data, at the KCL and the UKF.

	KCL	**UKF**
EC approval	16/LO/2209	538/16
enrollment period	Jun. 2017 - Aug. 2019	Jul. 2017 - Mar. 2020
# of participants	29	172
recruitment age range	18 - 80	7 - 80
actual age range	38.0	30.0
(median, [95 % CI])	[20.4, 63.8]	[14.7, 64.0]
sex (% female)	48.3 %	47.1 %

of adult patients were admitted for a 5- to 10-day stay, with overall shorter durations for children.

Trained personnel are needed to perform standard vEEG monitoring, including electrode placement according to the 10-20 international system, and to maintain high-quality recordings. EEG recordings were fully reviewed, and seizure onset and offset were annotated, in addition to supporting information including seizure semiology and ictal focus. The two centers where data were recorded, UKF and KCL, jointly developed and adhered to a review and annotation protocol specifying reviewing terminology and methodology to guarantee consistency in reporting clinical phenomena across patients. This included, for example, definitions of autonomic features such as tachycardia, which is ambiguously defined in epilepsy-related literature; determination of duration of impaired awareness, which is not always actively tested for; and an agreement on how to consistently store this information in a shared database for collaboration. The labeled vEEG recordings were then transferred to a secure server for storage and analysis, and seizure onset and offset times were applied to the simultaneously collected wearable recordings.

Wearable data collection and device integration

Data collection with wearables is generally done in one of two approaches: offline collection, where the data are stored locally on the device and then downloaded at a later time, or online collection, where the data are streamed continually via a wireless connection to an external device.

During online collection, the wearable device usually has a much shorter battery life, since wireless data transmission adds substantially to the overall energy consumption. However, the maximum recording time in offline collection is constrained by the internal storage capacity of the device. Furthermore, the data must be manually downloaded from the device, potentially requiring regular patient participation.

This process can be automated during online collection, at the expense of a potential for data loss due to connection problems. An added benefit to data streaming is the possibility of live data processing and visualization, allowing caretakers and study personnel to evaluate data as they come in. Live data streaming is also a key requirement for any intervention or alarm system not directly built into the wearable device.

For the studies included here, performed at the epilepsy wards of KCL and UKF, online data streaming was used. The wearable devices were constantly connected to a companion device

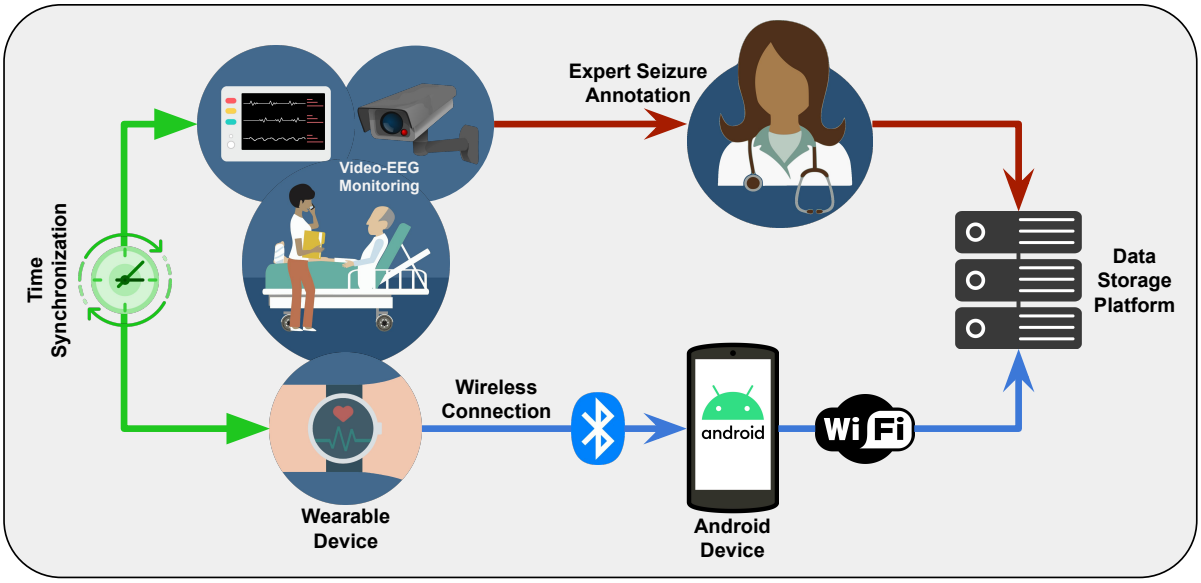


Figure 3.1: Setup of the technical environment for in-hospital recording studies with wearables for seizure detection in an EMU. Remaster of Figure 2 in Bruno et al. [103].

via Bluetooth, and a custom-built Android application was used to receive the raw data directly from the wearable device and upload them to a data storage server on the clinic premises (Figure 3.1). All components of this system like the Android app and the server framework are open-source software available on GitHub [291]. Wearable devices were exchanged twice per day, in the morning and evening, to allow for battery charging given the shorter battery life in streaming mode. There were also frequent problems with the devices' Bluetooth connectivity. The wearables often disconnected from the companion device, either due to the patient walking out of range or due to other, sometimes unexplained reasons. This would lead to frequent and extensive data loss (see Section 3.1.4, data quality and completeness), especially since the wearable device did not offer an on-device data buffer or automatic reconnect to the companion device.

Synchronization between wearable and vEEG data

Time synchronization between an external device and the vEEG is particularly important in the field of epilepsy research. The clinical seizure onset, used as the ground truth in developing models for seizure detection and prediction, can often be pinpointed with sub-second precision by clinical experts. Thus synchronizing the internal time of the wearable device to the time of the vEEG system is essential for data analysis. Furthermore, depending on the specific device used, internal inaccuracies can cause small shifts in the timekeeping between individual biosignal data streams.

There are two principal ways of achieving synchronization between a wearable device and a vEEG system. The most accurate and technically more advanced way is to directly and precisely adjust the on-device timekeeping of the wearables to the time used in the clinical vEEG system, for example, by some wireless connection. This will give millisecond synchronization between the two time bases, but may require some technical set-up beforehand, and it might not even be available as an option if the wearable device does not support this operation. The

second way of achieving synchronization is through the study staff, who can manually induce a visible and recognizable change in the wearable's recorded signals while also showing this action on the video or EEG signal. Alternatively, an artifact or label can be placed simultaneously during the device and EEG recording, and then be confirmed by EEG, as some standard vEEG systems suffer from an occasional minor desynchronization of the EEG and video. The data streams can then be synchronized retrospectively by adjusting the wearable data timestamps to align the events with the vEEG. Although the data streams can be synchronized to sub-second precision with this method, it requires manual modification of the data.

Both methods are susceptible to the internal drift of timekeeping in the wearable, caused by inaccuracies in the real-time-clock circuits in these devices. This drift can accumulate over time, up to several seconds of inaccuracy over several hours of recording. Therefore, it is advisable to repeat the synchronization process periodically during the recording. The automated method is more suitable for this, as the synchronization could be triggered, for example, every few minutes. Another method to deal with drifting timestamps directly is to measure individual calibration parameters for each device that is used in a study. Thereby, the precise sampling rate for a device is found by a calibration procedure, to a degree of accuracy that allows for a later recalibration of the timestamps in the recorded signals. Synchronizing the wearable data with the vEEG system can also enable integration of both into a common data viewer, which facilitates a better understanding of abstract wearable data in the context of the actual clinical setting.

In the studies included here, the automated method was employed. Because the devices are programmed to synchronize themselves to the clock of the companion Android device whenever they are first connected via Bluetooth, it is only necessary to synchronize the Android devices to the vEEG time base, which can be done easily via a network connection. Consequently, each center synchronized their wearable devices each time they were exchanged for battery charging, with intervals ranging from twice per day to every 2 days. Section 3.2 gives a more detailed description of the wearable device and technical infrastructure implemented for the studies.

3.1.3 Devices

Across the two study sites, KCL and UKF, several different wearable devices were used for data collection from study participants. Among the most prominent devices were *Biovotion's Everion*, *IMEC's sensor bracelet*, *Epitel's Epilog*, *Byteflies' Dots*, and *Empatica's E4*. The data quality and patient acceptance of some of these devices have been reported previously [159, 173, 292]. In the studies included in this thesis, only data from the Empatica E4 device (Section 3.2.1; Empatica Inc., Boston, MA, USA) were used for further analysis, as it was used in the overall largest amount of participants.

Wearable devices of the types used in clinical epilepsy studies can be categorized in various ways, all of which should factor into the decision when selecting a device for a study:

Medical certification: Wearables, in general, are employed in many different fields beyond medicine, so for use in studies as described here, the certification as a medical device can be an important factor. IMEC's sensor bracelet for example, as a prototype device, is not independently certified, whereas the Empatica E4 has a European CE class IIa certification as a medical device.

Modalities: Different devices record different biosignals at different sample rates, so an informed decision needs to be made about exactly what is needed to facilitate the outcomes of the given study. Multimodal devices are generally regarded as more effective and versatile [293–295], whereas a device recording only one modality may be sufficient for a very specific task. Epitel's Epilog, for example, provides only a single-channel EEG signal, whereas the Empatica E4 records four different types of raw biosignal data.

Data mode: Generally, there are two modes in data collection, online or offline, as described further in Section 3.1.2. In most cases a given device supports only one mode for recording data, so either the study protocol needs to be adjusted to support the device, or an appropriate device needs to be chosen for an already established study protocol. The online streaming mode is a requirement for systems that should include any kind of alarm or intervention. Byteflies' Dots, for example, support only offline recordings, whereas the Empatica E4 has the option to employ both methods.

Battery life: With current battery technology, the battery life of smaller devices or those that employ online raw data streaming is usually measured in hours, whereas somewhat larger devices with offline, on-device data storage can sometimes be active for days without the need to recharge. IMEC's sensor bracelet, for example, has a typical battery life of seven days, while the Empatica E4 has a manufacturer-specified battery life of $24\,\mathrm{h}$ to $48\,\mathrm{h}$, although in these studies empty batteries after half that time were often observed. This was in part due to the shorter battery lifespan when the E4 is used in streaming mode.

Device placement: Wearables are usually placed at a specific part of the body, which can be influenced by the study protocol and should be considered when choosing a device. In turn, the placement of the device may affect both the sensitivity and specificity of a prospective seizure detector. The Empatica E4, for example, is normally worn around the wrist, while the Byteflies' Dots can be attached to any part of the body by use of an adhesive patch.

Furthermore, research-grade devices, such as the Empatica E4, often have several advantages and disadvantages over other devices that are marketed directly to consumer end-users. Access to raw data is a necessity for many research studies, but something that consumer-grade devices and services rarely provide. Furthermore, companies offering research-grade devices are sometimes open to collaboration, for example, by supporting researchers with specialized knowledge of device capabilities.

On the other hand, research devices are often more expensive than their consumer counterparts and can be more cumbersome and uncomfortable to wear, since the device's aesthetic design is not a priority for the manufacturer. However, patients consistently gave more positive feedback on the wearability of the Empatica E4, as compared to the Biovotion Everion, which is a device on the market for regular consumers to buy [292].

3.1.4 Reporting Results

Usability challenges and users' perspectives

Wearable devices are progressively becoming an available and innovative tool for continuous seizure monitoring. People living with epilepsy have expressed interest in using new tech-

nologies in their daily life [125] and several unmet needs might be addressed by adopting digital solutions into health care services [125, 296]. The research focused on hypothetical scenarios has highlighted that motivation to use wearables is not driven only by the accuracy and reliability of the device performance. A design incorporating comfort and ease of use is also essential for acceptance and long-term adoption [292]. Obtaining feedback from patients after direct experience wearing devices is the only way to fully understand the practical and technical issues faced [297]. However, feedback on device comfort and usability has been collected only sporadically in previous studies, and information reflecting the direct experience of study participants is missing. The limited number of investigations exploring users' direct experience reported improvement of quality of life for both patients and caregivers [297], a benefit to autonomy and increasing independence in activities [226, 297], as well as a generally good evaluation of technology usability [226, 298]. Barriers to use, as reported, include discomfort in wearing the device during sleep, technical difficulties, and the burden of adding another aspect to routine epilepsy care [297].

In addition to the key requirements of a reliable and accurate performance, a successful integration of digital solutions into a patient pathway requires acceptance of the technology. The latter is required for long-term engagement, which is essential to a good detection performance, and to optimize the benefit to the patient. To identify and avoid potential barriers to a long-term engagement with the technology, patients' views and requirements need to inform the development of the technology and study design, and users' opinions on usability and acceptability should be collected systematically. Methods to obtain feedback from study participants range from a focus group (useful during the first stages to guide research questions and research development), interviews (at set time points during the study, for example, study end or in case of participants' withdrawal), collection of participants' observations (any time in the course of the study), and questionnaires (allowing direct comparisons between subjects and the identification of subject-related factors influencing their experience in the study). At KCL and UKF, participants' experience and the perceived ease of use and comfort of the technology were assessed at the end of the study using a self-administered Technology Acceptance Model Fast Form (TAM-FF) [299]. Moreover, in a group of study participants, the experience of wearing multimodal sensor devices was also assessed via semi-structured interviews covering questions on their experiences and concerns using the wearables, their thoughts about ambulatory use of wearables, and their reasons for stopping to wear the device if applicable [72].

Data quality and completeness

The value of collected data can be assessed by data completeness and data quality. Data quality measures evaluate properties like the noisiness, accuracy, and potential information gain of the data, whereas data completeness gauges data loss during recording. In the context of exploratory research, both data quality and completeness are of utmost importance, and several steps were taken to reflect that need. Collecting raw, unprocessed data from wearable devices, forgoing any internal processing, can facilitate the assessment of data quality. This will give a complete and clear picture of the suitability of the device for the task at hand. Furthermore, sharing raw data across different research sites and groups can enhance the value of the data set, advance the understanding of data complexities, and facilitate scientific exploration of the data. Another important tool to effectively assess data value is the use of data dashboards. These dashboards usually take the form of a website that aggregates data completeness and quality measures as new data come in and displays it with intuitive charts and

tables. Especially in the context of live data streaming, they can monitor system function and user adherence.

Gaps in the data can be caused by several issues related to data collection. A common cause of data loss is the limited battery life of the device. Charging the battery takes time (typically hours), and even if a second device is used to replace the one with an empty battery, this creates a small but noticeable gap in the recording.

Another common source of data loss is connection problems with wireless data streaming. With a Bluetooth connection, the maximum range between the wearable and its companion device is usually $10\,\mathrm{m}$ within the same room. Whenever a wearable device is disconnected, it needs to automatically reconnect and transfer any buffered data, otherwise, any data collected while the device is disconnected will be lost. The Empatica E4 device used in the studies included here does not implement such functionality in Bluetooth streaming mode. When this device loses its Bluetooth connection it powers off completely and must be manually restarted for the connection to be re-established, leading to significant data loss.

Finally, data gaps can be introduced by human interaction. Taking the device off for a short time, for example, during a shower or neuroimaging causes several minutes of data loss. Incorrect operation of the device can also lead to lost data. Some of these causes for data incompleteness can be avoided, for example, by the careful preparation of a study protocol detailing proper usage of the device. Others are inevitable, and some gaps in the data set are unavoidable.

Data coverage was determined in two different categories: the overall data coverage and the number of missed seizures during each patient's recording. Data coverage is computed by counting the number of samples per modality collected from the wearable device, per patient, and dividing by the number of expected samples given the recording time. This method potentially undercounts the data loss because it ignores any loss when the device is not worn. The same methodology is applied to counting missed seizures, that is, only seizures that happened within the start and end of the recording are counted toward the expected amount.

Among patients who wore the Empatica E4 device in the UKF and KCL sites, the average data coverage over all participants was $52\,\%$ and $40\,\%$, respectively, with the loss of data attributed in large part to the live data streaming functionality, but it was also affected by device recharging and the patients bathing during their in-hospital stay. In a different study which specifically and more extensively investigated data and signal quality of wearable device data, the data coverage was reported as $54.2\,\%$ and $51.5\,\%$ for the same clinical sites UKF and KCL, respectively, including similar patient cohorts and data sets [174].

Figure 3.2 highlights data completeness considerations for a patient in the UKF cohort. Two gaps in the data as well as missing seizures can be seen in this example. The recording for this patient is missing approximately $30\,\%$ of its expected data, and 17 of 33 seizures ($52\,\%$) were missed as a result.

Data quality is an important property for any scientific data set. The quality of data collected from wearable devices can be degraded by several issues related to the sensor hardware and application. Any physical sensor has mechanical or electrical imperfections that can produce sensor noise. Imperfections can also be caused by external stimuli introducing an unwanted variation of the data, a so-called artifact. These artifacts can sometimes be corrected after data collection, but other times completely disrupt the underlying data. A relevant example is motion artifacts in the photoplethysmography (PPG) data collected from the Empatica E4 device. A PPG sensor works by measuring the light reflection of the skin, which changes

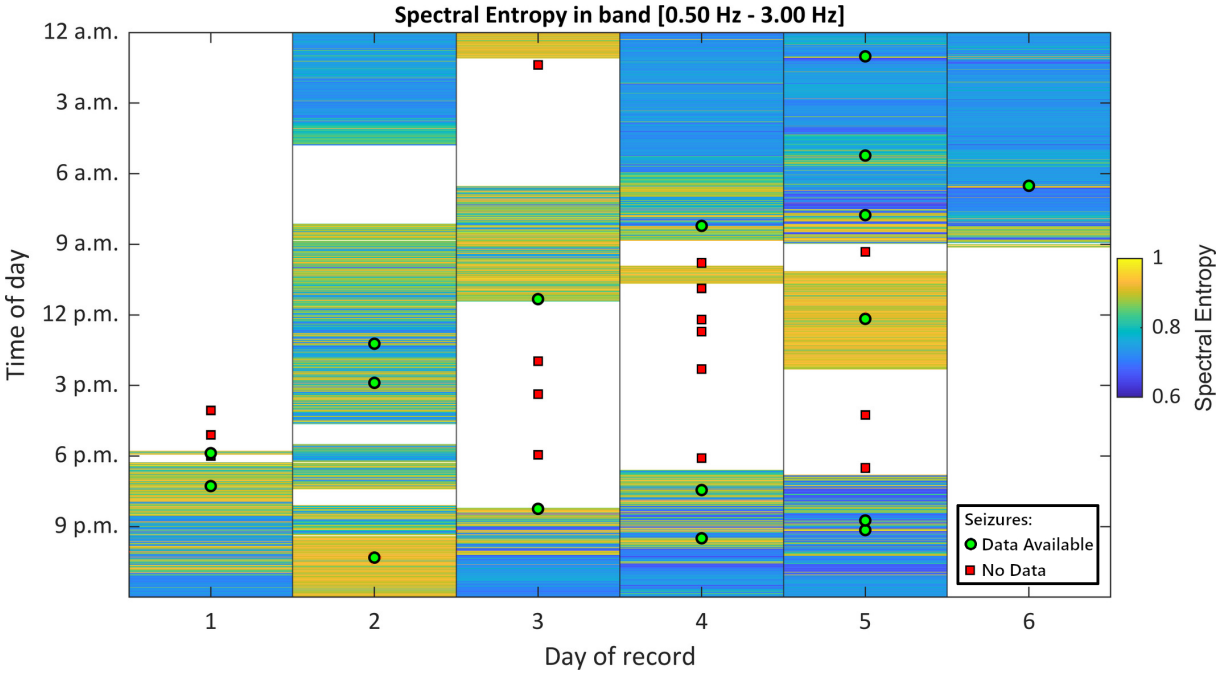


Figure 3.2: Spectral entropy of the blood volume pulse signal of the E4 device during the recording of a single patient recruited at the UKF site. The signal gives an idea of the quality of the blood volume pulse (BVP) data for HR calculation; low spectral entropy (blue) indicates a higher quality signal with fewer artifacts. The gaps show times when there was a problem with the recording and no BVP signal was present. The green circles mark seizures during which wearable data were recorded; the red squares mark seizures where no data were available. Remaster of Figure 3 in Bruno et al. [103].

with blood volume, that is, with each pulse. However, light from an external source, for example, sunlight, can compromise the reflective value measured by the photodiode of the sensor. If the device was not tightly fastened around the wrist, the actual BVP data for that segment are not recoverable. This can be a considerable problem for data collection during physical activity or during convulsive seizures. Another source of poor data quality is inaccuracies introduced by the sensor, for example, caused by faulty or deteriorated hardware.

To measure data quality, numerous methods can be found in existing literature, and are usually specific to a certain sensor modality [159, 212, 257, 300, 301]. Figure 3.2 shows a plot of the spectral entropy calculated from the Empatica E4's BVP signal collected from a single patient at the UKF site, as an example of a data quality measure applied to the wearable device recordings. Spectral entropy gives an idea of the quality of the BVP data for HR calculation. Lower values mean the signal is of higher quality, that is, contains fewer artifacts. The signal quality is generally higher overnight when patients rest. During the daytime, patients tend to be more active, and the signal is prone to movement artifacts, represented by higher values in spectral entropy.

Seizure detection evaluation

The common goal of most epilepsy-related studies with wearable devices is to achieve robust seizure detection and prediction. Reporting results of evaluations of these methodologies is

an important part of any study and should follow a defined protocol and refer to specific standards [212]. Sensitivity and specificity are the two cornerstones of reporting results of binary classification, especially in a medical context. Sensitivity, also often called recall in a machine learning context, measures the proportion of true positives (TPs) to all expected positive instances. It must always be reported as a study outcome, because for seizure detection it directly describes the respective methodology's ability to robustly detect seizures from the wearable data. On the other hand, specificity measures the proportion of true negatives to all expected negative instances. To report measures like specificity based on negative instances in the context of wearable seizure detection, the data stream must be segmented into equal-sized portions of either the seizure or non-seizure class. Due to the large data imbalance of these two classes that is usually observed in epilepsy studies, with sometimes multiple days of non-seizure portions in the data interrupted only by often minute-long seizure portions, the specificity measure is artificially boosted to consistently report values of, for example, $>98\,\%$, even if there are many false positives (FPs). Because of this lack of informative value, specificity is often omitted when reporting on the performance of a seizure detection system. Instead, the false alarm rate (FAR) or positive predictive value (PPV) can be reported as inverse measures of a seizure detection system's ability to correctly identify non-seizure periods. The FAR reports the number of false detections over a certain timespan, often chosen as a day ($24\,\mathrm{h}$). For example, a FAR of 0.5 per $24\,\mathrm{h}$ would mean that the system, on average, produces one false alarm every 2 days. FAR can also be separately reported for daytime and nighttime periods, as false nocturnal alarms may be much more disrupting and less acceptable to patients and caregivers. The PPV, also often called precision in a machine learning context, is the proportion of TPs to all detected positives. It thus gives a measure of the number of FPs to TPs, for example, a PPV of $50\,\%$ would describe a result of the same number of FPs as there are TPs. At least one of these measures, FAR and PPV, must always be reported as a study outcome, to properly convey the number of FPs a system is likely to produce. One possible way to counteract false alarms could be to ask patients to perform specific periodic movements like brushing their teeth. These movements, recorded by the wearable, could then be used to adjust a model to be more robust against non-epileptic activities of daily living.

To visualize the results of an evaluation of a seizure detection model, or to compare the performance of multiple models, the receiver operating characteristic curve is a widely used and accepted tool. It plots the probability of detection against the probability of false alarm (FP and TP rates) of a binary classifier at varied discrimination thresholds. Thereby, it visualizes the trade-off a model makes between detecting true events and producing false alarms.

For all of these measures, there is generally a trade-off between reporting them on a per-patient basis and taking the mean across patients or reporting the overall value over the whole applicable data set. Optimally, both aggregations should be reported in the outcomes of a study, as they often both provide slightly different but equally worthwhile conclusions.

3.1.5 Data Sharing

Free and open platforms for sharing data and facilitating collaboration are important research resources. Open databases (from which data can be explored and downloaded), and novel algorithms and source code (that can be shared between collaborators) are important tools in neuroscience projects. Different examples can be cited, including openneuro.org, epilepsyecosystem.org, ieeg.org, and physionet.org. Research teams should be encouraged to share raw

data and data processing scripts to allow replication and validation of results. Online competitions have also been successful at fostering the development of high-performance seizure detection and forecasting algorithms based on intracranial EEG [238, 253, 254], and similar results with wearable data could be expected. Moreover, sharing data, methodologies, and results with partner organizations, like other clinical centers or even device manufacturers, can be greatly beneficial to the advancement not only of the research field of wearable seizure detection in general but also the usability and development of new devices and technologies. This includes the sharing of raw data collected during studies, as well as any scripts and software used in the processing and scientific analysis of the data, especially concerning seizure detection. To facilitate data sharing, a standardized data format and schema should be adopted to prevent the use of different and potentially not compatible formats. This would promote the replication and validation of results in a collaborative manner and encourage the aggregation of data across research groups. In the long run, giving valuable and constructive feedback on device performance and usability to manufacturers, and sharing these experiences with other organizations, could be a huge boon to possibilities in the treatment of epilepsy, and patients with epilepsy by extension. To accommodate and facilitate the aforementioned sharing of data and experiences, however, a need for open and structured systems and forums exists. Here, clinicians, researchers, developers, manufacturers, as well as patients could collaborate and contribute to the advancement of the treatment of epilepsy with the use of wearable devices. And although strict data protection rules like the EU GDPR may hinder collaboration, these restrictions do not negate the major benefits of sharing pseudonymized or anonymized data for research progress and patient care.

3.2 Data Collection and Preprocessing

In this section, further details on data collection procedures and data preprocessing are compiled. This section is not part of the publication included in Section 3.1. The device used most prominently in the data collection studies at both clinical centers, the University Medical Center Freiburg and the King's College Hospital London, and which was selected for the data analysis and evaluation studies included int his thesis, was the Empatica E4 (Empatica Inc., Boston, MA, USA). Section 3.1.2 and Section 3.1.3 already highlight some information on these topics, which is extended here by more technical detail.

3.2.1 Empatica E4 Wearable Device

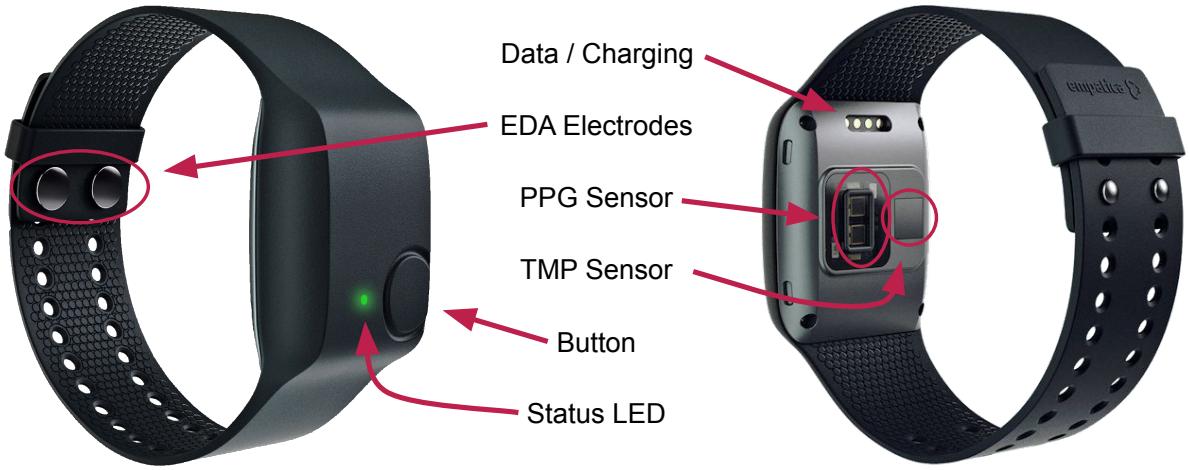


Figure 3.3: The Empatica E4 wrist-worn wearable device. **(left)** Front view, including the status light-emitting diode (LED) and button in the foreground and the two electrodermal activity (EDA) electrodes in the background, attached to the wrist strap. **(right)** Back view, including the photoplethysmography (PPG) and skin temperature (TMP) sensors at the backside of the device body, as well as a data and charging connector. Source: empatica.com, own labeling.

The *Empatica E4* wearable device, developed and manufactured by *Empatica Inc., Boston, MA, USA*, is a wrist-worn, research-grade monitoring device with a variety of integrated biosignal sensors. It was independently certified by a notified body of the European Union as a Conformité Européenne (CE) class IIa medical device. Figure 3.3 gives a high-level overview of the device and its major components, and Table 3.2 compiles the key technical specifications of the device.

The Empatica E4 integrates sensors recording ACC, EDA, PPG, and TMP biosignals. ACC raw data are recorded at a sample rate of $32\,\mathrm{Hz}$ with a sensitivity range of $-2\,\mathrm{g}$ to $2\,\mathrm{g}$. Ranges of $\pm 4\,\mathrm{g}$ and $\pm 8\,\mathrm{g}$ can be set with custom firmware in agreement with the manufacturer, but this was not done for any of the data collected in the studies included in this thesis. The EDA signals are recorded at a sample rate of $4\,\mathrm{Hz}$ with a measurement range of $0.01\,\mu\mathrm{S}$ to $100\,\mu\mathrm{S}$, using the exosomatic methodology with an alternating current source (see Section 2.2.2). The silver electrodes are integrated into the wristband and placed on the inner wrist. The PPG signals are captured at $64\,\mathrm{Hz}$, with an array of two green and two red LEDs, and two photodiodes.

Table 3.2: Technical specifications of the Empatica E4 device.

	Empatica E4
Manufacturer	Empatica Inc., Boston, MA, USA
Certification	CE class IIa medical device
Body Position	commonly wrist, ankle possible
Biosignals	ACC, EDA, PPG, TMP
Sampling Rates	32 Hz, 4 Hz, 64 Hz, 4 Hz
Measurement Ranges	$-2\,\mathrm{g}$ to $2\,\mathrm{g}$, $0.01\,\mu\mathrm{S}$ to $100\,\mu\mathrm{S}$, N/A[1], $-40\,^{\circ}\mathrm{C}$ to $115\,^{\circ}\mathrm{C}$
Battery Life	$12\,\mathrm{h}$ to $48\,\mathrm{h}$, dependent on recording mode and battery
Recording Mode	on-device storage[2] or Bluetooth streaming[3]
Other Features	IP22[4] drip-resistant; status LED; event marker button

[1] The blood volume pulse (BVP) output data are given in arbitrary units and has no relevant measurement range. [2] Storage capacity of up to $60\,\mathrm{h}$. [3] Wireless connection range of up to $10\,\mathrm{m}$. [4] Ingress protection code (International Electrotechnical Commission standard): protected against ingress of solid foreign objects of $\geq 12.5\,\mathrm{mm}$ diameter (e.g., fingers) and dripping water at angles of $\leq 15°$.

The device output is BVP data, computed by a proprietary algorithm from both green and red light reflectance values, and given at the same sample rate. Finally, the TMP data are sampled at $4\,\mathrm{Hz}$ with a measurement range of $-40\,^{\circ}\mathrm{C}$ to $115\,^{\circ}\mathrm{C}$.

The Empatica E4 supports two data acquisition modes: a local storage recording mode and a Bluetooth streaming mode. The two modes are mutually exclusive; If the device is connected to a Bluetooth device no data will be recorded on the internal storage. In local storage mode, the data are stored on an internal device memory and will be transferred to a Windows or Mac application running on an external computer whenever the E4 is connected to it by Universal Serial Bus cable. The raw data are however not directly accessible at this stage, and are transferred by internet connection to cloud-storage servers operated by the device manufacturer, to be viewed and downloaded via an online data dashboard.

Alternatively, the device can be connected via Bluetooth to an app running on a mobile device, which in turn reads previously stored or currently streaming data and uploads them to the same manufacturer-controlled servers. In addition, software development kits (SDKs) are available for the Android and iOS platforms to develop a custom data streaming application. For the realization of the studies such an application was used, avoiding the usage of third-party cloud storage (see Section 3.2.2). While in streaming mode, the Empatica E4 does not employ a data buffer, such that any data that are not transmitted to the companion device will be lost. Furthermore, it does not feature an automatic reconnect functionality, rather powering off once the Bluetooth connection has been severed.

In streaming mode, which was used for all the data collection included here, the device has a battery life of up to $24\,\mathrm{h}$, depending on the quality of the battery. The device has one physical button, serving as an on/off/reset switch as well as an event marker button, which was however not used in the data collection procedures. A status LED at the front of the device notifies the user of the current data acquisition mode as well as a low battery, a full internal storage, or other potential issues. The Empatica E4 is rated as IP22 drip-resistant, meaning it can withstand some light water dripping vertically or at a slight angle on the device, but it should not be used, for example, in the shower or submerged in water.

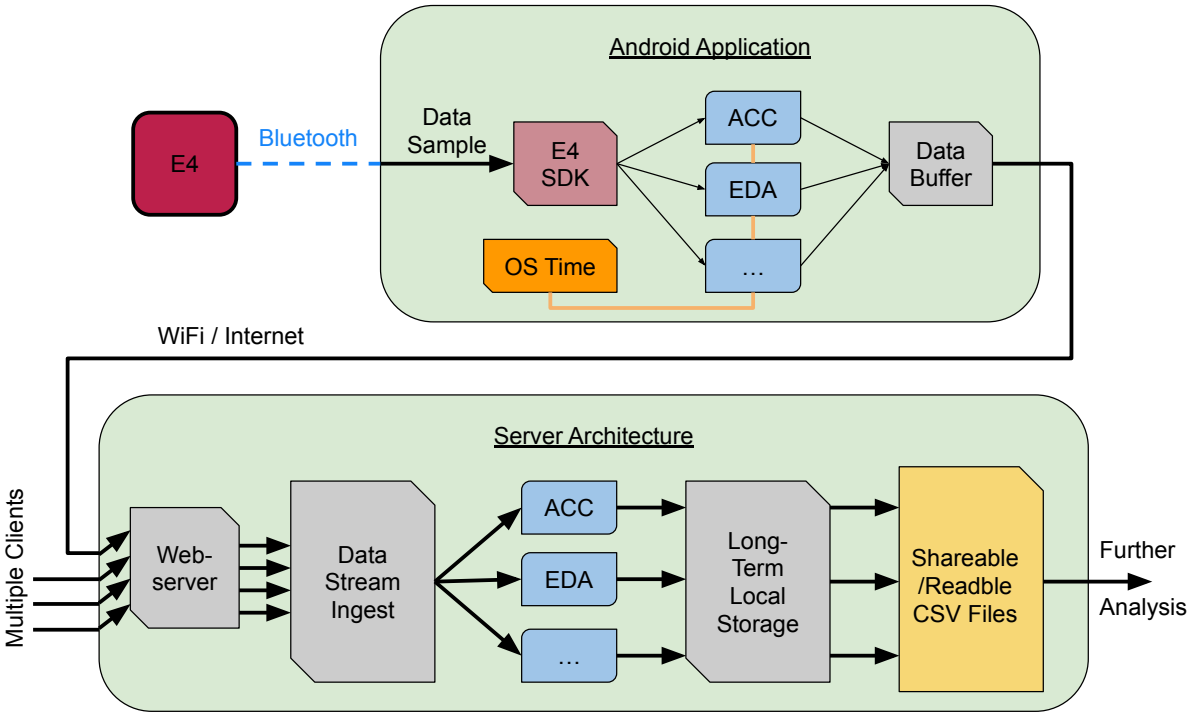


Figure 3.4: High-level overview of the complete data collection pipeline, from Empatica E4 wearable device to raw data CSV files that can be used in further data analysis.

3.2.2 Implementation of Data Collection

In conjunction with the centerpiece of the data collection, the *Empatica E4* wearable device, several other systems were deployed to facilitate biosignal data acquisition[1] [291, 302]. As a direct companion to the wearable, an Android device[2] was used to run a custom open-source application that connects to the wearable by Bluetooth. Furthermore, a data ingest and processing infrastructure was deployed on a server machine, handling the data collected and sent by the app via standard wireless network connection. The raw data were then stored on a network-attached storage server in human-readable, comma-separated text files for later processing and analysis. Figure 3.4 gives an overview of the full data collection pipeline, as also further described in the following.

The Android app uses a SDK provided by the manufacturer to connect to the E4, receive new raw data, and decode them from a binary format to appropriate floating-point numbers. The Empatica E4 sends data in batches, which the app receives together with a time stamp per sample that derives from the wearable's internal real-time clock. This clock is updated to the current time of the Android companion device only at the start of each connection session. The app then collects all of these samples of raw data, grouped by biosignal modality, supplementing each with a secondary time stamp of the Android operating system (OS) time at the moment of receiving. This OS-specific time is regularly synchronized with the internal time of the video-electroencephalography (vEEG) system of the epilepsy monitoring unit. The app then regularly sends the collected data to an ingest-, processing-, and storage server.

[1] radar-base.org

[2] *UDOO NEO* all-in-one mini computer, or *Samsung Galaxy Tab A (SM-T580)* tablet

The server application is divided into multiple consecutive parts. A web server handles all incoming requests or data packets (sent over the internet via HTTPS), forwarding them to the correct component. A distributed stream processing platform (*Apache Kafka*) ingests new data packets on a variety of data topics, that is, it receives data in batches and grouped by biosignal modality, serializes and schematizes that data, and stores it for the short-term. Thereby, it can efficiently handle multiple clients sending data in parallel and in real-time while keeping references to client identifier, biosignal modality, and other metadata, per sample. In a next step, the received and processed data are transferred to a more long-term storage component (*Apache Hadoop distributed file system (HDFS)*), saving the data on disk and easily accessible to other processes on the server. In order to facilitate data analysis and sharing, in a final step the data are encoded into human-readable comma-separated values files, which can be copied to other computers and easily read into standard data analysis software.

3.2.3 Data Preprocessing

All data processing and analysis performed in the studies included in this thesis was done using *MATLAB*, developed by *MathWorks Inc, Natick, MA, USA*, in different software versions depending on the most recent release version available at the time. After raw sensor data were collected from study participants, several preprocessing procedures were applied in order to prepare the data for further analysis, for example, feature extraction (see also Section 1.3.3). The specific actions relevant to the analysis performed here were deduplication of data samples, correction of timestamp drift, and discovery of viable data blocks.

The timestamps attached to each sample of wearable data, both the native E4 stamp and the Android OS time, are given as Unix time[1], that is, the number of seconds elapsed since `1970-01-01 00:00:00` UTC. Here, the timestamps are numbers in floating-point format at millisecond precision. The raw Empatica E4 data can sometimes contain two data samples with the exact same timestamp, due to unknown device-internal reasons. Data analysis methodologies generally expect time series data to be unique with respect to sample timestamps. Thus, immediately after reading the data into memory, samples with the same timestamp as the one before it were removed.

As outlined in Section 3.1.2, inaccuracies in the real-time-clock circuits of wearable devices may lead to some differences between the device time and the real time (or vEEG time). The two times may drift apart by several seconds over multiple hours of recording, and the rate of drift is not necessarily linear and not the same for different E4 devices. Furthermore, since the E4 device does not allow a regular automatic synchronization of the internal device time with some external time source, the recorded timestamps needed to be corrected retroactively. To this end, the timestamp drift was estimated to be linear over short timespans and between points of greater than normal change in time. The drift itself was calculated by taking the difference between the Empatica E4 and Android OS reference time, which was regularly synchronized with the internal vEEG time. The estimated short-term linear segments of the drift were then subtracted from the same, resulting in new timestamps for each sample that are in line with the reference time, within a margin for error of less than one second. Figure 3.5 gives an example of this procedure for a multi-day Empatica E4 recording, including several disconnects and reconnects showing as jumps of the drift to zero due to the one-time

[1] Also known as “POSIX time” or “Epoch time”

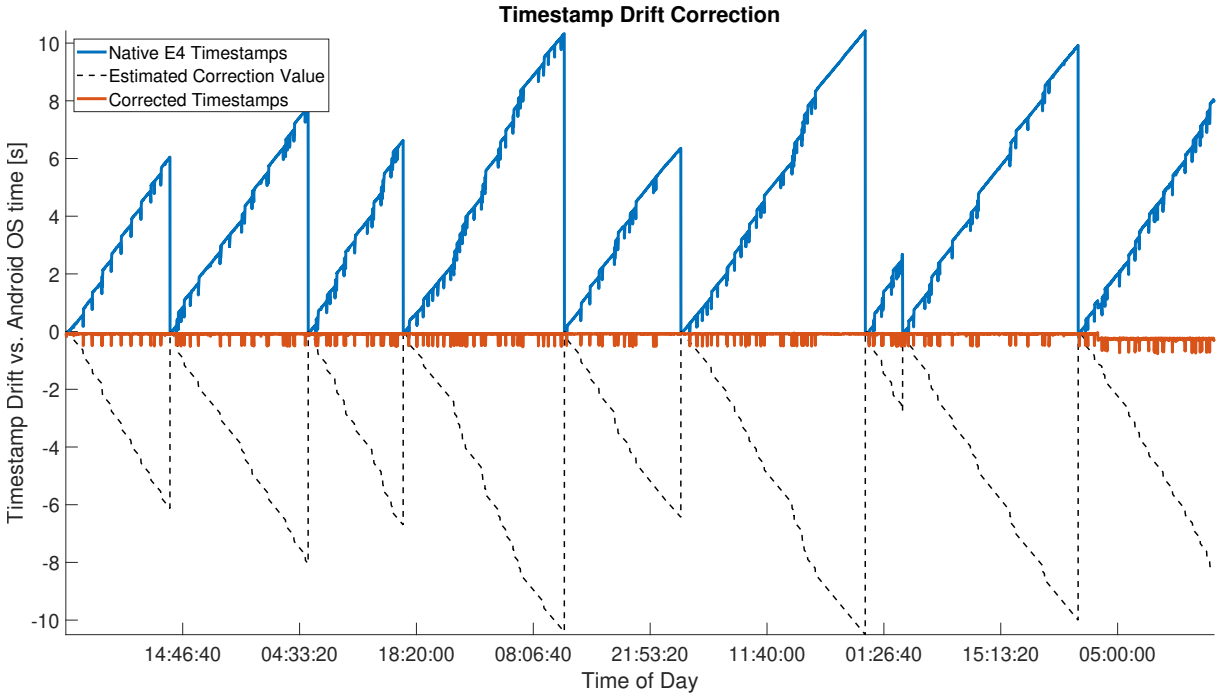


Figure 3.5: Example of the timestamp drift correction for multi-day E4 data. Shown are the increasing drift of native E4 time vs. the Android OS reference time (blue); The estimated correction applied to the native E4 stamps (dashed); And the difference between the corrected and the reference time (red). The visible reoccurring small changes in drift originate in the reference time and are not present in the native E4 or corrected time. Points in time when the drift reverts to zero are related to the E4 device having been reconnected to the Android companion device, triggering a one-time automatic update of the internal device time.

automatic synchronization with the Android companion device whenever the E4 connected to it.

The seizure detection methodologies and machine learning algorithms used in the analysis rely on gap-less data for all included biosignal modalities. However, the Empatica E4 does not reliably provide data from all sensors at all times during the recording. In fact, the raw data set contains several sometimes hour-long periods where data of one modality is present, but not of another. These gaps were furthermore found to always be larger than one second. Thus, to facilitate efficient analysis, gap-less data blocks including full data from all three relevant biosignal modalities (ACC, EDA, BVP) were discovered. To this end, the raw data were traversed in one-second steps, and those parts that did not contain at least one sample from each modality within one second were disregarded.

3.3 Summary

This chapter highlighted clinical study design principles, data collection procedures, and considerations concerning the data quality and performance evaluation. Methodologies with respect to feature extraction and seizure detection were distinctive per analysis, and are further explained in each respective part of the thesis. The data collection studies referenced in this thesis were conducted at two European epilepsy centers, the University Medical Center Freiburg and the King's College Hospital London.

Study aims for wearable device studies in epilepsy inform aspects like patient selection, choice of device, and data analysis. Specifically, the seizure types targeted for detection are the decisive element. For example, seizure types that do not predominantly include movement of the limbs as a symptom may require different biosignal sensors to collect meaningful data for seizure detection, as compared to convulsive seizures. Here, convulsive tonic-clonic seizures and focal seizures with motor components were the target of seizure detection, and as such the *Empatica E4* wearable device was considered to be the best choice. Furthermore, patients with known motor seizures of any kind were included as participants in the studies, whereas patients with a history of only non-motor semiologies were not included in the subsequent data analyses.

The Empatica E4 is a research-grade wrist-worn wearable certified as a medical device for epilepsy monitoring, and provides access to the raw data of the biosignal modalities accelerometry, electrodermal activity, and blood volume pulse (BVP), which were used in the seizure detection methodologies presented here. Study participants also underwent video-electroencephalography (vEEG) monitoring at the two epilepsy centers the studies were conducted at. Epileptologists created expert annotations of seizure onset and offset times as well as type and included semiologies, serving as a ground truth for the evaluation of the seizure detection algorithms. The wearable data were collected via an open-source Android application and server infrastructure, with the E4 device streaming data via Bluetooth. The raw data time series were synchronized to the vEEG data in a preprocessing step, among other preparing procedures like deduplication and viable data block discovery.

Several aspects with respect to the effective reporting of evaluation results were also introduced. The data and signal quality provided by the data collection procedures are relevant. Hardware and firmware constraints of the wearable device, along with the data streaming collection mode, resulted in a substantial loss of data samples for most study participants. Furthermore, the signal quality of especially the BVP data is heavily impacted by motion artifacts, which needs to be kept in mind for those analyses utilizing this signal.

To evaluate seizure detection methodologies, several established metrics were introduced enabling gauging of the performance of a detector, as well as comparing it to other detectors. The sensitivity of a detector measures how many events were correctly classified out of all true events. The false alarm rate gives an estimate of how many false detections a model would make per a span of time. The positive predictive value is the proportion of true detections to all detections, informing about the relative amount of false alarms.

CHAPTER 4

Detection of Major Convulsive Seizures

Convulsive tonic-clonic seizures, while not the most predominant seizure type among patients with epilepsy in terms of incidence and prevalence, are among the most dangerous, and also widely known, seizures that can occur. With regard to automatic seizure detection with wearables, early warning, alarm, and automated diary systems are all relevant in some way. This thesis, however, focuses primarily on the context of seizure diaries. Both generalized tonic-clonic seizures and focal to bilateral tonic-clonic seizures are included in the studies presented here, considered to be reasonably straightforward to detect with data from wearable biosignal sensors.

The first part of this chapter includes a study centered around the assessment of a convulsive seizure detection methodology, employing features from accelerometry (ACC) and electrodermal activity biosignal data and a boosting ensemble machine learning model. The evaluation shows that the technique performs at least as good as some of the current best published detection systems in similar contexts.

The second part of the chapter examines the possibility to use wearables to detect and monitor post-ictal immobility, a state of unconsciousness and no movement that often occurs after patients have had a convulsive seizure and can foreshadow sudden unexpected death in epilepsy (SUDEP). A simple heuristic approach based on ACC data can robustly detect this period in a data set of convulsive seizures recorded in the hospital.

The detection of post-ictal immobility highlighted in the second part is specifically dependent on prior knowledge of a seizure, and thus would need to be joined with an automatic seizure detection in order to be used in a real-world alarm system for heightened susceptibility to SUDEP. While the detection methodology presented in the first section is not immediately compatible, it could potentially be adapted for this purpose, which is further explored and discussed in the summary of this chapter.

4.1 Detection of Tonic-Clonic Seizures

[104] ⇒ **Böttcher, Sebastian**, et al.
Detecting Tonic-Clonic Seizures in Multimodal Biosignal Data From Wearables: Methodology Design and Validation
2021, JMIR MHealth and UHealth, DOI:10.2196/27674

Parts of this publication were removed or edited to fit into the composition of this complete thesis. No substantial changes altering the results were made.

Own Contributions:

- All, except clinical expertise and the data collection at the KCL site

4.1.1 Introduction

Background

Epilepsy is one of the most common chronic neurological diseases, with a reported yearly worldwide incidence of more than 60 per 100,000 individuals [15]. Epilepsy also has a remarkably diverse set of indications, with several different types of symptoms and characteristic seizures of varying severity. Seizures are usually distinguished by their onset in the brain, focal or generalized. They can involve a variety of different combinations of symptoms, including impaired awareness or loss of consciousness; cognitive, emotional, or sensory abnormalities; sudden changes in the autonomic nervous system; or motor manifestations such as spasms, automatisms, or tonic and clonic movements of the limbs [27]. These convulsive seizures, particularly focal to bilateral or generalized tonic-clonic seizures (TCSs), are the most dangerous type of epileptic seizures. They imply loss of consciousness and loss of motor control with considerable risk for physical harm and can transition to life-threatening *status epilepticus* or sudden unexpected death in epilepsy [303].

For the diagnosis and treatment of epilepsy, clinicians rely on patient self-reporting and structured diaries, counting the number of seizures a patient had in a certain time frame. However, personal diaries filled out by the patients themselves have been proven to be very unreliable, with frequent undercounting because of a lack of awareness of seizures [54, 304]. An objective seizure diary is therefore needed to obtain valid data on seizure occurrence, contributing to improved guidance for the treatment of people with epilepsy. Wearable non-electroencephalography (EEG) devices (wearables) could provide data for such a diary. They are discreet and unobtrusive, contrary to many wearable EEG devices that are often cumbersome and stigmatizing [305], although some less obtrusive wearable EEG systems are in development [119, 306]. Moreover, a robust detection of convulsive seizures with wearables, paired with identification of seizure-related risk factors [105], could be of great clinical importance and provide essential information for the identification of seizure-related sudden unexpected death in epilepsy risk factors.

Although seizure detection with non-EEG wearables is a relatively new field in epilepsy research, there have already been some studies that have demonstrated the viability of this

kind of system. To date, most studies have concentrated on a single biosignal modality for training a seizure detection model, with a minority using a multimodal approach [106, 295].

Objective

In this study, an automatic seizure detection system for TCSs is presented, using supervised machine learning that is straightforward to implement and reproduce. The detection model is evaluated on a newly recorded data set from a multicenter clinical study with wearable non-EEG devices. Finally, the detection system, its performance, and its limitations are discussed and an outlook of possible further applications for this detection approach are concluded.

4.1.2 Methods

Data Set

During the course of the study, between July 2017 and February 2020, studies collected wearable device data from 243 patients diagnosed with epilepsy: $70.7\,\%$ (172/243) of patients were recruited at the epilepsy monitoring unit (EMU) of the University Medical Center Freiburg, and $29.2\,\%$ (71/243) of patients were recruited at the EMU in the neurophysiological department of the King's College Hospital London. Patients with a diagnosis of epilepsy in the age range of 7 to 80 years were recruited, unless they had vigorous involuntary non-epileptic movements. Consecutive patients were admitted to their respective EMU as part of their standard epilepsy clinical care, for differential diagnosis or for presurgical evaluation, and may have had their antiepileptic medication reduced during the recording. All patients were continuously monitored via a video-electroencephalography (vEEG) system during their stay in the EMU. Clinical experts manually reviewed the video and EEG data for all participants and labeled type, onset, and offset for all seizures. Specifically, they also labeled the onset and termination of every motor manifestation, including the tonic and clonic phases of each seizure. These labels were then used as the ground truth in the training and testing phases of the evaluation.

Participants wore a variety of different wearable devices across the 2 sites. However, the only device worn by participants from both sites was the Empatica E4 (Section 3.2.1; Empatica Inc., Boston, MA, USA). Owing to battery limitations, each participant was assigned 2 devices, between which they changed twice daily to ensure continuous recordings. The wearable device recorded 3-axis accelerometry (ACC) at a sample rate of $32\,\mathrm{Hz}$, electrodermal activity (EDA) at $4\,\mathrm{Hz}$, and photoplethysmography (PPG) at $64\,\mathrm{Hz}$, which was processed on the device to a blood volume pulse signal. Participants generally wore the device on the arm that was most involved in motor semiology during seizures, that is, the arm that presented the most substantial movements. In the set of 10 participants with TCSs included here, each wore the device on their non-dominant hand, except for 2 participants included in this study who specified that they were ambidextrous. The study and recording procedures are further described in Chapter 3.

All recruited patients provided written informed consent, and the study procedures were approved by the ethics committee at the University of Freiburg (538/16) and the London Fulham Research Ethics Committee (16/LO/2209; Integrated Research Application System project ID216316).

Features

An extensive feature set was created from the ACC and EDA signals, encompassing 141 ACC and 10 EDA features, at sliding window sizes of 2, 10, and 20 seconds for the ACC features, and 5, 10, and 20 minutes for the EDA features. PPG signals were not analyzed in this study because of major ictal movement artifacts resulting from the convulsive TCSs. Although artifacts in PPG data can still convey information, in that the presence of noise itself can be information, the choice was made to omit it here in favor of focusing on the other 2 biosignals, because the information of PPG motion artifacts is naturally included in the ACC signal as well. The ACC features included a variety of different time and frequency domain features. The EDA features represented the skin conductance level (SCL), that is, tonic low-frequency EDA changes, and skin conductance response rate (SCRR), that is, phasic or higher-frequency EDA changes, calculated against a baseline.

As detection models usually perform most effectively with smaller feature sets, both in terms of computational cost and prediction performance [307], the number of used features was reduced considerably. For this feature selection, related literature in the field of wearable seizure detection was consulted to narrow down window sizes that effectively capture relevant signal changes in time and identify feature types that were successfully used previously. Thereby, a window of 10 seconds for the ACC features was selected [214, 215, 224], and a longer window of 5 minutes for the EDA features to capture the tonic changes in the EDA signal that evolve over longer periods [153]. The feature data were then visualized in a period around the seizure, overlaid over each other, and for all features separately. In addition, the mean and standard deviation (SD) for each data series was plotted. The data that were used for these graphs were taken only from the seizures of participants that were not included in the test set to be used in the out-of-sample performance evaluation (see Section 4.1.3). Features showing recurrent typical ictal changes were then visually selected for further analysis (Figure 4.1). Variable seizure durations were handled by upsampling shorter seizures by linear interpolation to the length of the longest seizure among those plotted.

The resulting feature subset for the ACC modality consisted of the magnitude, zero crossing rate, and recurrence plot features (Figure 4.1) [308]. For the EDA features, the area under the curve and the maximum of the SCL within the window, and the SCRR were chosen, all corrected against a baseline, which is an interval of the same duration as the feature window, ending immediately before the beginning of the feature window. Thus, the resulting feature set can be divided into 4 main feature groups:

1. Magnitude of the ACC signal ($ACC = \sqrt{x^2 + y^2 + z^2}$):
   (a) Raw ACC signal, over a $10\,\mathrm{s}$ window.
   (b) Band pass filtered ACC signal over a $10\,\mathrm{s}$ window. The band pass filter had a frequency band of $0.1\,\mathrm{Hz}$ to $10\,\mathrm{Hz}$, representing the linear component of the ACC signal, and was applied before segmentation into windows.
   (c) Low-pass filtered ACC signal over a $10\,\mathrm{s}$ window. The low-pass filter had a cutoff frequency of $1\,\mathrm{Hz}$, thus preserving only the gravitational component of the ACC signal, and was applied before segmentation into windows.
2. Zero crossing rate of the ACC signal over a $10\,\mathrm{s}$ window, for each of the 3 axes, respectively. The zero crossing rate is the number of times in a certain period the signal crosses the value 0 over the same period.

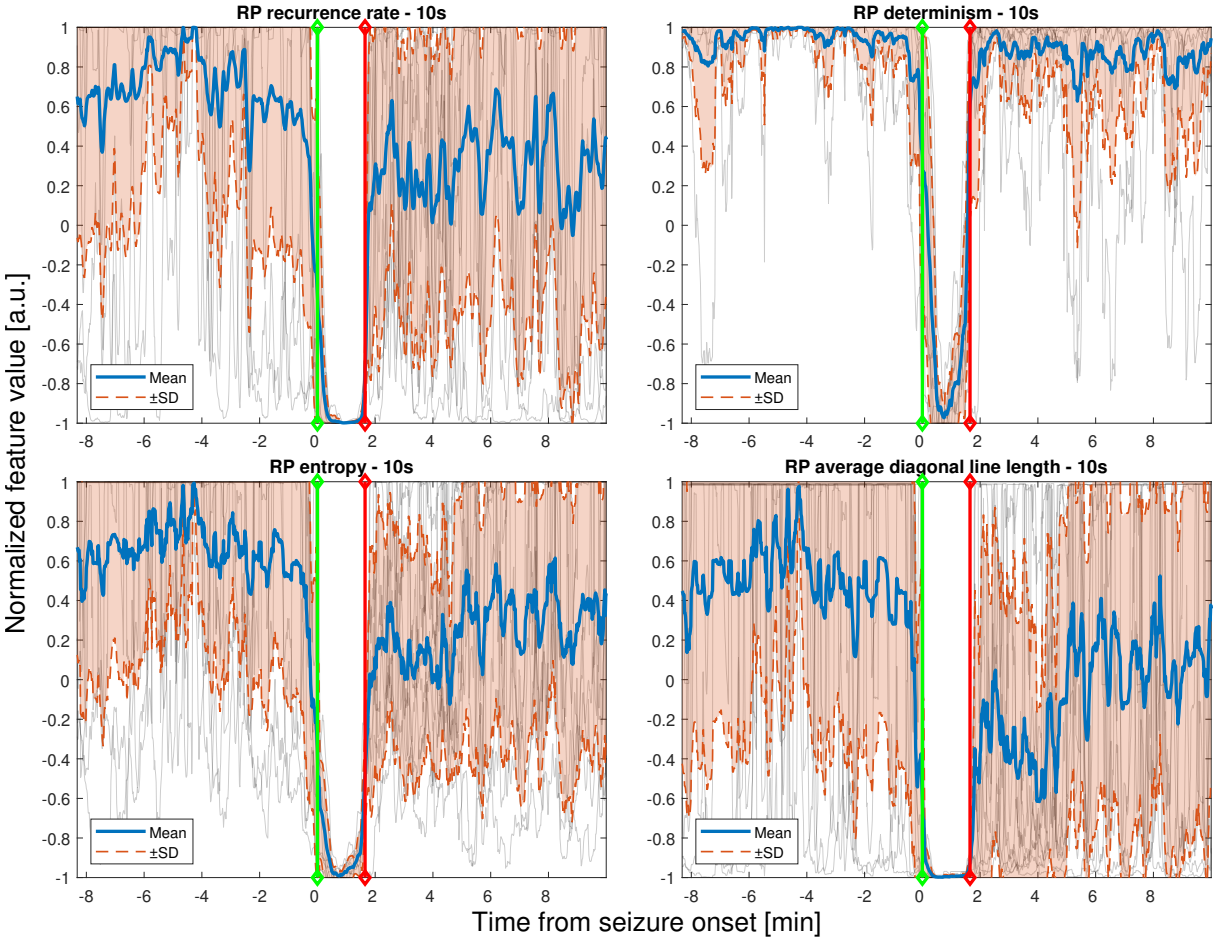


Figure 4.1: The overlaid feature value graphs for the four recurrence plot features calculated from 10 s windows of the ACC data. Graphs representing feature values for each individual seizure (gray, background) are overlaid by the mean and SD. The green and red vertical markers represent the seizure onset and offset, respectively. All features are normalized between −1 and 1 for this plot, independent of each other. Remaster of Figure 1 in Böttcher et al. [104].
RP: recurrence plot.

3. Four features calculated from the recurrence plot of the ACC signal:
   (a) Determinism, that is, the percentage of points that form diagonal lines of a minimal length.
   (b) The Shannon entropy of the probability that a line has a certain length.
   (c) The average diagonal line length.
   (d) Recurrence rate, that is, the density of recurrence points.
4. EDA-based features over a 5 min window, minus the same value in the 5 min before the feature window:
   (a) The area under the curve of the SCL was calculated as the moving mean of the raw EDA signal over a 1 min window.
   (b) The maximum value of the SCL calculated as above.

(c) The SCRR was calculated as the number of threshold crossings of the first derivative of the smoothed EDA signal within the window.

To accommodate the different window sizes over which the ACC and EDA features are calculated, a fixed interval between feature window applications was applied. Thus, all features are calculated at fixed time points, with their respective windows centered on each consecutive point, creating the same number of feature vectors for both the ACC and EDA features over a segment of data. This enables the use of the complete, merged feature space as the single input into a detection model for training [106]. The interval between the fixed time points for feature calculation was chosen to be 2 seconds. Figure 5.2 and Figure 5.5 give a graphical representation of the feature computation methodology.

Seizure Detection

A gradient boosted decision trees (GBT) model [309] was used as the detection model for the TCSs. Although similar to the well-known random forest (RF) method in being a set of trees that are grown with training data, a GBT model builds trees as weak learners in an additive manner. The model is improved with each new weak learner that is added to the ensemble, whereas the RF model trains all trees in parallel and independent of each other. Weak learners in this case are trees with a very low number of splits, down to decision stumps with just 1 split. This results in an overall lower bias and similar variance for GBT models compared with RF models at the cost of higher parameter tuning effort. Therefore, GBT models generally perform better than RF models if tuned sufficiently, and they have been successfully used in machine learning problems [95]. To tackle this tuning effort, a hyperparameter optimization was performed over several of the model parameters in a leave-one-participant-out (LOPO) manner. To this end, the data set was split into a training set and a test set. The training set consisted of the $10\,\mathrm{min}$ peri-ictal data of 10 TCSs from 8 patients with epilepsy recruited at the Freiburg site. The basic test set consisted of the complete data from 2 patients, 1 from the Freiburg site and 1 from the London site with 11 TCSs (see Section 4.1.3). The hyperparameter optimization only used the training set to keep the test set unknown to the model before testing. All feature data were normalized between -1 and 1 before training and testing. For training, the combined feature input for the model, that is, the peri-ictal feature data of 10 TCSs, were normalized, and for testing the complete feature data from the recordings for a participant were normalized independent of the feature data of the other participants in the test set.

The hyperparameter optimization was performed in a LOPO nested cross-validation manner on the training set. The data for 1 of the 8 participants in the training set were kept back as a validation set, and the model was trained on the seizures from the other 7 participants, using only $10\,\mathrm{min}$ peri-ictal data for each seizure. This reduction of the training data to only a small period around seizures helps with the large imbalance in the data set when comparing ictal and non-ictal epochs. Once the model was trained, it was then tested on the complete data of the validation participant in the respective round, and the process was repeated 7 more times, cycling through the participants for validation. The mean score of the 8 validation runs was then saved as the performance of the current parameter combination, and the entire validation process was repeated for the next parameter combination. The parameters that were tuned in the optimization and their divisions are listed in Table 4.1, with the resulting optimal

Table 4.1: Parameters optimized in the gradient tree boosting machine hyperparameter optimization and their optimization ranges.

Parameters	Value range	Description
Learning rate	1, **0.1***, 0.01, 0.001	The step size in the iterative learning process, also called shrinkage
Number of trees	25, 50, 100, **250***, 500, 750	The maximum number of trees to produce in the model
False positive cost	1, 10, 20, 30, 40, **50***	Specific misclassification cost for false positives when weighting during the learning process
Tree depth	**1***, 2, 4, 8, -1	The maximum number of splits in the decision tree, where -1 denotes one less than the number of samples in the training set, that is, the maximum possible value

* The chosen optimal parameter combination.

parameter combination highlighted. In total, 720 parameter combinations were evaluated in the hyperparameter optimization process.

Furthermore, the GBT model also had some fixed parameters that were the same for all optimization runs. The boosting method used in the model was adaptive boosting for binary classification [310], and the misclassification cost for false negatives was always 1. The hyperparameter optimization resulted in an optimal set of parameters that were subsequently used in all the testing steps. The optimal parameter combination was chosen as the combination that achieved the highest sensitivity and lowest false alarm rate (FAR) during the LOPO validation run of the parameter combination, prioritizing sensitivity. Model parameters not specified here were left at their default values.

Evaluation

To process the model output and score its performance when compared with the ground truth, the same method was used both in the validation during hyperparameter optimization and later during the testing phase (see Section 4.1.3). Owing to the method of feature extraction at fixed time intervals of 2 seconds, the output of the GBT model is a prediction vector containing the predicted label every 2 seconds. The input labels, that is, the ground truth, and the predicted labels were binary, denoting the classification of each 2-second interval to either belong to a seizure or not. Comparing the ground truth and the prediction labels for evaluation can be done sample-wise by comparing each 2-second interval, or event-wise, by combining consecutive intervals of the positive class to distinct events. In this analysis the latter method was chosen, which requires postprocessing of the model output.

First, the prediction output of the model was smoothed with a hysteresis-like filter to avoid single-sample positives or gaps in consecutive positive predictions. To this end, all gaps between consecutive positive predictions smaller than 20 seconds in duration were filled out as positive, thus creating continuous, longer events from short neighboring positive predictions. Thereafter, all consecutive positive predictions of a certain length were discarded. This value was chosen to be 4 seconds, as it provides a good balance between discarding short, single-sample predictions and still keeping possible relevant events. Thus, the prediction output of the model can be matched to the ground truth per participant by counting overlaps of predicted positive events with a positive ground truth event as true positives (TPs) and predicted positive events with no overlaps in the ground truth as false positives (FPs). The number of false negatives is then the difference between TPs and the number of seizures a participant recorded. The number of true negatives was not considered for this evaluation, as the sensitivity and FAR are sufficient to evaluate a methodology for seizure detection. Unless otherwise stated, the reported sensitivity $Sens = \frac{TP}{TP+FN}$ and $FAR = \frac{FP \cdot 24}{hours\ rec}$ are calculated across all relevant participants as a whole, not the mean over single participants.

All calculations for signal processing, feature extraction, and model development and evaluation were performed using MATLAB 2020a (MathWorks Inc, Natick, MA, USA).

4.1.3 Results

Overview

For the study presented here, only study participants with focal to bilateral or generalized TCSs were included. This resulted in a data set of 21 TCSs from 10 participants, 9 from the Freiburg site with 19 seizures captured, and 1 from the London site with 2 seizures captured. The mean length of convulsive motor phenomena was 64 (SD 23) seconds. Table 4.2 lists the clinical and demographic information of the participants. They were $40\,\%$ (4/10) female and on average 32.7 (SD 11.2) years old. The etiology of epilepsy for 2 participants was unknown at the time of recruitment. A total of 1 participant was diagnosed with generalized epilepsy, and the other 9 were diagnosed with focal epilepsy. For all captured seizures, wearable device data for at least $30\,\mathrm{min}$ before and after the ictal period were recorded in good quality; that is, the recorded data showed no major artifacts or intervals with constant 0 amplitude on visual inspection. A total of $612.6\,\mathrm{h}$ of data were recorded for the included participants with seizures.

Cross-validation Training

The training set used for hyperparameter optimization included 10 seizures from 8 participants and covered $414.7\,\mathrm{h}$ of wearable device data. With the best parameter combination, as described above, the LOPO cross-validation could detect all 10 seizures (sensitivity = $100\,\%$) with a total of 8 FPs (FAR = 0.46 per $24\,\mathrm{h}$). The FP rate was calculated as the ratio of total FPs across all participants to the number of hours of recordings multiplied by 24, and not the mean FAR across participants. In the training set LOPO cross-validation, $75\,\%$ (6/8) of FPs were produced from the data of 1 participant and 2 by another. Thus, the other 6 participants were free of FPs. All 8 FPs detected by the model during the LOPO cross-validation occurred when the patient was off camera, for example, in the morning or evening when they were in the bathroom for their daily washing routine.

Table 4.2: Participants with recorded tonic-clonic seizures that were included in this study. Wearable data recorded from these participants were used in the evaluation of the seizure detection model. The recording duration is the duration that participants were wearing the device, without accounting for data loss.

ID	Gender	Age	Recording Duration	Epilepsy Origin	Epilepsy Type
FR1	Female	35 years	5 days	Unknown	Focal (TLE)
FR2	Female	26 years	6 days	Structural	Focal (TLE)
FR3	Male	22 years	4 days	Genetic	Generalized (IGE)
FR4	Female	34 years	4 days	Unknown	Focal (FLE)
FR5	Male	56 years	8 days	Structural	Focal (TLE)
FR6	Male	38 years	7 days	Structural	Focal (TLE)
FR7	Male	25 years	4 days	Structural	Focal (xTLE)
FR8	Male	16 years	7 days	Structural	Focal (FLE)
FR9	Male	37 years	12 days	Structural	Focal (xTLE)
LO1	Female	38 years	6 days	Structural	Focal (TLE)

TLE: temporal lobe epilepsy; IGE: idiopathic generalized epilepsy; FLE: frontal lobe epilepsy; xTLE: extratemporal lobe epilepsy.

Out-of-Sample Testing

The model was also tested using a previously unseen test set from the overall data set. This test set included 11 seizures from 2 participants, 1 from the London site with 2 seizures recorded, and the other from the Freiburg site with 9 seizures recorded, for a total of $197.9\,\mathrm{h}$ of test data. The choice of training and test set was deliberate: With the relatively low number of seizures and their distribution among participants in this data set, the goal was to train on as many participants as possible while also having approximately the same number of seizures in the test set. This allocation ensures a model that is not patient specific while keeping the training and test sets balanced in terms of the number of seizures.

The GBT model with the optimal parameters and trained with all 10 seizures from the training set could detect 10 of the 11 seizures in this test set (sensitivity = $91\,\%$), without any FPs. However, this test set was limited in that it was biased towards participants who had convulsive seizures. Therefore, the test set was expanded to also include data from all 30 patients with epilepsy recruited at the London site that had data recorded with the wearable device. Although this does not add more seizures to the test set for the model to detect, it does add a considerable amount of data, improving the assessment with respect to the FP rate. The expanded test set thus encompasses $1935.9\,\mathrm{h}$ of wearable device data from 31 participants, including the same 11 seizures as before. In this data set, the same model produces 30 FPs (0.37 per $24\,\mathrm{h}$). Further investigation of the FP distribution among the participants showed that 15 false detections resulted from a single participant who used a stepper during monitoring as physical activity to trigger her seizures. All FPs for that participant were related to this activity. Removing this participant, performing unnatural repetitive movements, from the expanded test set lowers the FP rate to 0.19 per $24\,\mathrm{h}$. Of the other participants in this expanded test set, the data of 2 participants produced 3 FPs, respectively, whereas 9 other participants each

Table 4.3: Per participant evaluation results, for participants with seizures recorded. The 3 totals given for the test set are (a) the total across the test set participants with seizures recorded ($N = 2$), (b) the total when including all patients with epilepsy recruited at the London site with data recorded (not listed, $N = 31$), and (c) the total when excluding 1 participant with an artificially disproportionate number of false positives ($N = 30$).

ID	Sensitivity	FP	FAR24	PPV	Recording Length	Seizure Type
			Training Set			
FR1	1 (100 %)	0	0	100 %	59.6 h	FBTCS
FR2	1 (100 %)	6	1.56	14 %	92 h	FBTCS
FR3	2 (100 %)	0	0	100 %	35.5 h	GTCS
FR4	1 (100 %)	2	1.34	33 %	35.8 h	FBTCS
FR5	1 (100 %)	0	0	100 %	36.3 h	FBTCS
FR6	1 (100 %)	0	0	100 %	88.5 h	FBTCS
FR7	1 (100 %)	0	0	100 %	40.7 h	FBTCS
FR8	2 (100 %)	0	0	100 %	26.2 h	FBTCS
Total	10 (100 %)	8	0.46	56 %	414.7 h	
			Test Set			
FR9	9 (100 %)	0	0	100 %	112.2 h	FBTCS
LO1	1 (50 %)	0	0	100 %	85.7 h	FBTCS
Total (a)	10 (91 %)	0	0	100 %	197.9 h	—
Total (b)	10 (91 %)	30	0.37	25 %	1935.9 h	—
Total (c)	10 (91 %)	15	0.19	40 %	1870.3 h	—

produced 1 FP, with the remaining 19 participants being free of FPs. Thus, the FAR, when calculated as the mean across all the included participants' individual FARs, was 0.45 (SD 1.1) per 24 h, and 0.29 (SD 0.53) per 24 h when excluding the participant with 15 FP. Table 4.3 provides a detailed overview of the results among the participants with recorded seizures.

Seizure Duration

The duration of detected seizures was significantly correlated with the vEEG-based seizure duration, as labeled by clinical experts (Figure 4.2). The true seizure duration here is based on its clinical manifestation, that is, onset until offset of ictal motor phenomena related to TCSs. In a Pearson correlation test, the correlation coefficient was $r = 0.55$, with $P = 0.01$. In general, the seizure duration was underestimated by the model by approximately half of the true duration, with a mean identified duration of 29 (SD 15) seconds versus the mean seizure duration of 64 (SD 23) seconds. This may be reflective of the less severe movement amplitudes during the tonic phases of the TCSs.

Feature Importance

Furthermore, the feature importance for the feature set was analyzed as a metric for the contribution of a specific feature to the performance of the model. It was calculated as the mean

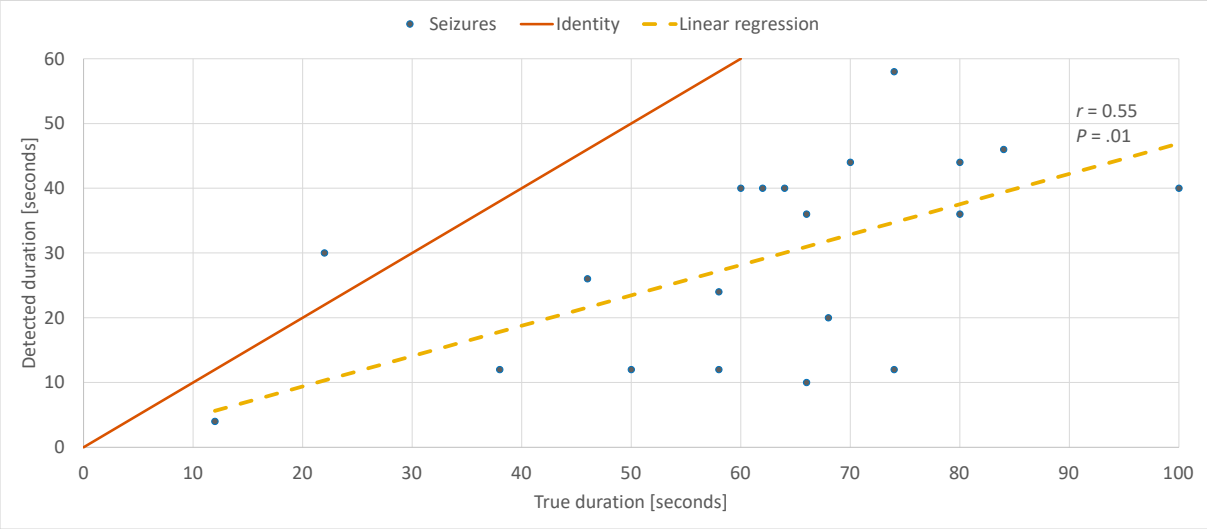


Figure 4.2: Correlation of the true seizure durations as labeled by clinical experts and the ictal durations detected by the gradient tree boosting machine model based on accelerometry and electrodermal activity. The dotted line shows the linear regression fit across the data points. The Pearson correlation coefficient was $r = 0.55$, with $P = 0.01$. The identity line shows that the seizure duration is generally underestimated by the model.

feature importance over all trained GBT models in the LOPO cross-validation (Figure 4.3). The feature importance was based on the Gini impurity, calculated such that the smallest possible value was 0 [311]. Overall, all 4 feature groups, as described in the Features section, are represented in the resulting GBT model to varying degrees of importance. The top 3 features among the feature set were that the Shannon entropy of the probability that a line in the recurrence plot had a certain length calculated over a $10\,\mathrm{s}$ window of the ACC signal, the magnitude of the band pass filtered ACC signal in a $10\,\mathrm{s}$ window, and the maximum of the SCL in a $5\,\mathrm{min}$ window of the EDA signal, corrected for a baseline.

4.1.4 Discussion

Principal Findings

The results show that the GBT model can robustly detect TCSs from non-EEG wearable device data. A sensitivity of $100\,\%$ (10/10) on the training set during a LOPO cross-validation, a sensitivity of $91\,\%$ (10/11) on the out-of-sample test set, and a FAR of less than 1 per 5 days in more than $1800\,\mathrm{h}$ of data indicates a sufficient robustness of this methodology to consider it in designing an automated seizure diary. A large percentage of FPs occurred in a small percentage of participants, with most other participants showing between 0 and 0.5 FPs per day. Furthermore, in participants of the test set who had TCSs, no FPs were reported by the model. In addition, all true detections of the model occurred within the ictal period of the respective seizure, showing that the system has high accuracy. Evaluating a test set that includes data largely from 1 site (London), while the model was trained exclusively with data from the other site (Freiburg), shows the generalizability of the model.

Although the data set contains continuous circadian data, most TCSs occurred during nighttime sleep. In the training set, $50\,\%$ (5/10) of seizures occurred while the patient was

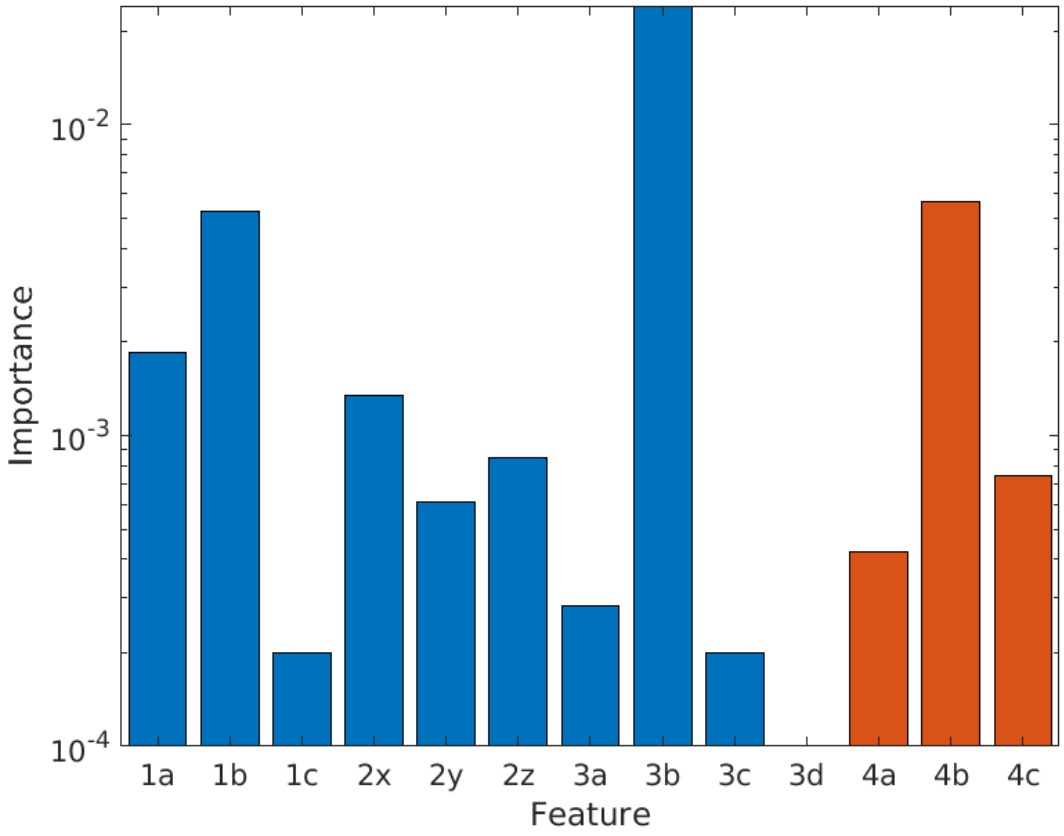


Figure 4.3: Feature importance, calculated as the mean feature importance of all models during a leave-one-participant-out cross-validation, with the optimal parameters of the gradient tree boosting machine as reported in the Seizure Detection section. All the features are shown as listed in the Features section (1: magnitude of accelerometry, 2: zero crossing rate of accelerometry, 3: recurrence plot features of accelerometry, and 4: electrodermal activity features). The feature importance is shown in logarithmic scale to better visualize smaller differences.

awake, and in the test set, only 9 % (1/11) occurred during wakefulness. Of these 6 awake seizures, 2 seizures occurred when the patient was outside the bed. All TP detections, both in the training set LOPO cross-validation and in the test set evaluation, occurred within the ictal phase of the respective seizure. Conversely, all FP detections occurred when the patient was awake and active, and most of them occurred during daytime. Patients were generally not confined to their beds but rather to their hospital rooms. They could freely perform a variety of activities of daily living, such as strolling across the room, going to the bathroom, brushing their teeth, eating and drinking, and washing themselves. Movement patterns during these activities, particularly if repetitive, could resemble those during convulsive seizures and may be a common source of FP detections. However, false alarms during these activities when the patient is awake could be ignored easily by way of patient validation and feedback to avoid inappropriate interventions.

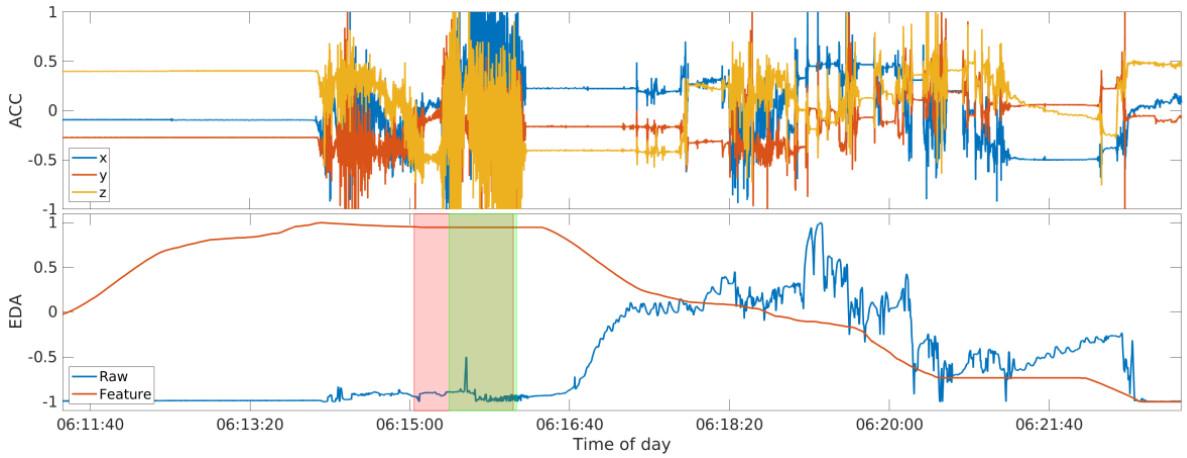


Figure 4.4: The seizure of participant LO1 that was detected by the model. The raw ACC signal is shown at the top, and the raw EDA signal as well as the best EDA feature (Section Features, Feature 4b) at the bottom; all are normalized between -1 and 1, independent of each other. The ictal tonic-clonic phase is overlaid in red, the true positive detection is overlaid in green.

Feature Importance

The distribution of feature contribution to the performance of the model shows that all selected features are used by the model to predict a seizure event, except for one, the recurrence rate in the recurrence plot of the ACC signal. The least amount of importance is assigned to the magnitude of the low-pass filtered ACC signal. This is an expected outcome, as this feature represents the gravitational component of the movement, which is minimal during convulsive seizures. During these seizures, almost all movements are part of the linear component, represented by the band pass filtered signal, which is also confirmed by this feature being one of the most important in the model.

Among the EDA-derived features, the highest importance was consistently assigned to the difference between the highest value in the feature and the baseline windows of the SCL. A typical EDA signal progression in the peri-ictal period is a steep increase from a low pre-ictal baseline during the ictal phase, followed by a shallow decrease in the post-ictal phase, spanning multiple minutes. Thus, the feature based on the difference of the highest value between pre-ictal, ictal, and post-ictal phases can sufficiently represent this trend, as evidenced by its high importance. Figure 4.4 shows the EDA signal progression and the respective maximum SCL feature during a seizure. The feature values are at their highest during the ictal phase, whereas the raw EDA signal shows the typical progression described above.

False Negatives

There was 1 seizure the model did not detect among the training and testing data sets (Figure 4.5). This false negative was produced by one of the participants recruited at the London site, and the seizure occurred during the night when the patient was asleep. The other seizure recorded for this participant was successfully detected by the model. The raw data before and after the seizure can shed light on why the seizure was rejected by the model, specifically looking at the ACC response during the seizure, and the EDA trend going from the pre- to post-ictal

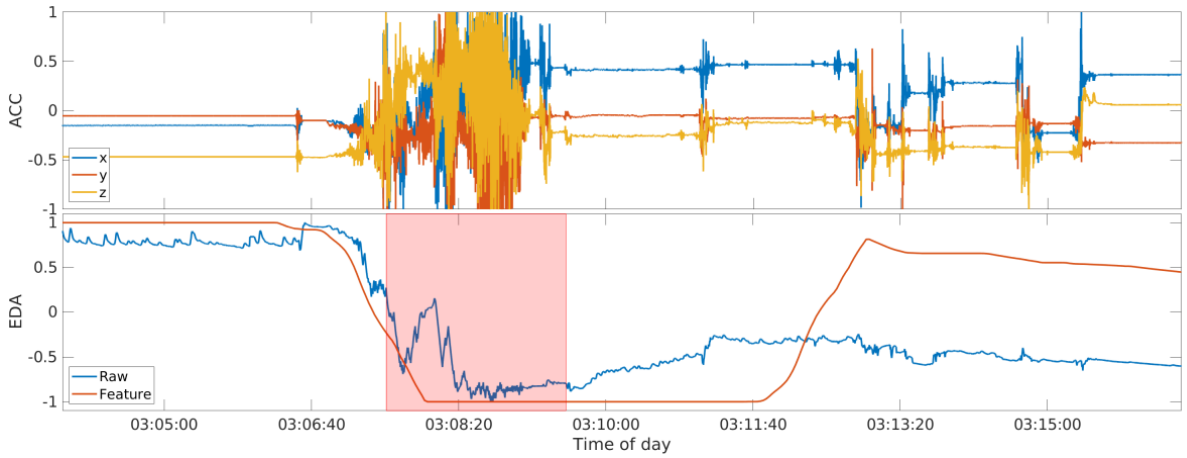


Figure 4.5: The seizure of participant LO1 that was not detected by the model and the single false negative that was produced during the evaluation. Note the differences in the EDA signal progression in comparison to Figure 4.4, which shows a typical response. The raw ACC signal is shown at the top, and the raw EDA signal and the best EDA feature (Section Features, Feature 4b) at the bottom; all are normalized between -1 and 1, independent of each other. The ictal tonic-clonic phase is overlaid in red.

phase. The motion response in the ictal phase of the rejected seizure was a typical progression from a short tonic phase at the beginning of the seizure to a longer, very pronounced, and violent clonic phase, stopping promptly with the seizure offset, followed by a short phase of post-ictal ACC silence. The raw EDA signal, however, follows a progression directly opposite to the signals from all other TCSs in the data set. The signal shows a steep decrease from a high baseline during the ictal phase and remains at a lower level in the post-ictal phase compared with the baseline in the pre-ictal phase. Figures 4.4 and 4.5 show the comparison of data from the 2 recorded seizures from participant LO1, with the detected seizure being representative of all other TCSs in the data set, especially those in the training set that created the model. Both seizures showed similar ACC data and a similar change in the ACC-based feature values. However, the EDA data and feature values were visibly opposite. This confirms that the model was trained properly on both the ACC and EDA features and that both modalities contributed to the model's classification of seizure occurrence. Thus, the misclassification of 1 event was due to atypical raw data and confirmed that the model included EDA features in its classification.

A possible explanation for the unusual EDA signal during this seizure could be that the EDA electrodes lost adequate contact with the skin, and it was not fully re-established after the seizure. This could be caused by an improperly worn wearable device, or a loss of contact owing to the wearable device coming into contact with an external obstacle such as being pressed into the bed, slightly raising the EDA electrodes off the skin.

Related Work

The research that is most closely related to the premise of this study is certainly that of Onorati et al. [224]. In their work, the Empatica research group developed a seizure detection model based on wearable data from the same device used in this study, Empatica E4. They used a support vector machine trained with 25 ACC as well as EDA features that were not further

specified to detect convulsive seizures and achieve a very good performance, with their best classifier reaching a sensitivity of $94.5\,\%$ and a FAR of 0.2 per day on 55 seizures from 22 patients. The work presented here is on par with their results and reinforces their findings. It is shown that results of this quality can be achieved with a relatively basic methodology, which is described in greater detail here, making it fully accessible and reproducible. The methodology may even be transferrable to other diseases with convulsive attacks, such as paroxysmal dystonia[1] or psychogenic non-epileptic seizures. Thus, the study described here could be used as a stepping-stone for further work not only in epilepsy research but also in other medical fields.

In a further study, Kusmakar et al. [215] used a monomodal support vector data description model on wearable ACC data to detect 21 generalized TCSs from 12 patients, with a total recording length of $966\,\mathrm{h}$. The outlier classification model could achieve a sensitivity of $95\,\%$ in a LOPO cross-validation, with a mean FAR of 0.72 per day. However, their model generated FP detections across almost all of the 12 included patients, showing a general trend toward FP detections independent of patient selection, whereas the model evaluated here could achieve a generally lower FP rate on both the training and test sets, also revealing certain patients with a disproportionate FAR.

Arends et al. [226] used the LivAssured NightWatch wearable device in a large ambulatory long-term monitoring study, collecting 908 convulsive seizures from 28 patients over more than 1800 nights. The device collects ACC and PPG signals from the patients' upper arm, specifically during the night. Their thresholding algorithm could detect $86\,\%$ of the recorded seizures, with a positive predictive value of $49\,\%$, indicating that roughly half of all predictions were FPs. Although the methodology introduced here produces slightly worse results with respect to the overall FAR, the studies differ in that the NightWatch study only assessed nocturnal data with patients at rest, whereas the assessment presented here, based on continuous data comprising wakefulness and sleep, showed the model's ability to correctly detect daytime seizures; notably, all the FPs were generated while the respective patient was awake and active.

In a more recent study, Johansson et al. [214] used wrist-worn ACC sensors to detect 37 TCSs from 11 patients with $666\,\mathrm{h}$ of data. They evaluated 3 different types of models on a test set of 10 seizures and obtained the best result using a RF algorithm, detecting 9 of 10 seizures with a FAR of 0.24 per day. However, the evaluation of FPs is constrained to patients with TCSs, introducing a certain bias in patient selection. In this evaluation, a control group of up to 29 participants without TCSs recorded was added, with the model achieving a similar FAR, while also only producing FPs on these participants without seizures, whereas the participants with TCSs had no false alarms.

Limitations

The methodology for TCS detection described here also introduces some limitations, one of which is the long feature window used for the EDA feature computation. A $5\,\mathrm{min}$ long window enabled the inclusion of tonic changes in the EDA signal spanning over multiple minutes in the post-ictal phase. However, this automatically introduces an inherent detection delay, as a real-time system would need to first collect these data before being able to extract the

[1] A neurological hyperkinetic movement disorder causing sudden pain, tremors, or abnormal movements of the body, limbs, or face

EDA features and detect a potential seizure. Thus, this methodology would not be suited as a real-time warning system. Another limitation is the constraint of the model to detect only TCSs. As the model training process relies on data from the accelerometer sensor, non-motor seizures cannot be detected with this set of modalities and features. Future work will be needed to assess the contribution of the PPG and EDA sensors in detecting non-motor seizures. Furthermore, the performance of the specific model trained here is likely not sufficient to be deployed directly as an automatic seizure diary, especially considering its constraint on TCSs, which can be infrequent in everyday life. Additional work and more training data would be needed to create a system that is usable in clinical practice, possibly even shifting to a semi-personalized model that can be reinforced over time by patient feedback.

One of the most prevalent limitations in many studies in this field is the controlled in-hospital setting in which wearable device data are collected. Although patients in this study were able to perform some activities of daily living in and around their bed and were able to walk within their hospital room, the likelihood of false positive generation can be assumed to be higher in an outpatient setting. False alarms during physical activity could be addressed by actively involving the patient through validation and feedback, for example, by giving them a chance to review seizure diary entries. Nevertheless, transferring this methodology to an ambulatory setting will require extensive modifications and reevaluation with data recorded in everyday living situations that include a gold standard for seizure labeling. In any case, a robust classifier that has a likelihood of working in the field must first be validated in an inpatient setting to progress to an ambulatory study, and the research presented here takes a clear step in that direction.

4.2 Potential Use of Tonic-Clonic Seizure Detection

[105] ⇒ Bruno, Elisa and **Böttcher, Sebastian**, et al.
Post-ictal accelerometer silence as a marker of post-ictal immobility
2020, Epilepsia, DOI:10.1111/epi.16552

Parts of this publication were removed or edited to fit into the composition of this complete thesis. No substantial changes altering the results were made.

Own Contributions:

- Contribution to methods (4.2.2)
- Wearable device data analysis (4.2.2)
- Contribution to description of results (4.2.3)
- Contribution to discussion (4.2.4)

4.2.1 Introduction

Movement analysis based on accelerometry (ACC) signals has shown an overall good performance for the detection of convulsive seizures (generalized tonic-clonic seizures (GTCSs) and focal to bilateral tonic-clonic seizures (FBTCSs)) in epilepsy monitoring units (EMUs) [312–315] and, to a lesser extent, in real-life settings [298]. A potential application of movement sensors may also be the identification of the diametrically opposite feature: the absence of motion. Immediately following a seizure, this phenomenon represents an interesting clinical manifestation, known as post-ictal immobility (PI). PI has been recognized by clinicians for more than a century [316], although the pathophysiological[1] mechanism has remained largely unclear. Frequently observed after convulsive seizures, PI has been associated with potentially life-threatening complications and with sudden unexpected death in epilepsy (SUDEP) [317]. In addition, PI has often been observed in association with post-ictal generalized electroencephalography suppression (PGES) [318–322], an electroencephalography (EEG) pattern recorded in SUDEP cases [303, 323–329].

The identification of PI through wearables has not yet been adequately explored. Because most convulsive seizures do not lead to SUDEP and the occurrence of a fatal seizure is unpredictable, the automatic, continuous, long-term identification of risk factors for SUDEP associated with each individual seizure assumes great clinical importance. In a population at high risk of SUDEP, this study aims to assess whether ACC could be used as a reliable digital marker of PI and of its duration after convulsive seizures. In addition, the association of post-ictal ACC silence with PGES, and with other physiological and clinical variables associated with SUDEP, was investigated.

[1] Functional changes occurring due to a disease

4.2.2 Methods

Study Participants

Like the other studies included in this thesis, this study was developed in the context of "*Remote Assessment of Disease and Relapse – Epilepsy*", a multicenter study designed to assess the clinical utility of multi-parametric remote measurement technologies in a clinical population with epilepsy, in the hospital and real-world environment [291]. The study population consisted of consecutive patients with epilepsy (PWEs) who were admitted, for diagnostic reasons or presurgical evaluation, to the EMU at either the neurophysiology department at King's College Hospital London, or the Epilepsy Center at the University Medical Center Freiburg. Participants presenting with convulsive seizures (GTCSs or FBTCSs) were included.

Ethics Approval

The trial and study procedures were approved by the London Fulham Research Ethics Committee (16/LO/2209; IRAS project ID216316) and in Freiburg by the Ethics Committee at the University of Freiburg (538/16). All participants provided written informed consent.

Recordings and Data Collection

Measurements: Participants were asked to wear a wrist-worn multimodal device (Empatica E4, Figure 4.6a; Section 3.2.1; Empatica Inc., Boston, MA, USA) at both sites for the entire duration of their stay in the EMU. Additionally, some participants at the London site were also asked to wear an upper arm-worn device (IMEC armband, Figure 4.6b). Among other sensors, both the devices have an embedded three-axis accelerometer capturing XYZ raw ACC at a sampling frequency of $32\,\mathrm{Hz}$ and an electrodermal activity (EDA) sensor measuring the electrical conductance of the skin through dry electrodes. Both the devices measure the conductance between two electrodes on the skin by applying a direct current to the stratum corneum[1] beneath the electrodes. The sampling frequency for the EDA measured at the ventral side of the wrist is $4\,\mathrm{Hz}$ (Empatica E4) and at the ventral side of the upper arm is $256\,\mathrm{Hz}$ (IMEC armband). The devices were worn on the nondominant hand. Video-electroencephalography (vEEG) recordings were performed using a minimum of 21 scalp electrodes.

Clinical and Seizure Characteristics: For each patient, demographic and clinical data (including age at onset of epilepsy, type of epilepsy, seizure frequency, and so on) were collected. For each seizure recorded, data on state of wakefulness (awake/asleep), ictal focus, duration of the clonic and of the tonic phases, as well as the entire seizure duration were recorded. Convulsive seizures were classified into three categories according to Alexandre et al. [330]. Early intervention by a nurse (during the seizure or within the first $5\,\mathrm{s}$ after seizure termination), early administration of oxygen (with oxygen mask during the seizure or within the first $5\,\mathrm{s}$ after seizure termination), and prone position at seizure end were also annotated.

Seizure, PGES, and PI Annotation: Two neurologists with EEG expertise reviewed the vEEG recording independently and manually labeled the start (first EEG or clinical manifestation) and the end of the seizures (EEG end) and the presence/absence, onset/offset, and duration of PI and PGES. A random sample of vEEG was reviewed by both the annotators to

[1] Outermost layer of the epidermis (skin)

(a) (b)

Figure 4.6: (a) Empatica E4 wristband. Source: empatica.com **(b)** IMEC upper-arm band. Source: imec-int.com

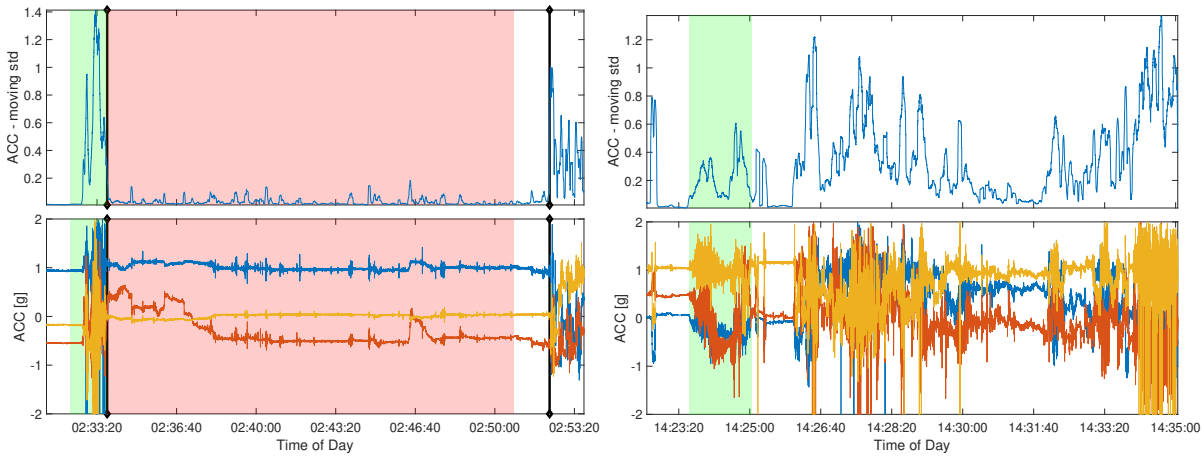
guarantee consistency of the labeling procedures. PGES was defined as the post-ictal generalized absence of electroencephalographic activity $<10\,\mu V$ in amplitude, allowing for muscle, movement, breathing, and electrode artifacts [327]. Seizures presenting with a PGES duration $\geq 20\,s$ were considered at higher SUDEP risk [327]. PI was defined as the post-ictal absence of movements (allowing for respiratory movements) on the video recording. The duration of PI was defined as the time from the onset of PI to the onset of the first post-ictal active non-respiratory movement [321].

Wearable Devices Data: Raw data from the wrist-worn wearable device (Empatica E4) were streamed via Bluetooth to an android app during the recordings and collected on a centralized data server [291], as described in Section 3.2. Raw data from the upper arm device (IMEC armband) were transferred directly from the device to a laptop via cable at the end of the recording. Both devices and the vEEG were synchronized with a timeserver at the beginning of each recording. If there was still time drift left on the signal during a seizure, the offset was manually determined by visual comparison of the expert-labeled video and raw data.

Data Analysis

Automatic Detection of Post-Ictal ACC Silence: The ACC data immediately around the seizure event were plotted and visually inspected. The information obtained from the plots was used to create an algorithm for the automatic detection of the post-ictal ACC silence. Given the seizure offset marked by clinical experts, the post-ictal ACC silence detection algorithm marked the "start" of the post-ictal ACC silence if the moving standard deviation was < 0.2 for $5\,s$. The algorithm then marks the "end" of the post-ictal ACC silence if the moving standard deviation was ≥ 0.2 for at least $5\,s$. Thereby, the duration of the post-ictal ACC silence was calculated. The moving standard deviation was calculated over a $5\,s$ window with maximal overlap.

Electrodermal Activity: The EDA was plotted around the seizure for visual inspection. To account for the typically slow changes in tonic EDA, an additional hour of data before and after the seizure events was also included in the plots. The analysis of the EDA signal was based on Poh et al. [153]. The signal was filtered and smoothed to reduce motion artifacts and



(a) (b)

Figure 4.7: (a) Post-ictal ACC silence detected by the algorithm and PI labeled by clinical experts, **(b)** post-ictal agitation without ACC silence. In each plot, the first graph reports the ACC moving standard deviation, the second the ACC raw signal, and the third (if applicable) the EDA signal. The seizure period is highlighted in green, the expert-labeled PI is highlighted in red, and the automatically detected post-ictal ACC silence start and end are indicated by the black vertical lines. Remaster of Figure 2 in Bruno et al. [105].

to obtain the tonic component of the raw EDA signal. The EDA baseline was computed over the $60\,\mathrm{min}$ pre-ictal segment. The start of the EDA response was defined as the point after the labeled start of the seizure where the filtered EDA signal first reached a value of baseline $+2\sigma$. The end of the response was defined as the point after the start of the response where the filtered EDA signal first dropped below $80\,\%$ of the response peak. The EDA duration was estimated as the time between start and end of the EDA response. Furthermore, the response amplitude (EDA amplitude) was defined as the difference between the response peak and the pre-ictal baseline.

Statistical Analysis

Linear regression was used to analyze the linear pairwise relationship between the automatically detected duration of post-ictal ACC silence and other variables, such as duration of PI, seizure duration, and others. The strength and direction of the linear relationship between the different pairs of variables were quantified using the Pearson correlation coefficient.

Variables with $P < .2$ at univariate analysis were entered in a multivariate model. A two-sided Mann-Whitney-Wilcoxon rank-sum test was used to assess the relationship between the duration of the post-ictal ACC silence and binary clinical variables, whereas the Kruskal-Wallis test was used for categorical variables. To account for multiple comparisons, the resulting P-values were adjusted using the Bonferroni method. Each test was performed at a significance level of $.05$. When seizures were captured with both the devices (Empatica E4 and IMEC armband), the signals collected from the wrist-worn device (Empatica E4) were included in the analysis. Moreover, a subgroup analysis on the EDA response captured by the wrist-worn device, producing better signals, was performed. Data were processed using

Table 4.4: Linear regression analysis of duration of post-ictal ACC silence and continuous clinical and seizure characteristics.

Variable	Univariate Analysis		Multivariate Analysis
	r	P-Value	P-Value
PI Duration	0.92	**<0.001**	**0.000**
Age	0.78	**<0.001**	**0.041**
Seizure Duration		0.88	
Duration of Tonic Phase		0.42	
Duration of Clonic Phase		0.38	
PGES Duration	0.40	**0.033**	0.75
EDA Duration		0.46	
EDA Amplitude		0.33	

r: Pearson coefficient. P-values in bold are statistically significant.

STATA 14.0 (StataCorp LLC, College Station, TX, USA) and MATLAB R2019b (MathWorks Inc, Natick, MA, USA).

4.2.3 Results

Participants and Seizure Characteristics

Twenty-two convulsive seizures were recorded from 18 study participants between September 2017 and October 2019. The mean age of study participants was 37.1 years (standard deviation (SD) 12.8). The mean disease duration was 19.8 years (SD 7.4) and the median number of anti-epileptic drugs (AEDs) taken was 2 (range 1-4). The majority of the participants ($66.7\,\%$) reported $\geq$ 3 convulsive seizures per year. All were admitted to the EMU for presurgical evaluation of their pharmacoresistant epilepsy. Eleven patients wore the wrist-worn device, four the arm band, and three both devices. Thirteen seizures were recorded with the wrist-worn device, five with the arm band and four with both.

PI and ACC silence

PI occurred in 20 of 22 seizures ($90.9\,\%$), whereas 2 seizures were followed by post-ictal agitation and confusion. PGES occurred following 15 seizures ($75.0\,\%$) and it was $\geq 20\,\mathrm{s}$ in 11 ($55.0\,\%$), which were considered at higher SUDEP risk.

Early nurse intervention was performed in all the seizures recorded and consisted mainly of assisting the patient into the recovery position or repositioning the head. Oxygen was administered early in 10 seizures ($45.4\,\%$). None of the patients was observed in prone position at seizure offset due to nurse intervention. In the post-ictal period, the automated estimation of post-ictal ACC silence identified all the 20 expert-labeled PI, discarding post-ictal agitation, and performing equally in both the wrist and upper arm-worn devices. Figure 4.7 illustrates an example of the post-ictal ACC silence detected by the algorithm and its correspondence with PI labeled by experts, and an example of post-ictal agitation. Figure 4.8 illustrates the ACC signals captured during the same seizure by the two different devices and the similar performance of the algorithm on the two different body sites. Occasionally, the post-ictal ACC silence lasted beyond the PI labeled by experts, as very subtle movements of the hands

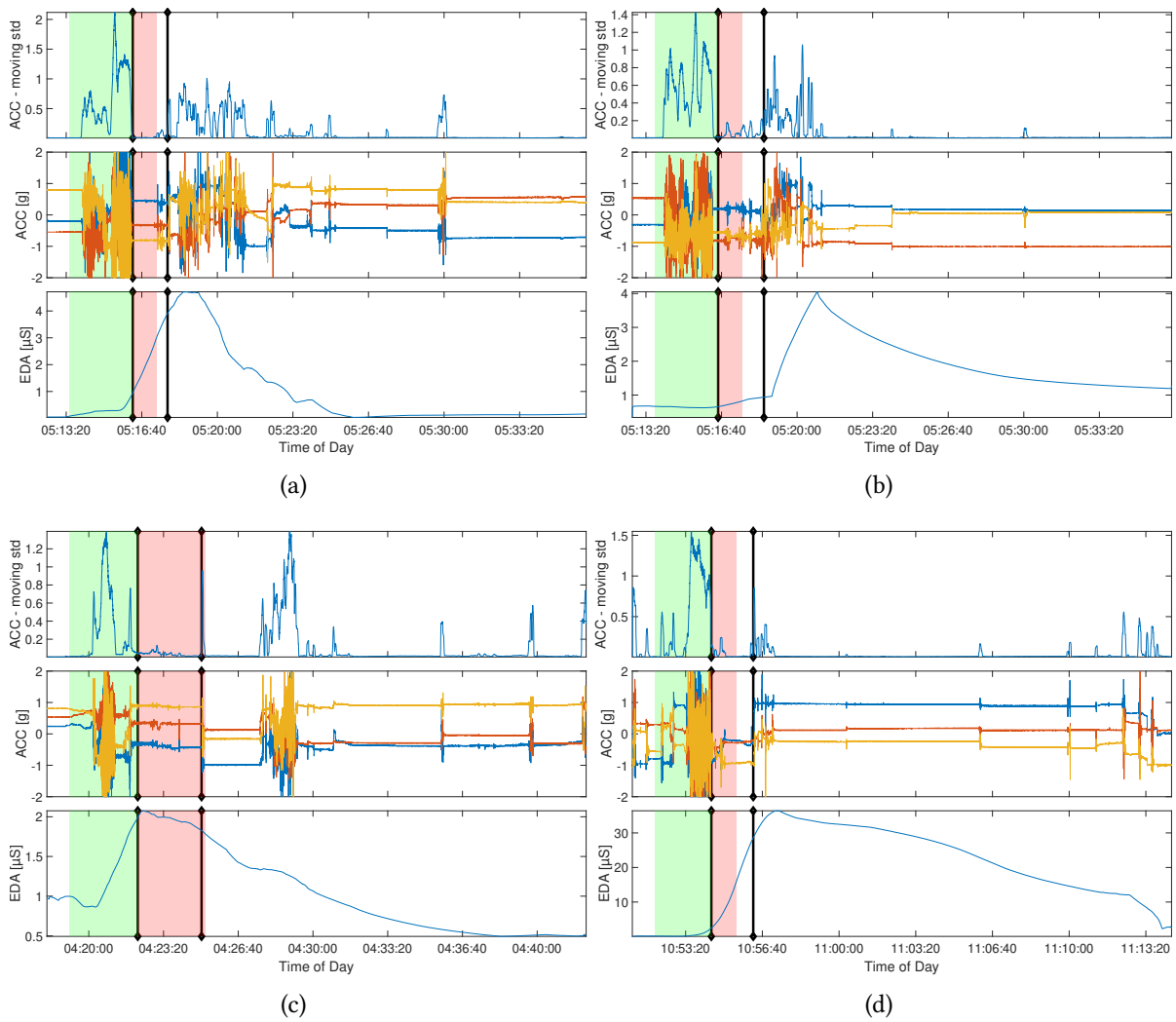


Figure 4.8: Post-ictal ACC silence detected by the algorithm in one seizure recorded simultaneously by **(a)** the wrist-worn device (Empatica E4), and **(b)** the upper-arm worn device (IMEC). **(c and d)** Peri-ictal EDA response observed in two different seizures from two study participants. Remaster of Figures 3 + 4 in Bruno et al. [105].

or neck were used by experts to establish the end of clinical PI, but were not captured by the ACC threshold algorithm.

The linear regression (Table 4.4) showed that the duration of the post-ictal ACC silence was correlated with the duration of expert-labeled PI (Pearson $r = .92$; $P < .001$; Figure 4.9), with the age of study participants ($r = .78$; $P < .001$), and with the duration of PGES ($r = .4$; $P = .033$). After inclusion of the duration of PI and age into a multivariate model, the association with the duration of PGES became non-significant. No relations were observed between the duration of post-ictal ACC silence and seizure duration, duration of the clonic phase, and duration of the tonic phase. The EDA signal was available for 16 of 20 seizures, whereas in 4 seizures the signal was either not recorded or corrupted. A post-ictal EDA response was observed in 13 of 16 seizures (Figure 4.8c+d). The duration and amplitude of the EDA response were not related to the duration of the post-ictal ACC silence (Table 4.4) or to the duration of the PGES (duration of EDA $P = .43$; amplitude of EDA $P = .67$). These results were also con-

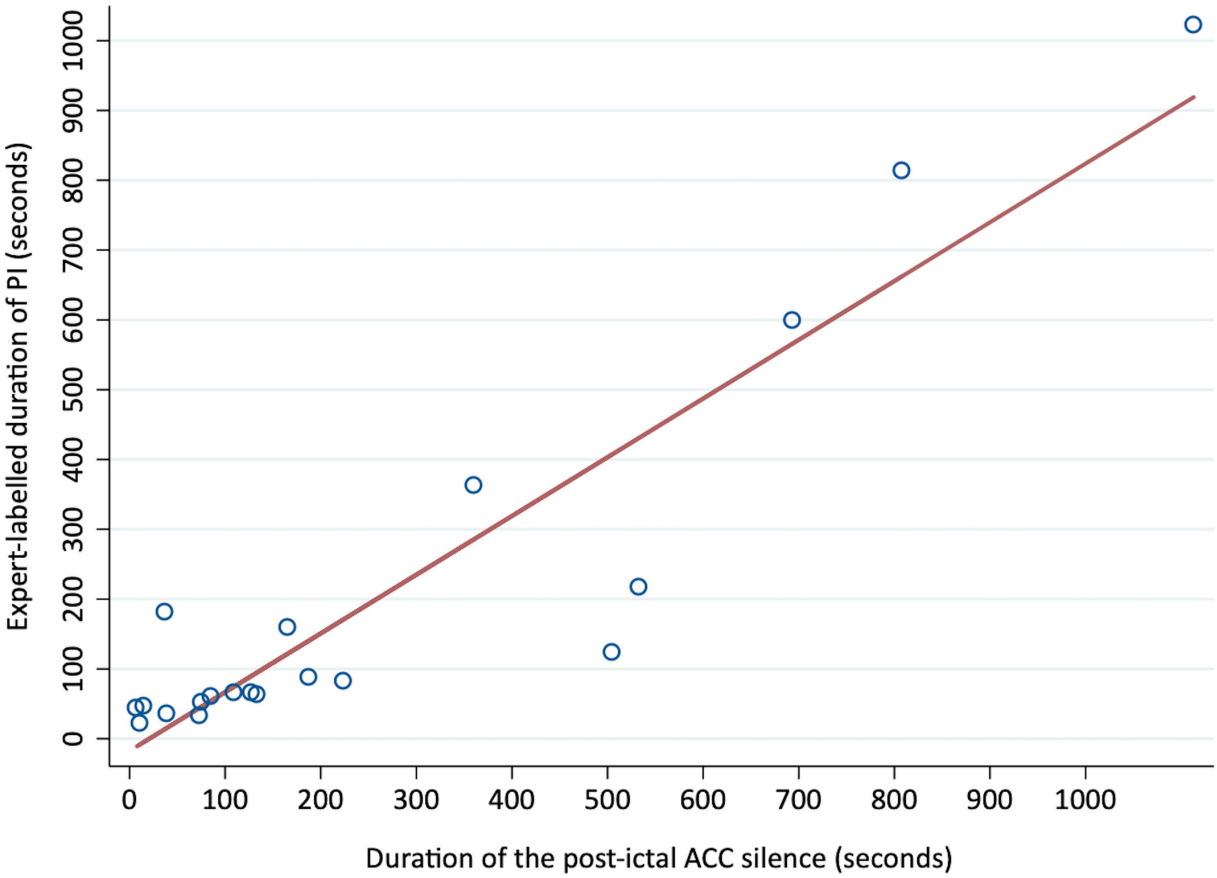


Figure 4.9: Correlation (Pearson correlation coefficient $r = 0.92$; $P < .001$) between the duration of the post-ictal ACC silence estimated using the algorithm and the duration of the expert-labeled PI. The red line represents the linear regression.

firmed when the analysis was restricted to data recorded with the wrist-worn device only. The relationship between the duration of post-ictal ACC silence and categorical seizure-specific variables were also analyzed. The post-ictal ACC silence was longer in seizures followed by PGES (adjusted $P = .05$), in seizures presenting a higher risk of SUDEP (PGES $\geq$20 s; adjusted $P = .037$) and in seizures showing an EDA response (adjusted $P = .038$), the latter also showed longer PGES durations (adjusted $P = .004$). There was no association with convulsive seizure type and number of AEDs taken, whereas a borderline significance was found in seizures originating from the temporal lobe (adjusted $P = .059$) and seizures arising from sleep (adjusted $P = .24$).

4.2.4 Discussion

ACC sensors built into consumer electronics such as smartwatches have been widely used for the identification of convulsive seizures at rest [134], additionally demonstrating good correlation with seizure motion duration [224]. This study demonstrated a novel application of wearable ACC. In a population at high risk of SUDEP, represented by patients with refractory epilepsy, potentially candidates for epilepsy surgery, and with a large majority reporting more than three or more convulsive seizures per year, wearable ACC was an accurate digital marker of PI and of its duration after convulsive seizures. Both these functions certainly

assume a great clinical value, providing information to identify and characterize potentially life-threatening seizures. In fact, although immobility following a convulsive seizure is a frequent post-ictal event, the occurrence of PI has been regarded as a precipitating factor for post-ictal cardiorespiratory dysfunction in recorded SUDEP cases [303, 323–329]. In the same context, the duration of PI has been considered as a factor contributing to the lethality of some convulsive seizures.

There is an urgent need for appropriate markers to delineate individual risk of SUDEP and to track the evolution of risk factors that might predispose to SUDEP over time [331]. The possibility of automatically gathering information on the presence of dangerous post-ictal states, such as prolonged immobility, is a step forward in this direction.

Alongside its role as a marker of PI, the post-ictal ACC silence was also correlated with other known SUDEP risk factors. The duration of the post-ictal ACC silence was correlated linearly with the age of study participants. This finding is of interest as age is known to affect the occurrence of seizure-related autonomic responses [332–334] and may play a role in autonomic dysregulation-supported phenomena, such as SUDEP. Notably, the average risk of SUDEP is age dependent, ranging from $0.2/1000\,\mathrm{PY}$[1] in children to $1.2/1000\,\mathrm{PY}$ in adults [19]. However, the correlation of PI with age has not been explored and additional investigations and larger samples, including children with epilepsy, are required to confirm this observation.

A long post-ictal ACC silence indicated the presence of seizures followed by PGES $\geq 20\,\mathrm{s}$ ($P = .037$), carrying a higher risk of SUDEP [327], and a linear correlation was highlighted between the duration of post-ictal ACC silence and the duration of PGES ($P = .033$). However, the moderate coefficient found ($r = .4$) suggests that a nonlinear relation may exist between these variables, and different models should be investigated in larger data sets. Moreover, the linear correlation disappeared in the multivariate model, probably due to the presence of a correlation between PGES and PI durations. These findings are consistent with, and replicate, previous studies analyzing the relationship between PI and PGES, where seizures associated with PGES had a longer duration of immobility as compared to those not followed by PGES [318, 319].

No association with seizure duration and duration of the clonic or tonic phase was found. The lack of association between PI and seizure duration was first reported by Gowers [335] in 1881 and confirmed in later studies [320, 321]. With some exceptions [322], additional observations demonstrated no correlation with either the duration of the convulsive phase or the tonic phase of the seizure [321].

Of interest, prolonged post-ictal ACC silence and PGES were observed in seizures showing an EDA response. However, there was no clear relation between the amplitude of the EDA and the duration of the post-ictal ACC silence or the duration of PGES, nor was higher amplitude of EDA observed in seizures associated with PGES $\geq 20\,\mathrm{s}$, as reported elsewhere [153]. The surge of EDA in the peri-ictal period is an index of sympathetic over-activation [150] that may be relevant in the pathogenesis of SUDEP [43, 153, 336].

As hypothesized in previous literature [331], the findings presented here reinforce the idea that the combination of multiple biosignals, such as post-ictal ACC silence and EDA, could increase the possibility of identifying seizures, and likewise PWEs potentially carrying a higher likelihood of seizure-related mortality. These findings highlight the importance of an automatic assessment of PI, which may be particularly relevant for nocturnal seizures, which are often unnoticed.

[1] Patient-Years

4.2.5 Limitations

This study is limited in several aspects, which are highlighted in the following. The small sample of patients and seizures may limit the generalizability of findings, which need to be interpreted with caution and replicated in larger cohorts and in real-life settings. Nevertheless, the small cohort included here is certainly of interest given that it is potentially at higher risk for SUDEP due to its characteristics. Moreover, the automatic detection of PI presented here depends on the prior marking of the preceding convulsive seizure. Thus, in a real-world system, this methodology would need to be preceded by an independent automatic seizure detection. This limitation is further explored in Section 4.3.

PI was defined as the post-ictal absence of movements, allowing for respiratory movements that are not detectable with wearable ACC and cessation of which is relevant for possible SUDEP events. However, the evidence on how PI contributes to SUDEP is uncertain and not exhaustive. It is likely that the absence of body movements represents an "early stage" of total cessation of movements (including respiratory movements), and that the automatic identification of such an "early sign" (alone or, preferably, in association with tools to detect hypoxemia) may be clinically significant.

Seizures were recorded with two different devices. Although this may have influenced the data recorded and in particular the EDA response, which is more easily detected at the wrist as compared to the arm, the ACC biosensors had similar characteristics and the signals obtained during periods of movement and immobility were comparable despite the different position of the device on the upper limb. False detection is a potential major weakness of wearable technology applied to both seizure and seizure-related phenomena detection. The trade-off between detection benefits and potential false detection should always be accounted for when dealing with digital technologies. Scenarios where a device has fallen from the patient during a seizure due to an incorrectly fastened wristband, or where at the end of the seizure the patient is lying on the limb to which the device is attached causing abnormal measurement noise, may occur, although they were not specifically observed in this cohort. Despite being possible, these events are probably infrequent, making the beneficial effects of PI detection greater than the nuisance of potential but rare false detections. ACC false detections may be mitigated by the simultaneous use of video and automated analysis. Video has been used as a sensitive way to quantify movement [269] and could enable remote detection of PI. However, although feasible especially in seizures happening from sleep (which are considered at higher risk of SUDEP), automated video detection may present disadvantages. In a mobile patient, video can be as uninformative as wearable ACC if the patient leaves the predefined space where the camera is placed and, at night and during sleep, when patients are covered by blankets.

The duration of post-ictal ACC silence sometimes lasts beyond the PI labeled by experts. According to the definition of PI used here, experts considered subtle movements as "PI end". However, these movements were mainly hand, finger, or neck movements, which are certainly insufficient for an adequate body repositioning after a seizure. It is then likely that the duration of post-ictal ACC silence is a better indicator of the duration of dangerous post-ictal states during which the patient is still unable to move. Nurse intervention was prompt in many of the events recorded and has likely interfered with the cascade of post-ictal events observed. PI in SUDEP cases has often been observed in combination with a prone position in the bed [303, 323–329], a variable that could add a layer in the assessment of SUDEP risk. However, none of the patients included in this study were prone due to nurse intervention, preventing the assessment of usefulness of the gyroscope in combination with ACC to identify the body po-

sition post-ictally. Conversely, the passive movements produced during nurse intervention had no impact on the detection of PI via ACC, as they consisted mainly of assisting the patient into the recovery position (body rotation) or in repositioning the head, resulting in either non-detectable movements or short, rapid accelerations ($<5\,\mathrm{s}$).

In conclusion, PI is one of the most common seizure-activated phenomena, often associated with negative outcomes, which this study shows to be easily monitored via new technologies at different body sites. In combination with other remote measures and paired with an automated identification of convulsive seizures, the detection of the post-ictal ACC silence could be regarded as a risk assessment tool in individual seizures, as a way to monitor disease progression and evolution and, possibly, as a potentially modifiable outcome when assessing the impact of preventive measures, interventions, and surveillance systems.

4.3 Summary

This chapter investigated the detection of convulsive tonic-clonic seizures and related clinical manifestations. Convulsive seizures such as generalized tonic-clonic seizures and focal to bilateral tonic-clonic seizures are what popular culture typically refers to when portraying an epileptic seizure, and they are among the most high-profile and dangerous seizure types in terms of their semiology. This, on the other hand, also potentially makes them the most straightforward to detect using biosignals captured by typical wearable sensors. High-amplitude and high-frequency movements are represented in accelerometry (ACC) signals as oscillations in all three axes with the selfsame characteristics. Changes of the autonomous nervous system that frequently occur peri-ictally in these seizures are reproduced in the traces of electrodermal activity (EDA) and blood volume pulse signals, although the latter are rarely usable as the strong motion of the body during tonic-clonic seizures (TCSs) also induces heavy artifacting in the raw photoplethysmography sensor signal.

The first part of the chapter presents a convulsive seizure detection pipeline, employing an ensemble machine learning model (gradient boosted decision trees) trained with features from ACC and EDA biosignal data. From a data set of 10 participants from two clinical cohorts with a total of 21 recorded TCSs, the methodology could detect $91\,\%$ of the seizures in an out-of-sample test set. Adding additional data from participants without recorded seizures to the test set, thus amounting to 78 days of wearable data, the model produced a false alarm rate of $0.19/24\,\mathrm{h}$. The study accordingly shows that supervised machine learning can achieve a high sensitivity and low false positive rate in detecting convulsive TCSs.

The second part of the chapter highlights a specific clinical manifestation of convulsive TCSs, post-ictal immobility (PI), a period of unconsciousness occurring after such a seizure that can be a major risk factor regarding sudden unexpected death in epilepsy (SUDEP). An automatic detection of this period could facilitate an alarm system, potentially alerting caretakers that the patient is unconscious after a seizure and needs immediate care. In 20 cases of PI from 18 study participants the heuristic algorithm based on ACC signals was able to identify all events, showing significant correlation between the detected and the true length of the PI period. The continuous detection of risk factors associated with seizures is of great clinical importance, and the study shows that wearables may be a useful tool for this task.

The two studies, while at first glance are both focused specifically on convulsive TCSs, are not immediately compatible. The identification of PI is based on prior knowledge of the occurrence of a convulsive seizure, and so a potential SUDEP warning system would need to be extended by an automatic seizure detection. The detection methodology presented in the first part is however not entirely applicable as the detection delay caused by the choice of features from the multimodal biosignal data makes it impracticable for this purpose. Most essentially, the long-term feature calculation window of $5\,\mathrm{min}$ into the future for the EDA features introduces a delay that is likely to be longer than the seizure itself. The long EDA feature window duration was chosen to meaningfully represent the long-term response of the signal after a seizure, with a drop of the signal down to a baseline typically occurring over the span of multiple minutes. As such, the detector would only recognize the seizure when the PI phase has already begun, or possibly even after it. A convulsive seizure detector as a pre-stage for a PI warning system would need to feature a detection delay of at most a few seconds to be useful. While not universally reported, some related studies also include detection latencies for their methodologies. Onorati et al. [224], for example, report latencies of $\approx 30\,\mathrm{s}$ for their

TCS classifiers based on ACC and EDA data, and Milosevic et al. [337] cite a median latency of $10.5\,\mathrm{s}$ over 22 nocturnal TCSs from 7 patients, based on ACC and electromyography data. For the detection methodology presented here in Section 4.1, some further feature engineering work and subsequent model optimization would be necessary to transform it into a system suited for a SUDEP warning system. In particular, the EDA features would need to be changed, specifically the long window duration as mentioned above. This would likely involve taking into account only the quick response to a peak value and not the slow minute-long decline back to a baseline. That is to say, such a lower-latency detection system seems to be within the realm of possibilities, but ultimately was not part of the research included in this thesis.

Overall, the research presented here shows that multimodal detection of convulsive TCSs is not only possible, but useful and necessary to tackle ultra-long-term monitoring of patients with epilepsy. The studies furthermore demonstrate that even very limited data sets can enable simple classical supervised machine learning algorithms to robustly detect convulsive seizures with high sensitivities and low false alarm rates. Yet, the automatic detection of focal onset seizures is another matter and requires separate analysis, introduced in the next chapter.

CHAPTER 5

Detection of Focal Onset Seizures

Focal seizures, meaning those seizures with a localized onset in only one hemisphere of the brain, are the most prevalent type of seizure, and at the same time the most heterogeneous in terms of seizure semiology. Thereby, they are typically also less convulsive or do not exhibit movement symptoms at all, and are thus considered harder to detect automatically by the kind of wearable biosignal data used here. In this thesis, and the studies included in this chapter, only those focal seizures are involved that encompass motor symptoms, specifically tonic or clonic movements of the limbs. Other focal seizure types, particularly those without any movements, require separate analysis and potentially the use of other combinations of biosignals.

This heterogeneity of focal seizures is highlighted in the first part of this chapter, which comprises an explorative investigation of different types of focal motor seizures recorded from three distinct patients with epilepsy. While this study has limited significance in terms of evaluating new focal seizure detection methodology, it gives a meaningful and deliberate insight into potential uncertainties these seizures hold regarding biosignal data. Chronologically it was the first major analysis work in the context of this thesis, and the data collection studies were still ongoing at that time.

The second part of the chapter, in turn, features the chronologically last major analysis work for this thesis, building upon the experiences of the other research conducted in the meantime. It investigates the feasibility of a detector for focal motor seizures both in an individualized context and across patients. To that end, the presented methodology modifies the existing detection pipeline from Section 4.1, adding blood volume pulse as a modality and selecting features specifically suitable for focal motor seizure detection.

5.1 Exploratory Analysis of Focal Motor Seizures

[106] ⇒ **Böttcher, Sebastian**, et al.
Using multimodal biosignal data from wearables to detect focal motor seizures in individual epilepsy patients
2019, Proceedings of the 6th International Workshop on Sensor-Based Activity Recognition and Interaction, DOI:10.1145/3361684.3361687

Parts of this publication were removed or edited to fit into the composition of this complete thesis. No substantial changes altering the results were made.

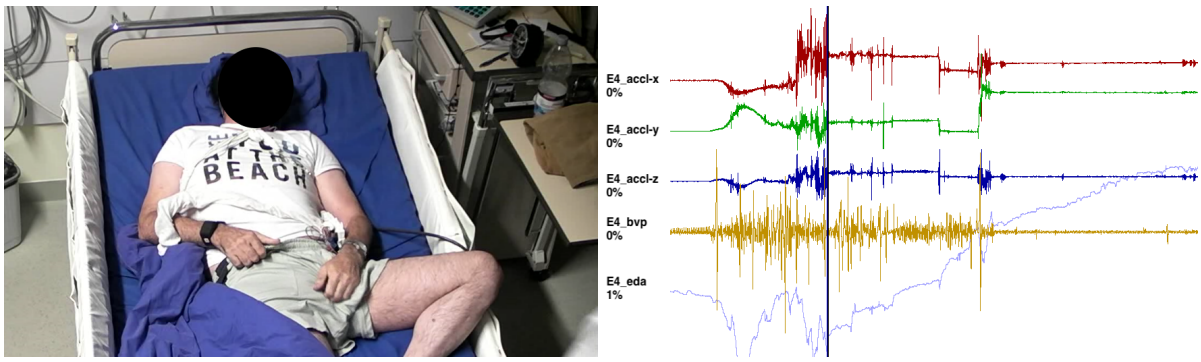
Own Contributions:

- All

5.1.1 Introduction

The few wearable devices that thus far have been used in epilepsy research are most commonly smartwatch-like devices or fitness trackers that record biosignals such as accelerometry (ACC), electrodermal activity (EDA), blood pulse via photoplethysmography (PPG), and electromyography (EMG). These biosignals have been shown to give sufficient indication towards epileptic seizures, with research focusing on monomodal and multimodal detection of generalized tonic-clonic seizures (GTCSs). GTCSs are one type of seizure that involves both hemispheres of the brain and present themselves in violent bilateral muscle contractions of the whole body. These are very different from focal seizures (FSs), which start in only one brain hemisphere and can present themselves in a number of different symptoms that are far harder to characterize. Other research has explored the detection of FSs with wearable data, however a majority of these efforts have focused on detecting a specific type of FS only, often by using a single modality.

This work proposes a multimodal approach to detect FSs, which has thus far been a new and underexplored avenue in epileptic seizure detection. It offers a first analysis into the challenges that lie ahead, especially in the analysis of the various subtypes of FSs and the implications this holds for classification tasks in particular. In the following, the current state of the art in epileptic seizure detection with wearables is explored, followed by the introduction of a new data set of biosignal data from wearables worn by three in-hospital patients that were monitored with video-electroencephalography (vEEG) in a medical epilepsy monitoring unit, along with wearable sensors (Figure 5.1). This study focuses specifically on showing the difficulties that may arise when implementing a multimodal seizure detection pipeline for variable types of seizures, using common biosignals such as ACC, EDA and PPG. The detection of seizures from three selected patients is evaluated, and the results are analyzed. Concluding, an outlook on the development of the detection pipeline is given.



(a) (b)

Figure 5.1: (a) A video frame from the vEEG epilepsy monitoring unit at University Medical Center Freiburg (UKF) during a patient's focal motor seizure, **(b)** the time series over a $5\,\mathrm{min}$ segment from the patient's right wrist. The time series shows from top to bottom: 3-axis accelerometry in x/y/z, blood volume pulse, and electrodermal activity, with the video frame's timestamp marked by the black line.

5.1.2 Related Work

This section on research work in the detection of epileptic seizures is structured along the two main types of seizures, GTCSs and FSs, as most research to date has explicitly focused on one or the other. Monitoring these two types of seizures has also shown to require very different modalities. Refer to Chapter 2 for a more extensive look at related work in the field of wearable biosignal monitoring and epilepsy.

Due to the severe manifestation in body and especially limb movements, GTCSs are relatively straightforward to detect using standard wearable biosignals like ACC or EMG. Moreover, GTCSs are a significant risk factor in sudden unexpected death in epilepsy, raising interest in the automatic detection of this type of seizures, especially in an ambulatory setting [338]. There are various examples of GTCS detection in literature, both with monomodal and multimodal data (see also Chapter 4).

One basic approach is evaluated by Kusmakar et al. [215] who use accelerometry data from a wrist-worn wearable to detect short-length GTCSs in 12 patients. Their approach with a support vector method and standard time domain features achieves a sensitivity of $95\,\%$ and false alarm rate (FAR) of $0.7/24\,\mathrm{h}$. Halford et al. [339] on the other hand use an upper arm wearable that records surface EMG signals on 199 patients with epilepsy. Their thresholding method detects $76\,\%$ of overall GTCSs, with a FAR of $2.5/24\,\mathrm{h}$. However, they also distinguished between properly and improperly placed devices, reporting that among properly placed devices $100\,\%$ of GTCSs were detected with a FAR of $1.4/24\,\mathrm{h}$. They conclude that proper placement of the device is important. EDA and ACC signals are used by Poh et al. [153] to detect $94\,\%$ of GTCSs in a data set from 80 patients, with a FAR of $0.7/24\,\mathrm{h}$. More recently, Regalia et al. [293] also used EDA and ACC signals to detect GTCSs, attaining a sensitivity of greater than $92\,\%$ with a FAR between 0.2 and 1 per day on varying data sets of inpatient and outpatient studies.

Also known as partial seizures, FSs are seizures that have their source in one of the brain's hemispheres, as opposed to GTCSs which spread over both. FSs are therefore usually not accompanied by severe motor reactions of the body like in GTCSs, but rather manifest in

a multitude of different symptoms: These can include autonomous reactions like heart rate increase (tachycardia), dyscognitive features like impaired awareness or unconsciousness, less severe motor components, or so-called auras, which are sensory phenomenons such as déjà vu sensations or dizziness. During the course of one FS, multiple of these symptoms may occur consecutively or simultaneously.

In literature, the detection of FSs with wearables has been attracting more attention in the recent past. Some research studies have considered single modalities to detect FSs of specific types. Jeppesen et al. [237] look at heart rate variability from electrocardiography (ECG) for 17 patients, detecting $74\,\%$ of seizures with their method. Poh et al. [153] on the other hand use an EDA sensor to analyze autonomic changes during and especially after FSs and GTCSs, concluding that the EDA response after GTCSs is much more severe and prolonged than in FSs. A different approach is taken by Vandecasteele et al. [74], who compare wearable ECG and PPG devices in the detection performance of temporal lobe epileptic seizures, which are a type of FS. They report sensitivities of $70\,\%$ for ECG and $32\,\%$ for PPG detection, with FARs of $2.1/24\,\mathrm{h}$ and $1.8/24\,\mathrm{h}$, respectively.

Recently, some studies have also expanded to multimodal detection of FSs. Cogan et al. [227] propose a multistaged detection system that uses heart rate, arterial oxygenation, ACC, EDA and temperature data, detecting $100\,\%$ of FSs in the sensor readings from nearly all 10 patients their data set consists of. However, they do not specify further what type of FS they worked with, only referring to the detected seizures as complex partial, an older term for focal seizures. The work presented here is most comparable with that of Onorati et al. [224], who use EDA and ACC data from 69 patients to detect GTCSs as well as FSs, comparing three different classification methods. Their best performing method reaches a sensitivity of $95\,\%$, with a FAR of $0.2/24\,\mathrm{h}$ and an F-score of 0.67 in cross-validation. However, their data set of 55 convulsive seizures only includes six focal seizures.

Among the above research works, there are several studies in literature that present monomodal and multimodal seizure detection on large data sets, however, these are often very generalized in what seizure types are included. The distinction between GTCSs and FSs is often made, but subtypes within FSs are rarely investigated or separated in the annotation. The work presented here illustrates the breadth of FSs by focusing specifically on the multimodal detection of three distinct types of focal motor seizures in three individual patients and therein identifies difficulties that may arise when analyzing a larger data set of focal seizures. Furthermore, this work explores a way of feature extraction that enables using multimodal data with multiple different window sizes per modality. In other studies only a single window size with a fixed overlap per modality is commonly used.

5.1.3 Data Set

The evaluation presented here uses selected data that are taken from a data collection clinical study. In this study, epilepsy patients that were continuously monitored in epilepsy monitoring units at two study sites were recruited and asked to wear a wearable wristband device, the Empatica E4 (Section 3.2.1; Empatica Inc., Boston, MA, USA), and a wearable upper-arm device, the Biovotion Everion. Both sensor units record ACC ($32\,\mathrm{Hz}/50\,\mathrm{Hz}$), EDA ($4\,\mathrm{Hz}/1\,\mathrm{Hz}$), and PPG ($64\,\mathrm{Hz}/50\,\mathrm{Hz}$) data continuously. The goal of the study was to capture a variable set of seizures for a population of at least $N = 200$ patients, while recording at least one seizure

Table 5.1: The selected participants for this evaluation, and the respective amount of seizures recorded. Seizure type are the most common types among n, where seizures can have multiple subtypes. All seizures have the "motor" subtype, indicating motor components during the seizure.

P	n	Data	Seizure Type	Comment
1	6	ACC/EDA	FS motor; FS auto.; FS dyscog.	Characteristic motor seizures with tonic/clonic arm movement
2	9	ACC/EDA	FS motor; FS auto.	Motor seizures with automatisms (most with arm movements)
3	7	ACC/EDA	FS motor; FS auto.; aura	Motor seizures with only automatisms (few with arm movements)

auto.: autonomic components, like tachycardia; dyscog.: dyscognitive components, like loss of consciousness; aura: aware seizure, usually with a specific associated feeling.

for $M = 96$ patients. The study is divided across two clinical sites, the King's College Hospital London and the UKF. Chapter 3 further details the study procedures.

The ground truth for seizure labeling is provided by a clinically trained expert, who scans through the vEEG recordings from the epilepsy monitoring unit and manually marks seizure onset and offset, as well as timings of various seizure phases, such as tonic movement, clonic movement, tachycardia, or unconsciousness. Patients are typically recorded for stretches of 5 to 7 days, and tend to suffer from any type of epilepsy. At the time of the writing of this part of the thesis, the data set included data from 174 patients from both sites, with 276 complete seizures recorded from 70 patients, respectively.

Selection of three cases

For the preliminary evaluation presented here, only a select set of participants in the study is considered from a single site, that is, three representative cases that are to be investigated. For each of the three patients in question, more than five FSs with varying types of motor components were recorded, and these patients were selected for their different seizure manifestations: One patient exhibited highly characteristic tonic arm movements that are clearly distinguishable on the raw ACC signal. The second patient had predominantly automatisms in their arms, which are often random movements of the limb that can be classified as neither tonic nor clonic. The third patient also had automatisms, however these were not located in the arms, but rather – more challenging for seizure detection – in the legs and also mouth region (oroalimentary). Among the three selected patients, a total of 22 seizures had motor features, and most of them also had autonomic or dyscognitive features, or auras associated with them. Table 5.1 gives a brief overview of the main characteristics for the three selected patients. The following evaluation highlights the ACC and EDA data from the Empatica E4 device.

Feature Set

In order to be able to train a supervised model with multimodal data with differing sample rates, the mixed modality feature set used here has variable window lengths per feature, but at

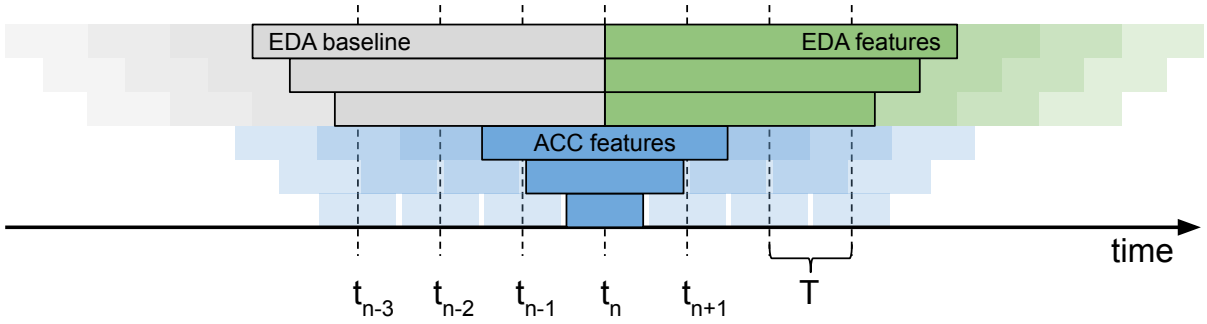


Figure 5.2: Graphical representation of the mixed EDA and ACC feature set used in the evaluation, with fixed feature points t_n and $T = 1\,\mathrm{s}$.

fixed time intervals $T = t_{n+1} - t_n$. Thus, in this approach the window lengths and interval size are the determining factors of the feature set, contrary to the usual method of defining window lengths and overlap. Since it is unclear from existing literature what window lengths are best for specific modalities for epileptic FS detection, this approach facilitates testing several window lengths at the same time for later analysis of the best combination of features and window lengths. More specifically, the resulting tables of features will have values at the same time points for all modalities and all window sizes, allowing the feature data to be concatenated into one table for model training. Figure 5.2 shows a graphical representation of this mixed feature set.

The feature set for this evaluation consists of 141 ACC features from the time and frequency domains (divided into 40 subgroups when grouping together x, y, z, and total features), and an additional 10 EDA features, some of which are also corrected for a baseline. With window lengths of $2\,\mathrm{s}$, $10\,\mathrm{s}$ and $20\,\mathrm{s}$ for the ACC features and $5\,\mathrm{min}$, $10\,\mathrm{min}$ and $20\,\mathrm{min}$ for the EDA features, a total of 453 features were calculated for each fixed time point. Note the large difference in window lengths for EDA vs. ACC; Since the EDA signal is primarily analyzed for tonic activity features, and the time frame of change in EDA signals is in the order of minutes, the window lengths for EDA were chosen like this. Furthermore, for this evaluation the time interval between feature points is fixed at $T = 1\,\mathrm{s}$. To avoid feature intervals with undefined values, feature extraction is only done on sections of the data where all modalities are present, that is, where there is a data point in all modalities for a given timestamp.

5.1.4 Evaluation

For evaluation, the described feature set was divided into sets per seizure per participant, each containing the feature data from the time interval $[s_{start} - 55\,\mathrm{min}; s_{end} + 55\,\mathrm{min}]$, where s_{start} and s_{end} refer to the seizure start and end, respectively, as labeled by a clinically trained expert via the vEEG recordings. One seizure data set thus consists of $55\,\mathrm{min}$ before the seizure start and after the seizure end, as well as the duration of the seizure itself, which for the 22 seizures of the three selected patients had a mean of $1\,\mathrm{min}\ 55\,\mathrm{s}$. Thus, one seizure accounts for approximately $112\,\mathrm{min}$ of data, amounting to approximately $41\,\mathrm{h}$ of data for all seizures of the three selected patients. The $55\,\mathrm{min}$ margin was chosen due to the large EDA window sizes and characteristically long EDA response time. Typical EDA response times can last up to an hour after the actual seizure has ended [153].

Table 5.2: Mean results (after 20 repetitions) of leave-one-seizure-out cross-validation using a RF model ($t = 50$). Shown are precision (p), recall (r), and F1-score (f) for both sets of experiments and the three patients highlighted in this study (P1, P2, and P3).

ID	First Experiment			Second Experiment		
	p_1	r_1	f_1	p_2	r_2	f_2
P1	0.78	0.5	0.56	0.92	0.7	0.77
P2	0.73	0.7	0.71	0.72	0.72	0.71
P3	0.29	0.22	0.19	0.42	0.28	0.28

On these seizure sets per individual patient, binary leave-one-seizure-out cross-validation was performed, using a random forest (RF) model with $t = 50$ trees. The evaluation was done sample-wise, that is, each time point is classified as either belonging to a seizure or not. The cross-validation for each patient was repeated 20 times to give a confident idea of the RF model performance. Additionally, a second round of tests was done, where the interval for the feature data now is $[s_{start} - 55\,\mathrm{min}; s_{end}] + [s_{end} + 5\,\mathrm{min}; s_{end} + 55\,\mathrm{min}]$, thus excluding data from detection for $5\,\mathrm{min}$ after a seizure was already recognized. This can be seen as the simulation of a post-detection pause of data analysis, which prevents false positive detections resulting from large uncertainty in data following immediately after a seizure.

Before the scoring of the tests, the predicted labels are smoothed by a hysteresis function with a threshold of $10\,\mathrm{s}$. Effectively, this means all consecutive positive predictions of less than $10\,\mathrm{s}$ are disregarded, and all consecutive negative predictions of less than $10\,\mathrm{s}$ within a larger positive block are still regarded as positive. The results for all tests can be found in Table 5.2 and will be discussed in the following.

5.1.5 Discussion

Since this is a sample-wise cross-validation, the scores from this evaluation give an overview of the performance of the RF model with respect to the classification of seizure status for each second in the test data sets; As opposed to event-based classification which would give an overview of the performance with respect to the classification of overall seizure events. Furthermore, since for this sample-based evaluation the train and test sets are highly unbalanced, only precision, recall and F1-score are regarded as measures. The imbalance derives from the choice of seizure data set, that is, data for one seizure includes $55\,\mathrm{min}$ of negative data before and after the seizure that typically has a length of $<5\,\mathrm{min}$, with a mean of $1\,\mathrm{min}\ 55\,\mathrm{s}$ for the selected seizures, or $1.7\,\%$ compared to negative data.

The performance for P1 is acceptable in data set 1, without the simulation of a post-detection pause, and increases considerably in data set 2, where $5\,\mathrm{min}$ of data are cut off after a seizure. This behavior is expected, as there often are detections right after the seizure in data set 1, as can be seen in Figure 5.3. Especially the precision score is affected by this, as primarily false positive classifications after a seizure are avoided. For P1 the recall score also improves substantially with data set 2, showing that the detection of clear and characteristic motor FSs may benefit the most from this method.

Contrarily, for P3 recall scores improve less than precision. Overall however, the seizures of P3 are not detected as reliable as those of P1. This is expected considering that those seizures are not extensively represented in the motion data, due to the seizure manifestation in automa-

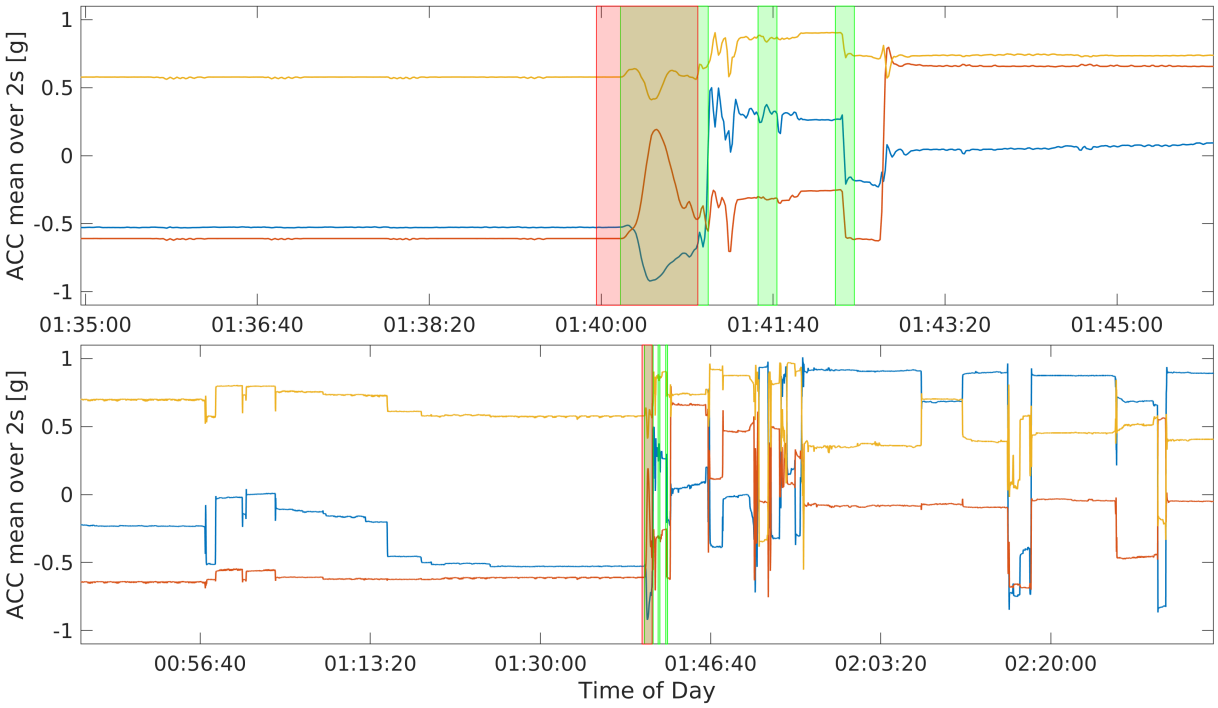


Figure 5.3: Example of the recognition performance on one seizure of P1, from a single cross-validation run on data set 1. **(top)** $5\,\mathrm{min}$ interval before and after the seizure. **(bottom)** Full data set of the same seizure including $55\,\mathrm{min}$ of data before and after, showing that the only positive predictions are in the immediate vicinity of the ground truth. Data series are the ACC means of each axis over a $2\,\mathrm{s}$ window. Overlay areas in red depict the ground truth, those in green mark the predictions. Remaster of Figures 3 + 4 in Böttcher et al. [106].

tisms occurring in the opposite arm that the device was attached to. There is no substantial change in scores for P2 after post-seizure classification pause, while the overall scores of that patient are comparable to those of P1. This may indicate that for the type of automatisms this patient was exhibiting, classification is invariant to post-seizure uncertainty, or – alternatively – that there is none.

Looking at the predictions from the point of view of time series event recognition, most predictions are in immediate time proximity to the seizure ground truth, with only few false positives. Figure 5.3 for instance shows that while there are some false predictions immediately after a seizure, the rest of the $55\,\mathrm{min}$ before and after the seizure is, correctly, free from seizure predictions. The results in Table 5.3 were attained from examining a single run of the cross-validation for all patients and counting the event-wise true and false positives. In these results, a true positive is any ground truth event that overlaps with a predicted event, and a false positive is any predicted event that does not overlap with a ground truth event. The FARs for each patient are rough estimates that were obtained by counting the false positives over the whole seizure data set: $FAR_{est} = n_{fp}/(112\,\mathrm{min} \cdot n_{sz})$. These results show that even this basic RF approach can already reach a performance comparable to that of related literature when regarding only sensitivity, while estimated FARs are still high, but may in reality be lower when testing on a whole patient recording.

However, the purpose of this study is not to compare performance to the current state of the art, but to analyze the challenges in the multimodal classification of epileptic focal seizures,

Table 5.3: Event-based results for a single run of the cross-validation. FARs are rough estimates, calculated from the number of false positive (FP) over the whole duration of the seizure data sets combined, for each patient.

ID	TP (%)	FP	estimated FAR
P1	6/6 (100)	2	$4.3/24\,\mathrm{h}$
P2	7/9 (78)	2	$2.9/24\,\mathrm{h}$
P3	3/7 (43)	11	$20.2/24\,\mathrm{h}$

and specifically those with motor features. These motor features can manifest themselves in many different ways. As the selection of patients in this study shows, there are motor features that are not captured by data from a single wearable. Even with multimodal ACC and EDA data, seizures that manifest themselves, for example, in a limb that the wearable is not directly attached to may be missed. While an additional modality like features from PPG may help with this, it is essential that the wearable collecting data is attached to the body part that the seizures are most predominantly located in, with respect to individual patients.

Post-seizure movement is another factor that makes some FSs difficult to detect accurately. Especially in a hospital environment patients may move in a way that makes accelerometry models less accurate, for example, due to nurse intervention. A further modality next to ACC may help with this, but the large timeframe in which EDA changes happen make it less ideal for that specific purpose. One way to counteract this is also to stop looking for seizures for some time after one was already detected, which is shown here to help with detection accuracy. Furthermore, multiple detections that are located within a certain time frame should be counted as one event, to reduce false alarm rates.

Lastly, due to the nature of epileptic seizures and their infrequent occurrence, the available data are highly imbalanced towards the negative class. The evaluation shown here tries to alleviate this problem somewhat, by segmenting out the seizures within a certain time interval. Yet, other measures could be taken to counteract the imbalance. For example, data during sleep may be cut out by looking for periods of very little activity in accelerometry data. In the end however, this problem remains somewhat unsolved in seizure detection with wearable data.

5.1.6 Conclusions

This study presents findings that focus on the multimodal detection of FSs from wearable sensor data. Data from three patients with epilepsy exhibiting different types of FSs were examined, showing that they manifest very differently in both the sensor signals and classification performance measures. This heterogeneity will inherently hinder accurate recognition of any FS from wearable assessment data, and needs to be taken into account when designing a learning model for seizure detection.

While the experiments shown here are promising for further work, it is clear that this is only a first step in building a system for multimodal detection of FSs using biosignal data from wearables. The results in Table 5.2 show that while this approach may work for individual patients with characteristic motor seizure manifestation, it may not work for patients exhibiting other seizure types. Furthermore, comparing the results of the first and second experiments shows that ignoring a certain amount of time after a seizure detection can substantially improve the detection performance. While the evaluation shown here is based on sample-wise

scoring, a fully implemented system has to be based on events, that is, consecutive positive predictions must be consolidated into one seizure event, which would be scored as a hit if it has some overlap with a ground truth event. An outlook on such a system is given by the results in Table 5.3, showing the performance of the presented system when scored on an event basis.

The selection of three specific patients with focal motor seizures illustrates some core problems that a more advanced detection system needs to deal with. In the future, the cross-patient seizure detection of such a system needs to be evaluated as well. While individual-based detection is one possible approach, the need for generalized models is apparent, and current state-of-the-art moves in the direction of individual-invariant models. Furthermore, a system's performance on different types of seizures like autonomic or dyscognitive FSs needs to be evaluated. Therefore, PPG features need to be considered in addition to the ACC and EDA features already implemented. Multi-class classification of seizure types and specifically recognition of phases within focal seizures are reasonable goals that may be achieved by a detection system that takes into account the high variance in focal seizure manifestations.

5.2 Detection of Focal Motor Seizures

[107] ⇒ **Böttcher, Sebastian**, et al.
Intra- and Inter-Subject Perspectives on the Detection of Focal Onset Motor Seizures in Epilepsy Patients
2022, Sensors, DOI:10.3390/s22093318

Parts of this publication were removed or edited to fit into the composition of this complete thesis. No substantial changes altering the results were made.

Own Contributions:

- All, except clinical expertise and the data collection at the KCL site

5.2.1 Introduction

Epileptic seizures are defined by a period of abnormal neuronal activity in the brain, and are generally divided into two main groups by their neurological onset [27]. Seizures with an early bilateral involvement are called generalized seizures, while seizures with just a single point of onset are denoted as focal onset seizures, but epileptic activity can propagate across the brain resulting in "focal to bilateral" seizures with characteristic motor manifestations. Bilateral tonic-clonic seizures have been assessed in numerous studies as to the viability of wearables for the detection during recent years, and detection has been demonstrated to be feasible in multiple retrospective studies [104, 226, 227, 280, 315, 340–343]. Conversely, focal seizures in general are still a relatively unexplored field with respect to wearable non-electroencephalography (EEG) detection [74, 106, 110, 219, 235, 344]. Symptoms and manifestations of these seizures are much more heterogeneous as compared to those of bilateral tonic-clonic seizures, with some barely or not at all captured by typical wearable biosignal modalities, such as accelerometry (ACC), electrodermal activity (EDA), or photoplethysmography (PPG).

Focal seizures can be roughly divided into two categories: those with motor and those without motor manifestations. Non-motor symptoms, that is, those without involuntary movement of the body, may include partial loss of awareness or consciousness, cognitive impairment, or emotional or sensory symptoms. Motor symptoms, on the other hand, can include tonic or clonic movements of the limbs or body in general, hyperkinetic movements, or automatisms. A single epileptic seizure can thereby be composed of multiple types of manifestations, in parallel or sequentially.

This work highlights difficulties in the detection of focal onset epileptic seizures, specifically those with focal tonic or clonic motor symptoms but without bilateral propagation, from biosignal data captured by wearable devices. A data set consisting of multimodal data from a wrist-worn wearable was recorded from patients with epilepsy during their in-hospital stay at an epilepsy monitoring unit (EMU). Classical supervised machine learning is applied in order to assess the utility of this kind of data for the detection of seizures, in the context of an automated seizure diary. As written diaries created by the patients themselves have been demonstrated to be very inaccurate and often severely undercount seizures even for convulsive seizures [54, 102, 304], an automated diary tool implementing an objective seizure identi-

fication and quantification is needed, for example, as a basis for treatment decisions made by epileptologists. In the supervised methodology employed here data are first labeled, based on parallel video-electroencephalography (vEEG) monitoring, as "seizure" or not "seizure", and then processed into meaningful features and given to the machine learning model for training. Thus, the trained model can then be used to automatically classify new data. Specifically, this study uses gradient boosted decision trees (GBT) as the seizure detection model. To evaluate such an approach, and subsequently also determine the application in a real-world system, two procedures can be applied: intra-subject or inter-subject evaluation. The intra-subject evaluation focuses on the performance of the methodology when applied to data from a single patient, while the inter-subject evaluation assesses the performance over multiple patients with potentially different types of epilepsy and seizure manifestations. The former requires multiple seizures recorded per subject and will produce individualized models tailored to a single patient, while the latter requires seizures recorded from multiple different participants and will give inter-subject models, to be used over wider populations. This study aims to determine which of these approaches may work best for focal motor seizures going forward, giving guidance for the design of future studies in the field. Following the study classification suggested by Beniczky et al. [212] in 2018, the study presented here could be classified as a phase 1 retrospective proof-of-principle study. The main contribution of this work is the evaluation of supervised machine learning methodologies on focal onset epileptic motor seizures in a data set recorded from a non-EEG wearable device. A comparison between two evaluation approaches, intra- and inter-subject, provides additional context and facilitates recommendations towards future studies in the field.

5.2.2 Materials and Methods

Data Set

Data from wearable devices were recorded from a total of 243 patients with epilepsy across two EMUs in the period between July 2017 and February 2020. Both at the neurophysiological department of King's College Hospital London (KCL) (71/243 patients), and at the University Medical Center Freiburg (172/243 patients), patients in the age range of 7 to 80 with a diagnosis of epilepsy were recruited sequentially as part of their standard clinical epilepsy care, for example, in the course of standard presurgical evaluation. Patients with predominantly (suspected) psychogenic non-epileptic seizures or other involuntary movements were not included in the study. As part of their stay in the EMU study, participants may have had seizures provoked, for example, by temporary reduction of their anti-epileptic medication or through other means, such as sleep deprivation or hyperventilation techniques. The vEEG data were retrospectively reviewed and labeled by clinical experts. Primarily, they marked seizure type and semiologies, electrographic and clinical onset and offset, and other metadata including state of vigilance and body position at seizure onset. While participants wore different kinds of wearable devices, and sometimes more than one in parallel, the retrospective study presented here only includes data from the wrist-worn Empatica E4 (Figure 5.4; Section 3.2.1; Empatica Inc., Boston, MA, USA). It is a research-grade device designed specifically for epilepsy seizure detection, recording 3-axis ACC at a sample rate of $32\,\mathrm{Hz}$, EDA at $4\,\mathrm{Hz}$, skin temperature at $4\,\mathrm{Hz}$, and PPG at $64\,\mathrm{Hz}$, the latter of which was internally preprocessed into a blood volume pulse (BVP) signal. The wearable has a European Union Conformité Européenne class IIa certification as a medical device. The data recording mode used in this study

Figure 5.4: One of the Empatica E4 wrist-worn wearable devices used in this study **(left)**, and the Android phone application that connects to the wearable via Bluetooth and records the data stream **(right)**.

was the online Bluetooth streaming mode. Battery life could range between 12 h to 48 h depending on the condition of the battery. Participants were given two devices, such that one would always be recording while the other was charging. The study and recording procedures are further described in Bruno et al. [103] (see Section 3.1) and Ranjan et al. [291]. As part of the study recruitment, all participants gave written informed consent, and the study protocols and consent forms were approved by the local ethics committees (London Fulham Research Ethics Committee — 16/LO/2209; Ethics Committee at the University of Freiburg — 538/16).

Feature Set

To facilitate the detection of focal motor seizures in this data set of non-EEG wearable data, a selection of derived features are calculated from each of the three raw data modalities. These features are chosen to meaningfully represent the changes in the signal between ictal (seizure) and inter-ictal (non-seizure) phases. Each feature vector is calculated consecutively from the raw time series data at a constant interval of two seconds, regardless of the actual length of the feature window (see Figure 5.5). The choice of features in this study was informed primarily by previous research in the field. The following details the feature calculations for the biosignal modalities, ACC, EDA, and BVP.

For the ACC features, a number of parameters calculated from the recurrence plot are used as features. Recurrence plots are a statistical tool to analyze recurrence in time series data [345, 346], and have been successfully used in the detection of motor movements from accelerometer data before [104, 308, 347]. Specifically, the determinism (percentage of points that form diagonal lines of a minimal length), the Shannon entropy (probability that a line

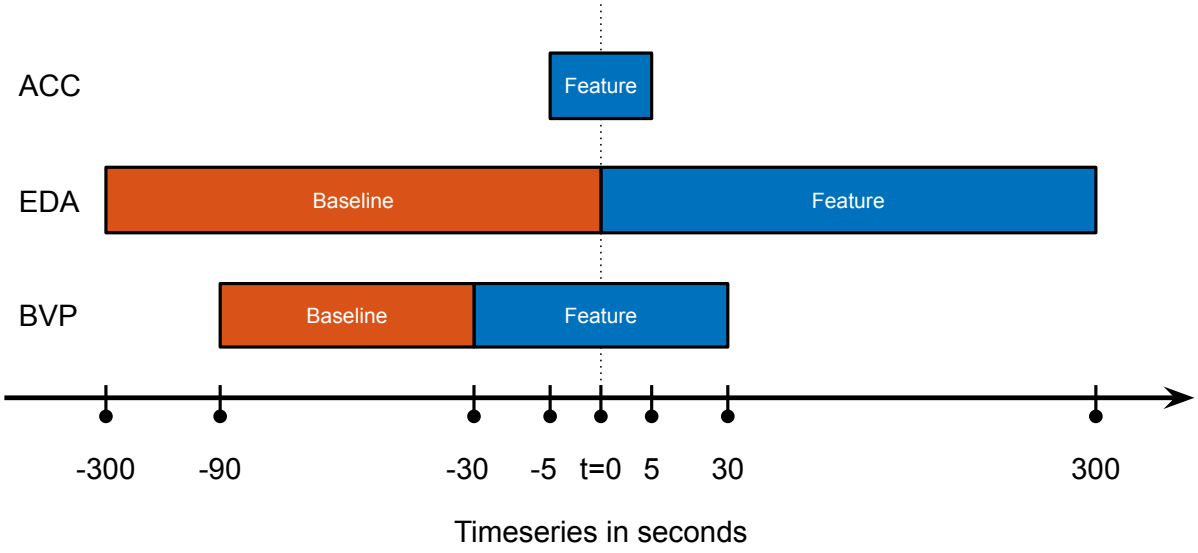


Figure 5.5: Overview of how the feature and baseline windows were chosen, for the three different groups of features by modality. This calculation would result in one feature vector, for the next the windows would all be shifted by an interval of $T = 2\,\mathrm{s}$ to the right. Abscissa not to scale.

has a certain length), the average diagonal line length, and the recurrence rate (density of recurrence points) were calculated from the recurrence quantification analysis. All of these values are derived from overlapping data windows of a length of $10\,\mathrm{s}$, centered at each $2\,\mathrm{s}$ interval.

The EDA features used here are calculated from the skin conductance level and the skin conductance response rate (SCRR) [42, 135, 150, 246]. The former is essentially a low-pass filtered version of the original raw EDA signal and thus represents the slower tonic changes in the EDA data. It is represented in the feature set by the difference of area under the curve and maximum between the five minutes before (feature window) and after (baseline) each two-second interval point. Additionally, the SCRR feature is calculated against the baseline in the same way, representing the higher-frequency phasic changes of the EDA signal. The SCRR is calculated as the number of threshold crossings of the first derivative of the EDA signal in the window.

Finally, the BVP raw data (derived device-internally from the PPG sensor) are processed to a heart rate (HR) estimation following the procedure described in Glasstetter et al. [348]: A peak tracking algorithm was applied to find local minima in the raw time series [349], and the resulting inter-beat-intervals were processed to the HR estimation employing several filters to produce a smooth and meaningful output. This HR estimation as well as a spectral entropy score representing BVP signal quality [159, 348] was used as feature values. As the BVP signal is highly sensitive to motion artifacts [173], using a signal quality index like this as a feature for classification follows the principle of regarding artifacts as additional information, instead of discarding them outright. Furthermore, this feature can be observed as a sort of indication for the quality of the model; a model that is highly dependent on the data quality of a signal may not be regarded as a particularly stable model. Additionally, the mean and maximum of

the calculated HR feature over a $60\,\mathrm{s}$ window are used as features as well, which are baseline-corrected by the difference of values between the feature and baseline window.

An overview of the different feature and baseline windows can be found in Figure 5.5. A comprehensive listing of the individual features is shown below:

1. Four features calculated from the recurrence plot of the ACC signal in a $10\,\mathrm{s}$ window:
   (a) **Determinism**, that is, the percentage of points that form diagonal lines of a minimal length.
   (b) The **Shannon entropy** of the probability that a line has a certain length.
   (c) The **average diagonal line length**.
   (d) **Recurrence rate**, that is, the density of recurrence points.
2. EDA-based features over a $5\,\mathrm{min}$ window, minus the same value in the five minutes before the feature window:
   (a) The **area under the curve of the skin conductance level** calculated as the moving mean of the raw EDA signal over a $1\,\mathrm{min}$ window.
   (b) The **maximum value of the skin conductance level** calculated as above.
   (c) The **skin conductance response rate** calculated as the number of crossings of a threshold by the first derivative of the smoothed EDA signal within the window.
3. Heart rate-based features calculated from the BVP signal:
   (a) The **local maximum of the heart rate estimation** in a $60\,\mathrm{s}$ window, minus the baseline value from the prior $60\,\mathrm{s}$ window.
   (b) The **mean of the heart rate estimation** in a $60\,\mathrm{s}$ window, minus the baseline value from the prior $60\,\mathrm{s}$ window.
   (c) The **spectral entropy data quality index** of the raw BVP signal, sampled at $2\,\mathrm{s}$ intervals.
   (d) The **heart rate estimation** calculated from the raw BVP signal, sampled at $2\,\mathrm{s}$ intervals.

Evaluation

To assess the possibilities of detecting focal epileptic seizures by wearable biosignal data, two different approaches were investigated: intra-subject and inter-subject. The distinction is an important addition to this work, as focal motor seizures have not been investigated to a degree that allows making the choice outright. While an inter-subject approach, that is, creating models that can detect seizures across a patient population without individual adjustments, would certainly be the best possible outcome, the heterogeneity of focal seizures may dictate an intra-subject approach using individualized models. To examine the effect that this might have with the given data set, the evaluation was divided into two parts.

Firstly, a subset of participants with at least three focal motor seizures recorded was isolated, and the detection model was evaluated per participant in a parameter-optimized leave-one-seizure-out (LOSO) cross-validation. As the data set did not provide data with more than

six seizures recorded for a single participant, or with multiple independent recordings of a participant, this was performed without a dedicated test set which is truly "out-of-sample". Rather, the model was trained with the data of all but one seizure and the respective peri-ictal data of 10 min before and after each seizure. These data were standardized using the z-score method before training, and the normalization parameters (centering mean and scaling standard deviation) were stored. The resulting model was then tested on the remaining participant data, standardized using the previously stored normalization parameters from the training step. This test data included the complete data set of the participant, including the left-out seizure, but not any of the data used for training the model. Nevertheless, due to the high imbalance between inter-ictal versus ictal phases, the proportion of data between the test and training set was typically far greater than 10:1. This process was repeated such that each seizure of the participant was left out once.

Secondly, for the inter-subject evaluation, the seizure data from all the participants with three or more seizures recorded, selected in the first step, were used to validate the performance on data from all the remaining participants with one or two focal motor seizures recorded. Thereby, the model was first parameter-optimized in a leave-one-participant-out (LOPO) cross-validation on those training participants. Thus, each of the participants in this optimization set was omitted from the model training process once and used as a validation data set. The mean performance scores over the cross-validation runs were then used to determine the optimal parameter combination. In a second evaluation step, the optimized model, now trained with all the peri-ictal seizure data from the training subjects, was then applied to all the data from the test set participants. During the training of this model, the data were again first standardized using the z-score method, and those normalization parameters were then applied to the incoming test data. Overall this resulted in a model trained and optimized on data from one set of participants, which was then tested on data from another separate set of participants.

Classification Model

The GBT [95, 309] methodology was chosen as the model used to detect ictal states, as it is relatively straightforward in its application and was already validated on the same cohort, albeit on data from patients with convulsive seizures [104] (Section 4.1). Due to this methodology's requirement for parameter tuning to achieve good performance, hyperparameter optimization was conducted in both the intra- and inter-subject evaluations, as described above.

For the optimization of the intra-subject model, an optimal parameter combination was found for each of the three included participants by performing a LOSO cross-validation. Thereby, the model was trained on the data of all but one seizure and the respective peri-ictal data, and tested on all remaining data including the left-out seizure for that participant, minus the training data. This was repeated for all seizures, and the performance scores were averaged. This procedure was then repeated for each combination of parameters. For the inter-subject evaluation, a similar procedure was implemented for the three selected participants, but in a LOPO manner. The model with the best parameter combination was then tested on out-of-sample data from previously unseen participants.

Four different model parameters were optimized: the learning rate, the maximum number of weak learners, the maximum tree depth per weak learner, and the misclassification cost for false positives (FPs). Conversely, the misclassification cost for false negatives (FNs) was not tuned and kept unweighted, and only one type of boosting was used, namely adaptive boosting

for binary classification ("AdaBoost") [310]. In total, the number of different parameter combinations over which the grid search optimization was performed added up to 600. The best parameter combination was chosen as the one with the highest sensitivity and lowest number of FPs, in that order. In the case of a tie, the parameter combination with a higher learning rate or lower number of trees was chosen as the best one, as it would be computationally more efficient.

To gauge the influence of the various features on the creation of the model, the feature importance of each of the optimized models was analyzed. Therefore, the importance scores for each of the models resulting from the single cross-validation runs was averaged in the intra-subject LOSO evaluation, resulting in one set of scores per participant included there. Moreover, for the inter-subject LOPO evaluation, only the importance scores of the optimal model, trained on all seizures from those intra-subject evaluation participants, were noted. The feature importance was based on decision tree node impurity, using the Gini diversity index and calculated such that the smallest possible value was 0 [311, 350]. Thereby, the reported importance scores for each of the features are the averages over all the trained trees in the boosting ensemble for the GBT model.

Performance Measurement

The main indicators of performance used in this evaluation are the mean sensitivity, false alarm rate (FAR) per $24\,\mathrm{h}$ (false alarm rate per 24 h (FAR24)) and positive predictive value (PPV). These scores were calculated from the number of overlaps of seizure events in the ground truth and predicted labels. The label data, analogous to the feature data, were stored at $2\,\mathrm{s}$ intervals, and seizure events here were defined as consecutive intervals of labels classified as a seizure of at least $6\,\mathrm{s}$ and at most $10\,\mathrm{min}$. The output of the classification model was furthermore smoothed before this scoring computation, by filling out gaps between seizure labels of at most $30\,\mathrm{s}$, and removing any orphan seizure labels. After this processing of the model output, it was compared to the ground truth and any overlaps of seizure events were counted as true positives. Seizure events in the ground truth but not in the detector output were counted as FNs, and vice versa, seizure events in the output that were not present in the ground truth were counted as FPs. Note that the comparisons described above are given a $2\,\mathrm{min}$ margin before and after seizure events in the ground truth, wherein overlaps with detected events still count as true positives. This was performed to account for some of the uncertainty related to seizure manifestations, as well as the in-hospital setting providing a certain degree of nurse intervention after a seizure. True negatives were not counted in this evaluation, as they do not give any more worthwhile information for performance measurement. The false alarm rate per night (FARn) calculates the FAR during a standard $8\,\mathrm{h}$ night between 23:00 and 07:00. Thereby, the number of false alarms produced by the model occurring during that time period were counted, and divided by the number of hours that were recorded during these nights, taking into account any data loss that may have occurred during these hours. The FARn was therefore calculated as:

$$\mathrm{FARn} = \text{number of FPs during night} \cdot \frac{\text{hours per night}}{\text{nightly hours recorded}} \tag{5.1}$$

All data analysis, feature extraction, and performance evaluation was implemented using MATLAB R2021b (MathWorks, Natick, MA, USA).

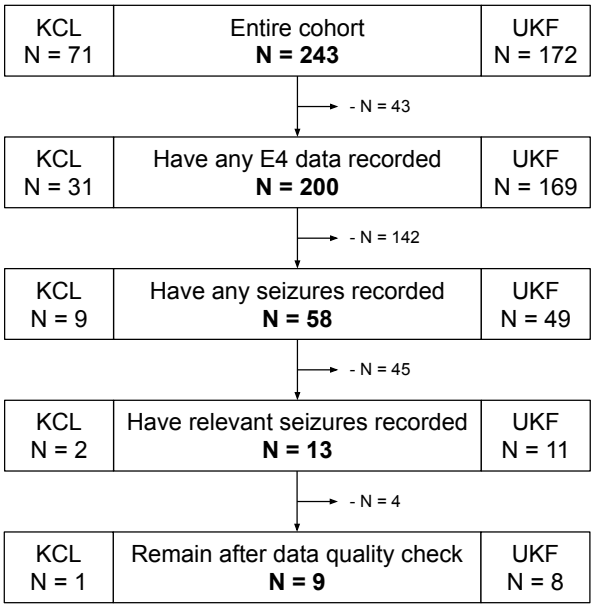


Figure 5.6: Data set flowchart of the participant selection process. KCL: King's College Hospital London; UKF: University Medical Center Freiburg; E4: Empatica E4 wrist-worn wearable device.

Data Set Selection

The complete data set of wearable data from 243 patients with epilepsy was filtered to include only data relevant to the premise of this study. Figure 5.6 visualizes the data set selection process, and Table 5.4 lists clinical and demographic information of the finally selected participants. First, only data from those participants who had at least one focal seizure recorded that involved tonic or clonic motor manifestations were included. Thereby, these manifestations could co-occur with other seizure manifestations; however, focal to bilateral tonic-clonic seizures were excluded. Moreover, the data set was not filtered further by overlap of symptom location versus device location. Therefore, the data set can, for example, include instances of motor seizures that manifest primarily on the right-hand side, but where the wearable device was attached to the left wrist. Excluding these seizure instances would substantially reduce the number of seizures and included participants for the analysis presented here, especially for the inter-subject evaluation, to the point of impracticality.

During the study recordings, the Empatica E4 device was used in a Bluetooth streaming mode [103], which unfortunately led to a considerable loss of data due to regular problems with connectivity of the wearable device to a base device that stores the data. Thereby, more than $50\,\%$ of the potential data to be recorded, and correspondingly as many potential seizures, were lost, leading to a substantially reduced number of relevant focal motor seizures recorded for this study. Furthermore, the data set was filtered for total length of recording per participant, where only those recordings with at least $24\,\mathrm{h}$ of data were included, and for length of seizures, where only those seizures with a duration between $10\,\mathrm{s}$ and $10\,\mathrm{min}$ are included. Limiting the duration of seizures excludes very short seizures of just a few seconds, such as myoclonic seizures, and very long seizures, such as status epilepticus. This is performed to exclude outliers and to have defined limits of duration within which to detect potential seizure events.

Table 5.4: Demographic and clinical information for the nine selected participants.

ID	Gender	Age	Total Recording Duration	Number of Seizures Recorded	Epilepsy Origin	Epilepsy Type
UKF1	m	55	84.4 h	6	Structural	TLE
UKF2	m	9	45.9 h	3	Structural	xTLE
UKF3	f	27	92.0 h	2	Structural	TLE
UKF4	f	69	120.8 h	1	Structural	TLE
UKF5	m	50	127.6 h	2	Structural	FLE
UKF6	f	34	35.8 h	1	Unknown	FLE
UKF7	m	48	105.1 h	1	Structural	TLE
UKF8	f	46	87.2 h	1	Structural	TLE
KCL1	m	65	50.2 h	3	Structural	TLE

TLE: focal temporal lobe epilepsy; FLE: focal frontal lobe epilepsy; xTLE: focal extratemporal lobe epilepsy.

Lastly, the data quality during all remaining seizures was visually checked, and specifically those with bad EDA signal quality were excluded. A poor EDA signal can either be a flat zero-line, indicating loss of contact of the electrodes with the skin, or multiple periods of high rates of amplitude change, indicating a loosely fitting device. The BVP raw signal was specifically not filtered for signal quality, firstly because it would filter out nearly every remaining seizure due to its high susceptibility to motion artifacts, and secondly because the feature set for this evaluation indeed includes a data quality index as a feature itself.

5.2.3 Results

Data Set and Examples

The resulting data set used for this evaluation thus included 20 relevant seizures from a total of nine study participants. 44 % (4 of 9) of the participants were female, and participants had a mean age of 45 years (range 9 to 69 years) at study enrollment. Three of these participants had more than two seizures recorded for a total of twelve seizures (Table 5.5), and the data from these were thus used for the intra-subject evaluation, as well as the training set for the LOPO cross-validation in the inter-subject evaluation. The remaining six participants had either one or two seizures recorded, for a total of eight seizures included in the inter-subject evaluation test set. The mean recording length per participant in this data set was 83.2 h (range 35.8 h to 127.6 h).

To give a better overview of the three participants with multiple seizures recorded and used in the intra-subject evaluation, Figure 5.7 presents one example seizure for each of these participants. Participant UKF1 had six seizures recorded with the wearable device, all of them focal onset motor seizures with tonic and clonic manifestations, ictal tachycardia, and impaired awareness. Furthermore, all of these seizures occurred while the patient was sleeping in his hospital bed, and all have a characteristic progression. Overall, these seizure symptoms are the most similar to focal to bilateral or generalized tonic-clonic seizures in the data set, yet noticeably lack the severity of the larger seizures, both regarding the vEEG and also the movements captured with the ACC signal.

Table 5.5: Clinical information on seizures recorded for the three participants used in the intra-subject evaluation.

Seizure ID	Seizure Duration	Motor Symptoms	Autonomic Symptoms	Awareness	Vigilance/Body Position
UKF1-1	82 s	Tonic, clonic	iTC	Impaired	Asleep/lying
UKF1-2	86 s	Tonic, clonic, myoclonic, automatisms (arms, legs)	iTC, UI	Impaired	Asleep/lying
UKF1-3	55 s	Tonic, clonic, myoclonic	iTC	Impaired	Asleep/lying
UKF1-4	73 s	Tonic, clonic, myoclonic, automatisms (legs)	iTC	Impaired	Asleep/lying
UKF1-5	43 s	Tonic, clonic	iTC	Impaired	Asleep/lying
UKF1-6	47 s	Tonic, clonic, myoclonic	iTC	Impaired	Asleep/lying
UKF2-1	23 s	Tonic	iTC	Aware	Awake/sitting
UKF2-2 *	39 s	Tonic	iTC, flushing	Impaired	Awake/lying
UKF2-3	107 s	Tonic	iTC, flushing	Impaired	Awake/sitting
KCL1-1	128 s	Tonic, clonic, automatisms (arms, face)	iTC	Impaired	Awake/sitting
KCL1-2	22 s	Tonic, automatisms (face)	iTC	Impaired	Asleep/lying
KCL1-3	22 s	Tonic, automatisms (face)	-	Impaired	Asleep/lying

* Seizure was not recognized by the model during evaluation. iTC: ictal tachycardia; UI: urinary incontinence.

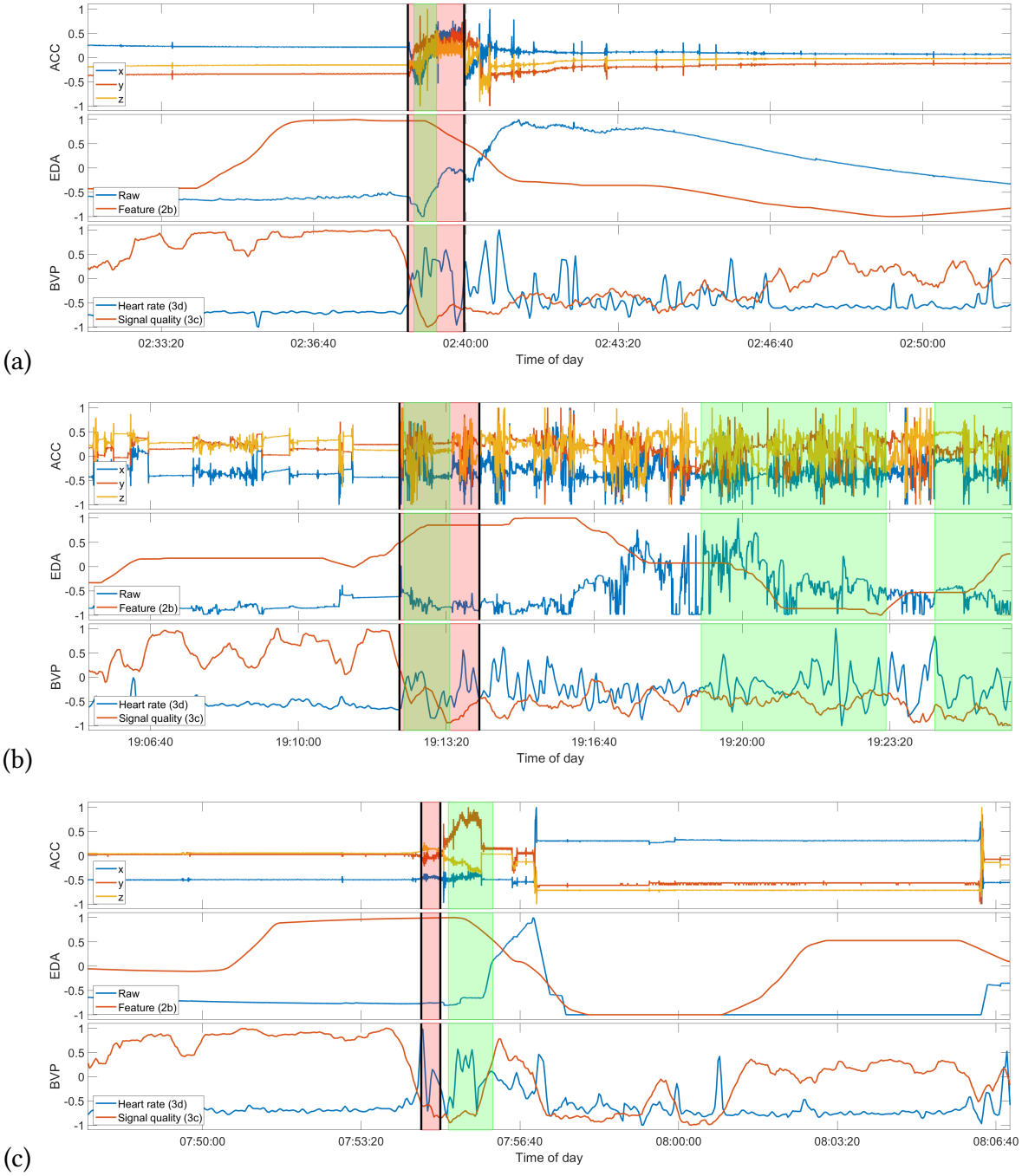


Figure 5.7: Selection of examples of true positive detections for each of the three participants in the intra-subject evaluation. Seizures shown are: **(a)** UKF1-4; **(b)** UKF2-3; **(c)** and KCL1-3 (see Table 5.5). Due to the grace period of 2 min around a seizure event, the detection for KCL1-3 counts as a true positive. Each plot of a seizure shows the raw ACC signal (top), the raw EDA signal and feature 2b (middle), and the estimated heart rate and signal quality index of the BVP signal (bottom). The regions highlighted in red mark the ground truth as labeled by experts, those highlighted in green mark the seizure intervals as predicted by the respective model. The seizure onset and offset are additionally marked by the black vertical bars. All signals shown are normalized between -1 and 1 only for these plots.

UKF2, on the other hand, had three focal onset motor seizures recorded with only tonic manifestations, ictal tachycardia, and miscellaneous awareness during the seizure. Notably though, all seizures occurred while the patient was awake. Another important distinguishing factor for this participant is that he was only nine years old at the time of enrollment, and as such the only pediatric patient in the relevant data sets regarded here. Epilepsy in pediatric patients generally manifests in different ways than for adults [23].

The sole data set from the KCL site, KCL1, had three seizures recorded that match the criteria for the seizure type. The motor manifestations for them were more heterogeneous than for the other two participants. All had tonic components, but there were also some oral automatisms, and one seizure also had clonic components. Furthermore, one seizure did not prompt ictal tachycardia, and there was a high variance between the seizure durations, with one being over $2\,\mathrm{min}$ and occurring while awake, and the other two only $22\,\mathrm{s}$, occurring from sleep.

Aside from the movements during the seizures, Figure 5.7 also gives a good overview of the typical EDA and BVP responses in the data, which can be observed most clearly in the first presented seizure UKF1-4. The EDA signal shows a clear response to the seizure, and the corresponding feature, the difference in the maximum of the skin conductance level, accordingly is at its highest during the seizure. The heart rate estimated from the BVP sensor signal also clearly demonstrates some response after the seizure onset for all three examples; however, at the same time, the signal quality also drops substantially, and as such the estimated heart rate should not be regarded as representative for these periods.

Intra-Subject Evaluation

Data from three participants were selected for the intra-subject evaluation. One patient was selected from the London cohort with three seizures recorded (KCL1), and two from the Freiburg cohort with six (UKF1) and three (UKF2) seizures recorded, respectively. Out of these twelve seizures, only one seizure, of participant UKF2, could not be identified in the individual optimized LOSO evaluation. All other seizures were consistently detected in the complete participant data by the optimized models, when trained on the other seizures of the respective participant. In terms of FAR however, the methodology exhibited vastly different performances over the three participants. The cross-validation runs for participant UKF1 showed the lowest number of false positives at just three on average, ranging from 1 to 5 depending on which of the seizures was left out for testing. Overall, this resulted in a low FAR of less than one per $24\,\mathrm{h}$ ($0.85/24\,\mathrm{h}$). For the other two cases, the FAR was considerably higher, at almost two per hour for UKF2 ($41.5/24\,\mathrm{h}$) and somewhat less than one per hour for KCL1 ($17.7/24\,\mathrm{h}$), on average. An overview of the per-participant results of this evaluation can be found in Table 5.6 under “Intra-Subject Evaluation”.

Inter-Subject Evaluation

To assess the performance of seizure detection across multiple patients, the GBT model was first trained using the peri-ictal seizure data of the 12 seizures from the three participants mentioned above. The model was thereby parameter-optimized in a LOPO manner, as explained in Section 5.2.2. In this cross-validation, the model with the best-performing parameter combination was able to recognize a total of eight of the twelve seizures (overall sensitivity $67\,\%$, mean $72\,\%$, and range $50\,\%$ to $100\,\%$) in the validation set, with a mean FAR of approximately one

Table 5.6: Evaluation results for the intra-subject leave-one-seizure-out evaluation, and the inter-subject leave-one-participant-out evaluation, respectively. Means and ranges are always across the single folds of the validations, that is, across the held-back data for the first part and across the held-back data, and test set participants, in the second part.

Patient ID	Sensitivity	Mean FP [Range]	Mean FAR24 [Range]	Mean PPV [Range]	Mean FAR per Night [Range]	Recording Duration	Device on Same Hand as Seizure
Intra-Subject Evaluation							
UKF1	100 % (6/6)	3 [1–5]	0.85 [0.28–1.42]	28.3 % [16.7–50 %]	0	84.4 h	100 % (6/6)
UKF2	67 % (2/3)	79 [18–126]	41.52 [9.42–65.94]	0.6 % [0–1.1 %]	6.3 [1–11.5]	45.9 h	100 % (3/3)
KCL1	100 % (3/3)	37 [0–58]	17.69 [0–27.72]	34.5 % [1.7–100 %]	1.9 [0–3.0]	50.2 h	0 % (0/3)
Inter-Subject Evaluation							
LOPO UKF1	50 % (3/6)	28	7.96	9.7 %	3.1	84.4 h	100 % (6/6)
LOPO UKF2	100 % (3/3)	124	64.9	2.4 %	9.5	45.9 h	100 % (3/3)
LOPO KCL1	67 % (2/3)	1	0.48	67 %	0	50.2 h	0 % (0/3)
LOPO test (N = 6)	75 % (6/8)	55 [16–87]	13.4 [4.4–22.7]	2.1 % [0–5.9 %]	2.0 [0.7–3.2]	568.6 h	38 % (3/8)

per hour, averaged over the three participants (mean $24.4/24\,\mathrm{h}$; range $0.5/24\,\mathrm{h}$ to $64.9/24\,\mathrm{h}$). The resulting model was then applied to the complete data sets of six other participants, including a total of eight epileptic focal motor seizures. In this out-of-sample test set, the model was overall able to detect six of the eight seizures (overall sensitivity $75\,\%$, mean $75\,\%$, and range $0\,\%$ to $100\,\%$) with a mean FAR of $13.4/24\,\mathrm{h}$ (range $4.4/24\,\mathrm{h}$ to $22.7/24\,\mathrm{h}$). Table 5.6 shows a summary of these across-participant results under "Inter-Subject Evaluation".

Feature Importance

The feature importance for each of the three optimal models trained on data from the three per-subject evaluation participants was calculated as outlined in Section 5.2.2. Figure 5.8 shows these importance scores per participant and feature, and the mean scores of each feature group by modality. These feature scores are unit-less and can be interpreted qualitatively to determine whether some specific feature or general modality is contributing more than the others. Here, the EDA features were more influential than the others in the two participants UKF1 and KCL1. Conversely, the BVP features were unexpectedly more meaningful for the model of participant UKF2 than the other two modalities. The same kind of feature importance scores for the inter-subject model trained on all three of these participants for the inter-subject evaluation can be found in Figure 5.8d. Here, the EDA features turned out to be more important than the others.

5.2.4 Discussion

Principal Findings

The main ambition of the evaluation presented here was to qualitatively assess the utility of multimodal biosignal data from wearables in creating worthwhile and robust seizure detection systems. Thereby, two principal avenues of potential study design were investigated: intra-subject and inter-subject schemes. Specifically, this evaluation focuses on focal motor seizures with tonic or clonic components, as opposed to bilateral tonic-clonic seizures. These focal seizures have a multitude of possible physical and psychological manifestations that can occur in sequence or in parallel, be repeated, or not occur at all, in a single seizure. Furthermore, while there may oftentimes be little change in the semiology of seizures for a single patient with epilepsy, they can be very heterogeneous across populations [27, 255]. These circumstances are also reflected in these results. Among the three participants with at least three seizures recorded, the individually optimized model could robustly recover the left-out seizures in the leave-one-seizure-out cross-validation for two participants.

In one other participant, however, out of three seizures, one could not be restored by the model when trained on the other two (Table 5.5 and Figure 5.9). These three seizures had roughly the same semiology with tonic manifestations and ictal tachycardia. In the wearable data, however, one clear difference can be found between this undetected seizure and the other two; that is, it showed no discernible EDA response before, during, or after the seizure. Additionally, this participant UKF2 had an important demographic difference to all other included participants in this data set, in that they were the only pediatric patient at nine years old. Age has been linked, for example, to significant changes in seizure semiologies [23]. These circumstances likely led to this specific seizure falling out of the scope of this methodology.

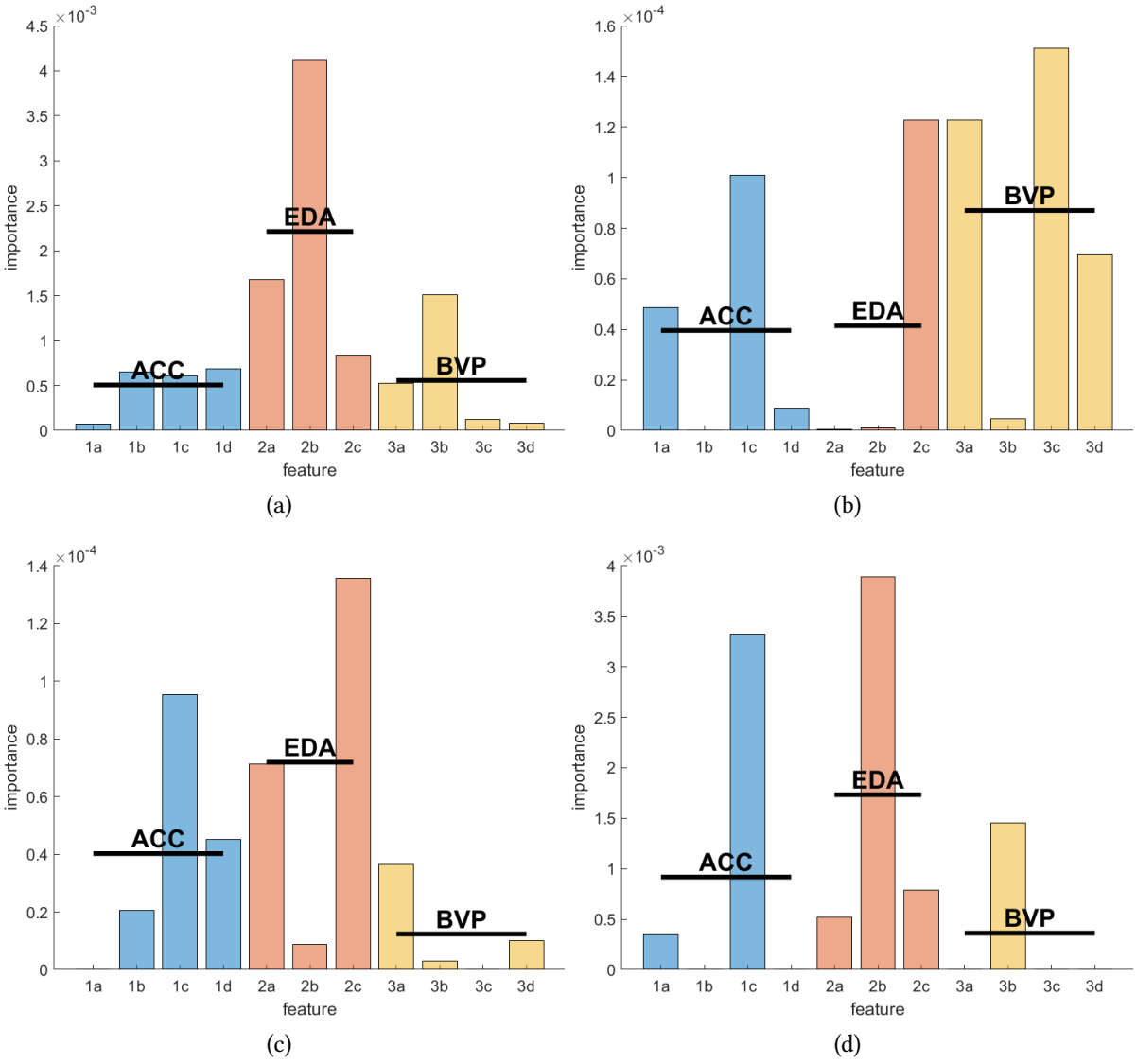


Figure 5.8: Feature importance scores per intra-subject evaluation for the seizure detection models of the three selected participants: **(a)** UKF1; **(b)** UKF2; **(c)** KCL1 (see Table 5.4); **(d)** Feature importance scores of the model resulting from training the GBT model on the seizure data of all three inter-subject training participants. Blue, red, and yellow bars show the importance scores for the features grouped by biosignal modality ACC, EDA, and BVP, respectively. Horizontal lines mark the mean scores of the groups. The ordinate is unit-less; the scores can be interpreted qualitatively. The feature labels correspond to the listing of features in Section 5.2.2.

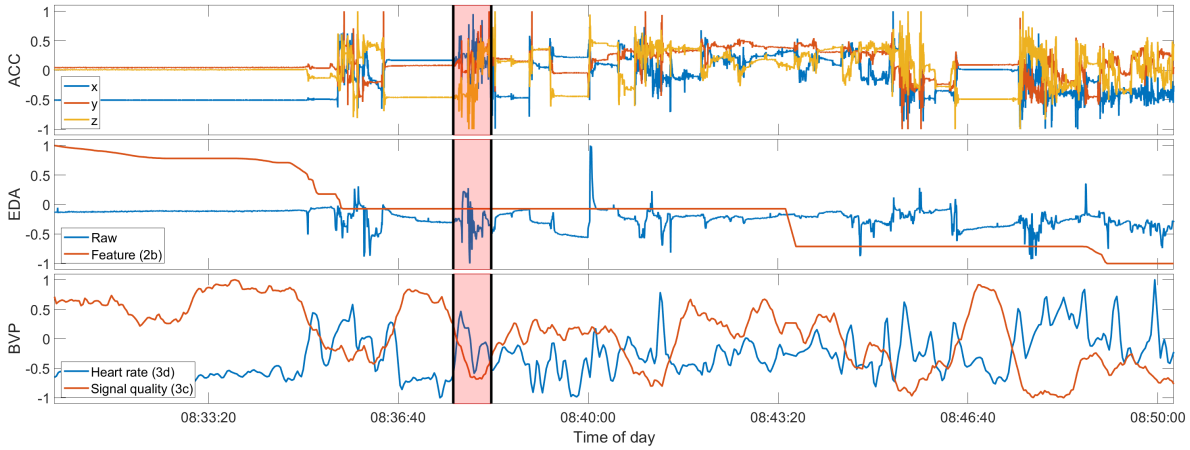


Figure 5.9: Seizure UKF2-2, a false negative. Compare also to Figure 5.7. Data shown from top to bottom: raw ACC, raw EDA and feature 2b, heart rate and BVP signal quality index. The red overlay is the seizure ground truth. The seizure onset and offset are additionally marked by the black vertical bars. All signals shown are normalized between -1 and 1 only for these plots.

These results suggest that a methodology such as the one presented here, optimized on individual participants, can robustly detect seizures for some patients with epilepsy, but it may fail, especially when the seizures have differing semiologies that are not represented in the training data for the model. Furthermore, when looking at the FAR24 and PPV, the heterogeneity of focal seizure detection is especially highlighted. The FAR24 performance of the seizure detection, ranging from less than one FP per day to almost two FP per hour, is an important factor when it comes to actually applying the methodology to a real-world setting. Similarly, this false alarm rate also carries over to the nighttime, with multiple false positives per night for some participants. Thus, a high sensitivity in detecting seizures is in vain if an automated seizure diary is filled with dozens of false seizure events per day. Yet, further data recordings and model optimization may produce robust seizure detection systems for individual patients.

With respect to the inter-subject evaluation across multiple study participants, the results for the methodology applied here further demonstrate the heterogeneity of the focal motor seizures in this data set, and clearly demonstrate the resulting difficulties. Inter-subject models applied in a leave-one-participant-out manner to data of the three selected participants from the intra-subject evaluation performed worse than if trained in an individualized manner, either in terms of sensitivity or FAR. Likewise, testing the model on out-of-sample data of six other participants resulted in a tolerable sensitivity but a high FAR and low PPV. A model such as this would be ineffective in real-world settings, be it as an automated seizure diary or an alarm system, and patients' and caregivers' needs, in particular, would not be fulfilled [72, 130, 131].

Overall, the results suggest that not only are focal onset motor seizures varied in their clinical manifestations, they are also sensitive to changes in common wearable biosignal modalities, when investigated across patients. However, in some individual patients, seizure semiologies are similar enough across seizures to enable robust seizure detection models for less-severe focal onset motor seizures, if the models are optimized in a personalized manner.

Related Work

To compare the results presented here to some of the current and past state-of-the-art seizure detection studies, a list of twelve related works featuring focal motor seizures in some form in their set of analyzed seizure types was compiled (Table 5.7). The main source of this list was the extensive literature review by Beniczky et al. [109], filtered for relevant seizure types and relevant biosignal modalities that are most closely related to the modalities ACC, EDA, and PPG. Furthermore, to add some variety, the list includes two recent studies employing wearable electromyography- and EEG-based focal seizure detection. All three of the performance measures included here are compared, namely sensitivity, FAR24, and PPV. However, as most of these works include focal motor seizures as part of a larger group of seizure types, almost always dominated by generalized and focal to bilateral tonic-clonic seizures, the list only includes three studies where this was possible for all three measures [74, 227, 341]. All three of these studies classify the focal onset seizures in their respective data sets as complex partial seizures (CPSs), which is an older type of epileptic seizure classification meaning an interval with ictal impaired awareness without giving information about motor manifestations during the seizures. Therefore, it is unclear if, with respect to movements during the seizures, the seizure types investigated in these works are comparable to those in the study presented here. Moreover, all three studies evaluated their seizure detection in an inter-subject manner across a population of patients with epilepsy.

Summarizing the relevant results from these three works, Cogan et al. [227] used a combination of two different wearable devices to record the biosignal modalities EDA, electrocardiography (ECG), and blood oxygen saturation, and reported an algorithm sensitivity of $50\,\%$ with a false alarm rate of $0.28/24\,\mathrm{h}$, in CPS only. Notably, during their analysis they also looked into personalization of their algorithm, and concluded that a minimum of 6 to 8 seizures per patient would be required to sufficiently train and optimize their algorithm parameters, based on a worst-case scenario. Kusmakar et al. [341] employed ACC sensors in a leave-one-participant-out inter-subject evaluation and reported a sensitivity of $67\,\%$ and a FAR24 of 4, regarding the participants with CPS. They duly concluded that these seizures are much more similar to inter-ictal data than they are to ictal generalized tonic-clonic seizure data, and as their data set included only a very small number of CPS, a good performance on this seizure type was not to be expected. Lastly, Vandecasteele et al. [74] compared two wearable devices recording ECG and PPG, respectively, and reported an overall sensitivity of $32\,\%$ with a FAR24 of 43.2 for the PPG-based wearable device, which was the same as the one used in this study. They concluded that the PPG-based detection was heavily impeded by motion artifacts even for these possibly non-generalized seizures.

Comparing these performances, and further results from other related work which did not report their outcomes per seizure type or per participant (Table 5.7), to this study's results (Table 5.6), it becomes clear that a sensitivity of $75\,\%$ in an independent test set of focal motor seizures is in fact among the top performing methodologies regarding this seizure type only. Furthermore, works that reported lower numbers of false positive rates also consistently had lower sensitivities, and some studies that reported similar numbers nevertheless still had lower sensitivities. Overall, a comparison of this study's results to any of these works should be taken with caution, as this study specifically analyzes focal motor seizures with multi-modal non-EEG wearable data.

Table 5.7: Related work compiled from Beniczky et al. [109], and this study as comparison. Only those works are included that involve seizure types relevant to this study, that is, any of focal motor seizures, SPS, CPS, or other non-generalized seizures.

Study	Modalities	Seizure Types	# Pat. w/Seizures	# Seizures	Sensitivity	FAR24	PPV
this study	*ACC, EDA, PPG*	*FS t or c*	*9*	*20*	67–100 %/75 %	*0.85–41.5/13.4*	0.6–34.5 %/2.1 %
[226] +	ACC, PPG	FS hyper/ o convulsive	(28 total)	5/14	73 %/84 %	Not reported per seizure type	
[280] +	ECG	FS/SPS/CPS/o	(31 total)	8/26/31/5	Not reported per seizure type		
[227] +,*,§	EDA, ECG, SpO2	CPS	8	23	16.7 %/50 %	0.7/0.28	6.25 %/50 %
[340] +	ECG	SPS/CPS	(16 total)	37/38	19 %/71 %	Not reported per seizure type	
[344]	ACC	t/t-c	15	22	67 %/100 %	Not reported per seizure type	
[341] +,*	ACC	CPS	3	5	67 %	4.19	22.5
[224] +	ACC, EDA	FS t-c	2	6	50 %	Not reported per seizure type	
[315] +	ACC	myo,t/ FS hyper/ FS min mot	(41 total)	140	6 %/24 %/2 %	Not reported per seizure type	
[343] +	ACC, ECG	FS hyper/ myo,t-cluster	5/5	18/9	Not reported per seizure type		
[74]	ECG/PPG	CPS	11	47	70 %/32 %	50.6/43.2	2.15 %/1.12 %
[231]	EMG	GTCS/t/c/ o-motor	20	18/9/3/17	83 %/56 %/ 33 %/76 %	-	83 %/50 % (t+c)/76 %
[230]	EEG, ECG, ACC	FS t/ FS non-motor	3	47/9 + 9	84 %/100 %	8/13 + 5	-

+ The study also contained other seizure types, most notably generalized seizures, however the presented data only relate to those seizure types specifically mentioned.
* Performance scores only include CPS, calculated by authors from original reported numbers. § Performance scores represent a non-optimized detection, and a refined analysis, respectively.
t: tonic; c: clonic; hyper: hypermotor seizures; myo: myoclonic seizures; FS min mot: FS with minimal motor component; o: other.

Modality Importance

Overall, the distributions of feature importance scores, as a proxy for the importance of biosignal modalities, seem to be heterogeneous, and no clear winner could be found among the three examples. The features calculated from the EDA signal, however, seem to be the most informative with respect to epileptic seizure phases, within this data set of focal onset motor seizures. This is concurrent with some prior research on generalized seizures as well [223, 293, 351, 352], and suggests that electrodermal activity could be an important clinical marker of epileptic seizures beyond highly convulsive episodes, which usually induce heavy sweating. However, it is not a universally applicable biomarker, as these results also suggest, with at least one participant and several seizures exhibiting no substantial (post-)ictal EDA response, as was also concluded in further literature on the topic [42, 43, 105].

In Glasstetter et al. [348], the authors explored the utility of wearable PPG signals for the detection of focal onset seizures with ictal tachycardia, and concluded that in some patients the tachycardia thresholds can be found regardless of seizure-related movements. It seems to be the case that in some seizures the ictal tachycardia presents itself some few seconds before the electrographic seizure onset, and could therefore be used as indicators for seizures, albeit not effectively in a monomodal system. The authors, however, also noted that arbitrary non-seizure-related movements may hinder this detection from PPG signals due to motion artifacts, and generally pre-ictal tachycardia seems to be an isolated phenomenon not generalizable over patient cohorts. The results presented here coincide with these conclusions, demonstrating that BVP features are important only for one of three patients with at least three seizures recorded.

Limitations

The main limitation of the study and analysis presented here is the limited size of the data set, with just 20 seizures recorded from nine patients with epilepsy. Unfortunately, publicly available data sets focusing on classical wrist-worn wearable data which combine movement and autonomous nervous system biosignals are scarce as it is, and practically nonexistent in the field of epilepsy research. To be considered a useful supplement to the data collected here, a public data set would need to provide at least ACC, EDA, and BVP data recorded from a wrist-worn wearable in a cohort of patients with epilepsy, and at least EEG seizure onset and offset would need to be labeled by experts. Such a data set does not currently exist publicly. The relatively small size of the data set used here is also the main reason the methodology used "classical" supervised machine learning with a set of specifically tailored features as opposed to some deep learning methodology. Deep learning may, however, be an avenue to pursue in a later study, with a more extensive data set.

As the seizure types regarded here were carefully chosen, a large percentage of the overall recorded seizures were not eligible for inclusion. Furthermore, during their usually week-long stay at the epilepsy monitoring units, patients rarely had more than a few seizures in the first place, as seizure provocation was often conducted only for a limited time until enough seizures had been observed to serve some clinical purpose. This lead to a generally low number of study participants with more than two seizures recorded, a necessary minimum requirement for any kind of analysis of personalized, intra-subject evaluations. Future research may include more seizures and seizure types from the already recorded data set, but new studies need to be conceptualized to aim for higher numbers of seizures recorded per participant, not just over

the whole cohort, in order to push focal seizure detection research to produce better, more realistically applicable results [130].

Another limitation of this study, as well as many other related works, was the confinement of participants to a hospital room throughout the data recording procedures. The data collected and analyzed here cannot be regarded as representative of real-world situations, and new ambulatory studies with a prospective goal of recording focal onset seizures with wearables from patients in their daily living environment are needed. However, even with generalized tonic-clonic seizures as a target, these phase 3 and 4 studies are still rare, and potential methodologies and pitfalls need first be explored in these in-hospital studies before worthwhile out-of-hospital studies can be designed.

5.2.5 Conclusions

Seizure detection for focal onset seizures without generalization by means of wearable non-EEG devices is a so far little-researched problem. This study demonstrated that for those seizures in this category with tonic or clonic movements, that is, those closest in semiology to generalized tonic-clonic seizures, robustly detecting seizures from wearable data may be possible for individual patients with epilepsy, depending on their specific seizure manifestation. Overall, electrodermal activity signals seem to provide the most informative features for seizure detection, suggesting it to be an essential part of any future seizure detection research. Detection across patients with purely inter-subject models without personalization was, however, not possible to a worthwhile degree, at least with the methodology and data set presented here. Both a low sensitivity that misses a quarter or more of the seizures and high false alarm rates in the order of one per hour make current results of these inter-subject models ineffective for clinical applications. This study's results thus demonstrated that individualized models are needed to robustly detect focal onset seizures, and future research in this domain should include at least some degree of personalization in the modeling process.

5.3 Summary

This chapter explored and evaluated the automatic detection of focal motor seizures by way of multimodal biosignal data from wearables. Focal seizures (FSs) are the most predominant seizure type in the general population and can have a multitude of different symptoms, ranging from movement of the limbs over tachycardia to sensory feelings. However, FSs are also typically less convulsive than their generalized-onset siblings, with some seizure types not including movement semiology at all. Thus, they are considered to be less unambiguously detectable with signals from wearable sensors like accelerometry (ACC), electrodermal activity (EDA), or blood volume pulse (BVP). Still, some types of FSs exhibit characteristic patterns both in movement and response of the autonomous nervous system, so a robust detection is not necessarily impossible.

In an exploratory investigation of focal motor seizures, the first part of this chapter highlighted their extreme heterogeneity in terms of semiology relevant to seizure detection with data from wearables. Focal motor seizures can manifest in characteristic tonic or clonic movements of the limbs as much as in random automatisms of the mouth or in exceedingly brief jerks of the whole body, oftentimes indistinguishable in the ACC signal from other non-epileptic movements. The study illustrates a core problem of seizure detection, the high variance in focal seizure manifestations, by evaluating a detection model on FSs from three different patients. While it performs acceptable on seizures with characteristic movement patterns of the limbs, it fails on other movements and those not manifesting in the limbs. This exploratory study represents a first glimpse into the conclusion of the subsequent part of the chapter, and parts of the eventual conclusion of this thesis.

That conclusion results from the final data analysis study in the context of this thesis, namely, that robustly detecting focal onset motor seizures with tonic or clonic movements from wearable data may be possible for individuals, depending on specific seizure manifestations. Based on the experiences of the previous studies, this study includes 20 focal motor seizures from 9 patients and evaluates both a patient-individual and a patient-agnostic approach. The detection methodology was similar to the one used in the convulsive seizure detection in Section 4.1, but with a new set of ACC, EDA, and BVP features. It achieved sensitivities of $67\,\%$ to $100\,\%$ and false alarm rates of down to $0.85/24\,\mathrm{h}$ for individualized detection models in three patients. Across patients, it reaches comparable sensitivity, however, false alarm rates are much higher, to the point of being impracticable for useful seizure detection systems.

Viewing both parts of this chapter in the context of automatic seizure alerts, it becomes clear that this problem is still to be solved, as multiple false alerts per day up to and beyond one per hour are unrealistic to be accepted by users. Furthermore, other related research reports similar performances across different combinations of focal seizure types, detection strategies, and biosignal modalities. Viewing this research in the context of automated seizure diaries, however, may be more interesting for the time being. For that purpose exceptional performance may not be the deciding factor, since the discovery of qualitative changes in seizure frequencies can already be an important outcome for clinicians to be able to help their patients. And as long as the detection model can reliably detect even just a portion of the occurring seizures, a robust qualitative assessment may be within reach.

CHAPTER 6

Conclusion

Wearable devices are a relatively new and promising tool to enable the remote, unobtrusive, and automated detection of epileptic seizures with multimodal non-electroencephalography (EEG) biosignal data. Affecting over 70 million people worldwide, with up to 100 new cases per 100,000 people per year [14], epilepsy is one of the most prevalent chronic neurological disorders. The severity of epileptic seizures can vary tremendously, ranging from major convulsive seizures accompanied by unconsciousness and a heightened risk of sudden unexpected death in epilepsy (SUDEP), to minor second-long jerks or a short sensory aura. Yet, the current gold standard of diagnosing, monitoring, and treating epilepsy are short visits to an epilepsy monitoring unit (EMU) at a hospital and undergoing video-electroencephalography recordings, with doctors provoking seizures by various different means in the hopes of capturing enough information to advise treatment. At home, between visits to the EMU, patients sometimes keep handwritten diaries of their seizures, noting down the approximate time and symptoms. However, due to the nature of epileptic seizures often involving an impairment of awareness or loss of consciousness, these manual diaries typically severely underestimate the amount and severity of seizures. This thesis investigated the utility of wearable devices for the ultra-long-term monitoring of patients with epilepsy, potentially improving their quality of life by enabling an automated way of keeping seizure diaries, and thus facilitating new and better treatments and lessening the burden of uncontrollable seizure events.

Wearable devices like smartwatches or activity trackers will often record a variety of biosignals that are relevant to the health and safety of their users. Among the potential modalities recorded are movement by accelerometry (ACC), changes in the autonomic nervous system by electrodermal activity (EDA), and the heart rate by photoplethysmography measurements. It is conceivable that epileptic seizures, especially major convulsive seizures and those with other motor manifestations of the limbs, can be detected automatically via changes in these signals. Indeed, there has been more and more research into this capability in the past ten years, making great strides towards autonomous, ambulatory, and abiding seizure detection with wearable devices. Nevertheless, this is still a relatively young field of study, and the amount of research that has gone into designing and evaluating different methodologies of wearable seizure detection is still in need of augmentation. This thesis compiled and framed several contributions of the author concerning the detection of epileptic motor seizures with multimodal non-EEG data from wearables.

Overall, the research topics presented in this thesis can be summarized to several key contributions, expressed along the investigations of Chapters 3 – 5:

Data Collection With Wearables: As a foundation for the subsequent data analysis studies, an extensive data set of biosignal data from wearables needed to be collected in a cohort of patients with epilepsy. An open data set of such data containing recorded epileptic seizures did not exist, so a total of 243 participants at two European epilepsy centers were recruited during their routine stay in an EMU. More than 300 seizures of various types were recorded with a wrist-worn research-grade wearable device (Empatica E4), including expert-labelled seizure onset and offset as well as other relevant metadata. Contributions include the dissemination of study design and procedure experiences, the implementation and setup of a local data collection architecture by way of an open-source mobile health platform[1], and the resolution of several data issues like timestamp synchronization.

Detection of Convulsive Seizures: A novel methodology employing features from ACC and EDA biosignals with an ensemble supervised machine learning model was developed for the automated detection of convulsive tonic-clonic seizures (TCSs). The technique could correctly classify $91\,\%$ of the seizure events in an out-of-sample test set of 11 seizures from 2 participants in different cohorts. Further, a false alarm rate of $0.19/24\,\mathrm{h}$ in 78 days of data places the methodology at the same or better level than other TCS detection methodologies employing multimodal non-EEG wearables. In a different analysis, a heuristic ACC-based detector of cessation of movement after convulsive seizures was developed. It was able to correctly classify $100\,\%$ of the 22 post-ictal events in 18 study participants, including two cases that showed post-ictal agitation, the opposite of the post-ictal immobility that is typically experienced after a TCS. Both works could be combined into a warning system for increased risk of SUDEP, but further optimization concerning the seizure detection latency is necessary. Still, by themselves the contributions show that monitoring convulsive epileptic seizures with multimodal wearable devices is feasible with high robustness.

Detection of Focal Seizures: While the detection of convulsive seizures leaned on the distinctive and extreme impact those seizures have on the biosignals recorded with wearables, the detection of focal seizures (FSs), even of those with movement manifestations, is more ambiguous. Two contributions explore the variety of motor manifestations in FSs and the impact of this circumstance on the feasibility of detection with wearable device data. FSs can be characteristic in individual patients, but are typically not comparable across patients, reflected in the result of patient-specific detection models performing overall better than those evaluated within entire cohorts. Related work on this topic is limited in number, and often needs to be dissected for results on specific FS types, as they are considered in combination with convulsive seizures more often than not. Yet, the contributions presented here show that there is substantial need for the automated detection of FSs from wearables, and current research shows ample promise that this is within reach.

[1] radar-base.org

In summary, this thesis presents research regarding the detection of epileptic seizures with multimodal non-EEG data from wearables. Specifically, those seizures with manifestations including movements of the limbs were investigated, and detection systems based on supervised ensemble machine learning using physiological biosignal data were found to be feasible.

The remaining sections of this final chapter will compile some additional lessons learned from the experiences made during the work on this thesis, and give an outlook on several aspects and issues to be tackled in the field of seizure detection with wearable devices.

6.1 Practical Lessons Learned

During the course of working on this thesis several worthwhile experiences were made with regard to epilepsy and wearable research. This section aims to impart some valuable lessons learned from those experiences that have not been specifically discussed elsewhere in the text. They may guide the design of new clinical studies and further research, as also discussed in Section 6.2. This meta-analysis does not communicate systematical, scientific research that has been published, but reflects the views and opinions of the author of this thesis.

6.1.1 Diversity of Epileptic Seizures

The extreme heterogeneity of epileptic seizures, especially those of focal onset without generalization, has been discussed in this thesis on various occasions. One aspect that has not been discussed further, however, is the clinical classification of epileptic seizures that is presented in Fisher et al. [27]. While this newest practical classification of seizures has made great strides in removing ambiguity in the naming of seizure types in a clinical context, there is still a lack of compatibility with respect to the classification from non-electroencephalography wearable data. Of course, that was not the purpose of this terminology (re-)assessment, but it brings up the issue of whether a supplementary taxonomy of seizures by their physiological semiology closer to what wearables may capture may be necessary. As it stands, the classification used universally by published studies both in general epilepsy research and in research concerning wearable systems is predominantly focused on the electrographical onset of a seizure and the peri-ictal awareness of the patient, and only secondarily on the motor manifestations. However, there is currently no evidence that suggests a direct relationship between the psychological state of consciousness of a patient and physiological data recorded from wearables like accelerometry, electrodermal activity, or blood volume pulse. Conversely, introducing a ubiquitous and commonly applied (sub-)classification based primarily on movement and responses of the autonomic nervous system may facilitate a more effective communication of wearable-based seizure detection methodologies and evaluations. The classification by Fisher et al. [27] already takes a step in this direction, but it is not yet universally implemented by current related work in this regard.

6.1.2 Selection of Meaningful Features

Again resulting from the high diversity of epileptic seizures, the performance of epileptic seizure detection by wearables seems to be highly dependent on the recorded biosignal data modalities and the choice of features computed from them. In the work included here, visual inspection of signal traces during seizures (see Figure 4.1), experience from related work, and

some common sense were all very helpful in selecting the feature sets used in the analyses. Furthermore, some research was conducted to investigate the value of certain biosignals and features in a deliberate manner. Glasstetter et al. [348], for example, assessed the utility of photoplethysmography (PPG) signals from wearables to detect ictal tachycardia, comparing it to the gold standard of electrocardiography-based heart rate estimation. They conclude that PPG-based estimation is feasible for seizures where tachycardia onsets before motor manifestations, but may be limited for other seizures due to motion artifacts impairing the signal. This was successively observed in the evaluation of focal motor seizure detection presented in this thesis as well (see Section 5.2). There is certainly room for improvement regarding the assessment of biosignal features, and more similar analyses are required beyond the current research. For example, studies applying metaheuristic feature selection and dimensionality reduction algorithms to wearable biosignal data recorded from epilepsy patients could facilitate better performances of seizure detection. Moreover, non-feature-based approaches like deep learning could be a viable path forward, but these typically require vastly more quantities of data than would be recorded in in-hospital settings.

6.1.3 Wearable Device Design

The current landscape of wearable devices for seizure detection is limited at best. By far the most prominent in related research is the Empatica E4 device, which was also used in the studies included in this thesis. Other devices exist and are in use [122], like the Byteflies Sensor Dots [353] or the Nightwatch bracelet [226], but they are often still in development, not certified as a medical device in the European Union, or limited to a few specific studies or contexts. Coincidentally, the Empatica E4 is also at the end of its lifetime and being phased out [354], to be replaced by a new wrist-worn wearable device from the same manufacturer ("*Empatica EmbracePlus*"), which however at the time of writing has not been featured in any peer-reviewed publications.

During the data recording studies included in this thesis, the Empatica E4 exhibited several issues related to device design that interfered with recordings. Primarily, the device had several problems with the continuous streaming of data via Bluetooth connection. While the maximum range of the wireless connection was just enough to include the whole hospital room, such that the patient was able to move roam around the room, there were frequent and unprompted disconnects even if the wearable was well within range of the companion device. Better, more robust wireless hardware is likely needed to mitigate these disconnects. Alternatively, the device could implement a data buffer that temporarily captures biosignals internally whenever it loses the connection, transferring the data only if it is reestablished. Moreover, the battery life of the device was an issue, especially since the batteries tended to degrade over time, with each new participant. However, this is a general problem in wearable devices and not specific to this one, and new technologies and approaches are forthcoming. Lastly, the wristband of the device was not flexible enough, both in its material and possible wrist sizes. Materials that are more adaptive the individual's anatomy, like fabrics or more pliant plastics, would facilitate better, cleaner biosignal recordings with sensors less prone to a loss of skin contact.

Generally speaking, wearable devices in this domain need to balance a trade-off between being general-purpose consumer-grade devices, and being specialized research-grade devices. The former are often designed with comfort and usability in mind, a necessity for ultra-long-

term recordings that are essential in epilepsy diagnosis and research, but they rarely give access to raw biosignal data. The latter give access to raw data, are flexible in terms of firmware and sensor parameters, and typically facilitate better overall data quality, but are impractical at best in the at-home context due to battery limitations, connection issues, and design concessions. All that is to say, that wearable device design, from the ground up, should keep both the patients and the researchers in mind, and not be exclusive to one group or mindset. Furthermore, open data sets of biosignal data from wearables during epileptic seizures are still a rarity, but necessary for the continued improvement of seizure detection systems. Device manufacturers could look to build and share such data sets as a result of their device validations, were they not focused solely on the design of proprietary, closed algorithms. In fact, entirely open-source ambulatory data collection and analysis with wearables was shown to be feasible, and could provide the foundation of future adaptive intervention systems [355].

6.2 Outlook

The data analysis studies included in this thesis each list some limitations that the respective evaluations were subject to. Naturally there was also some overlap between those studies' limitations. Some prominent aspects that were listed included the generally low numbers of study participants and seizures included in the analyses, the "lab" conditions of the data collection, that is, in-hospital recordings, or data quality considerations. This section discusses some of these aspects in the context of potential further research work.

6.2.1 Ambulatory Data Collection

All data recorded for the studies in this thesis were recorded from patients visiting an epilepsy monitoring unit (EMU) at a hospital, where they are usually constrained to a room with electroencephalography (EEG) electrodes attached to their scalp. Thus, they are heavily restricted in terms of the movements they can or will do, impacting the nature of the biosignal data recorded. For example, patients will rarely do physical exercise while at the EMU, even though they would regularly do so in their daily lives. Accordingly, seizure detection methodologies must be validated in real-world data at some point. Ground truth information, however, is difficult to record in ultra-long-term ambulatory studies. Current options include at-home video systems, which are often only viable during the night at the patient's bed and thus are problematic with respect to their privacy. Furthermore, there are wearable EEG systems like caps or single-channel behind-the-ear EEG, which are cumbersome, uncomfortable, stigmatizing, and inaccurate, and implantable EEG sensor systems which are invasive and have a high cost. Some few phase 4 ambulatory validation studies are ongoing [109, 223], but more studies with different devices, better ground truth, and larger patient cohorts need to be conducted.

6.2.2 Assessment of Data Quality

Seizures are typically short, often as short as a few seconds, and so seizure monitoring requires a high temporal resolution of the biosignal data as compared to many other clinical applications [342]. Furthermore, data accuracy and reliability depend on the quality of the recorded signals [173]. However, clinical epilepsy studies with wearables frequently underreport quantitative measures of the quality of the recorded raw data. In particular, tools and

standard metrics determining the data and signal quality of wearable signals are currently missing. Yet, knowledge of artifacts and tools to assess their impact are crucial for subsequent analysis and outcome reliability, and there is a need for extensive, automatic, standardized data quality checks. Some recent studies have investigated wearable data quality in the context of epilepsy [159, 174], but more research is required. Comprehensive stand-alone studies are necessary to properly validate signal quality metrics, but this is easier said than done. Most of the biosignals recorded with the wrist-worn wearable used in this thesis, for example, do not have an immediately obvious direct gold standard that could be used as a ground truth. For electrodermal activity (EDA), dry electrodes are already the best existing methodology to record skin conductance changes, and there is no accepted best practice for other parameters like the position of the electrodes on the body or the type of electrode and mechanical contact. Additionally, the blood volume pulse data could be validated only indirectly by estimating a heart rate (HR) and comparing it to an electrocardiography-based HR.

6.2.3 Detection of Non-Motor Seizures

In this thesis only seizures with some type of movement manifestation were considered. However, many focal seizures do not include epileptic movements, but are limited to symptoms of the autonomic nervous system, impairment of awareness, or sensory auras. Automatic detection of those events is likely to be more problematic. For one, the accelerometry biosignal will be meaningless and possibly even counterproductive, considering that random non-epileptic movements of the patient can still occur during those times. Furthermore, judging from the results of the focal motor seizure detection, individualized analysis is likely to be the only worthwhile pathway, at least with current wearable biosignal modalities. Nevertheless, the data set recorded for the studies presented here included some individual patients who had non-motor seizures with characteristic semiology like, for example, one with ictal goose bumps on the arms, which might be eligible for robust detection by EDA signals. Future studies or specific case reports on these examples may benefit the general understanding of some epileptic seizure types and advance non-motor seizure detection with wearables.

6.2.4 Other Machine Learning Approaches

In the seizure detection analysis contributed here, only supervised machine learning based on expert-labeled ground truth data was employed. This does not mean, however, that other approaches like unsupervised or reinforcement learning may not be equally, or even more so, relevant and effective. Many epileptic seizures are extreme events compared to the everyday life of patients, and so they can in a way be seen as outliers. Some research has already investigated this approach based on EEG signals [356–358], but it remains an open issue and studies based on wearable data are yet to be published.

Deep learning is another approach to machine learning that was not specifically included in this thesis. Artificial neural networks can adhere to any of the three main learning paradigms, supervised, reinforcement, or unsupervised. Supervised neural networks are typically trained with vast amounts of raw (non-feature) data, and thus would be most appropriate for ultra-long-term data sets from ambulatory settings. Data sets like this with non-EEG data from wearables are currently not available, but could be recorded with new devices suitable for daily at-home use, such as smartwatches.

Furthermore, methodologies that take feedback from patients into account, similar to how reinforcement or semi-supervised learning algorithms work, have thus far not been proposed or investigated thoroughly. Such a system would directly involve patients in the continuing improvement of a seizure detection model. For example, the model would be initially trained in a supervised manner on in-hospital data of a patient, and in the ambulatory setting a mobile application could notify the patient of a detected seizure. The patient would then give feedback if the detection was true or false, or if a seizure was missed and when it occurred. Based on those observations the model would then perpetually re-train itself on new data. Of course, the validation of such a system would still require ambulatory ground truth recordings, but current research shows that this is within reach.

In conclusion, patient-involved detection methodologies that individualize over time have the potential to eventually become the best option for ultra-long-term seizure detection. Future improvements of biosignal sensors, batteries, and the design of wearable devices will substantially advance the possibilities of a generic, ambulatory seizure detection system. Such a system could leverage ultra-long-term data and personal seizure logs as well as expert-labeled seizure episodes from in-hospital recordings to provide an automated seizure diary and alarm. And ultimately, this could even develop into a universal personal assistant that is tailored towards a specific patient's wants and needs.

List of Figures

List of Tables

List of Abbreviations

AC . . . . . alternating current
ACC . . . . . accelerometry
AED . . . . . anti-epileptic drug
ANS . . . . . autonomic nervous system
BVP . . . . . blood volume pulse
CE . . . . . Conformité Européenne
CI . . . . . confidence interval
CPS . . . . . complex partial seizure
CSV . . . . . comma-separated values
DBS . . . . . deep brain stimulation
DC . . . . . direct current
EC . . . . . ethics committee
ECG . . . . . electrocardiography
EDA . . . . . electrodermal activity
EEG . . . . . electroencephalography
EMG . . . . . electromyography
EMU . . . . . epilepsy monitoring unit
EU . . . . . European Union
FAR . . . . . false alarm rate
FAR24 . . . . . false alarm rate per 24 h
FARn . . . . . false alarm rate per night
FBTCS . . . . . focal to bilateral tonic-clonic seizure
FN . . . . . false negative
FP . . . . . false positive
FS . . . . . focal seizure
GBT . . . . . gradient boosted decision trees
GDPR . . . . . General Data Protection Regulation
GTCS . . . . . generalized tonic-clonic seizure
GYR . . . . . gyroscopy
HR . . . . . heart rate
HRV . . . . . heart rate variability
HTTPS . . . . . Hypertext Transfer Protocol Secure
ID . . . . . identifier
ILAE . . . . . International League Against Epilepsy
KCL . . . . . King's College Hospital London
LED . . . . . light-emitting diode
LOPO . . . . . leave-one-participant-out

LOSO . . . . . leave-one-seizure-out
MS . . . . . multiple sclerosis
OS . . . . . operating system
PD . . . . . Parkinson's disease
PGES . . . . . post-ictal generalized electroencephalography suppression
PI . . . . . post-ictal immobility
PNES . . . . . psychogenic non-epileptic seizure
PPG . . . . . photoplethysmography
PPV . . . . . positive predictive value
PWE . . . . . patient with epilepsy
RADAR . . . . . Remote Assessment of Disease and Relapse
RADAR-CNS . . . . . Remote Assessment of Disease and Relapse - Central Nervous System
RF . . . . . random forest
RNS . . . . . responsive neurostimulation
SCL . . . . . skin conductance level
SCRR . . . . . skin conductance response rate
SD . . . . . standard deviation
SDK . . . . . software development kit
SPS . . . . . simple partial seizure
SUDEP . . . . . sudden unexpected death in epilepsy
SVM . . . . . support vector machine
TCS . . . . . tonic-clonic seizure
TMP . . . . . skin temperature
TP . . . . . true positive
UKF . . . . . University Medical Center Freiburg
UTC . . . . . Coordinated Universal Time
vEEG . . . . . video-electroencephalography
VNS . . . . . vagus nerve stimulation

Bibliography

[1] Rajendra Kale. "Bringing epilepsy out of the shadows". In: *BMJ* 315.7099 (July 1997), pp. 2–3. DOI: 10.1136/bmj.315.7099.2 (cit. on p. 3).

[2] Gonzalo Alarcón and Antonio Valentín. *Introduction to Epilepsy*. Cambridge University Press, 2012, p. 605. ISBN: 9781107376359 (cit. on p. 3).

[3] World Health Organization, International Bureau For Epilepsy, and International League Against Epilepsy. *Atlas: Epilepsy Care in the World*. World Health Organization, 2005, p. 91. ISBN: 9241563036 (cit. on p. 3).

[4] Emmanouil Magiorkinis, Kalliopi Sidiropoulou, and Aristidis Diamantis. "Hallmarks in the history of epilepsy: Epilepsy in antiquity". In: *Epilepsy & Behavior* 17.1 (Jan. 2010), pp. 103–108. DOI: 10.1016/j.yebeh.2009.10.023 (cit. on p. 3).

[5] D. F. Scott. "The discovery of anti-epileptic drugs". In: *Journal of the History of the Neurosciences* 1.2 (Apr. 1992), pp. 111–118. DOI: 10.1080/0964704920952522 (cit. on p. 3).

[6] Ryosuke Hanaya and Kazunori Arita. *The new antiepileptic drugs: Their neuropharmacology and clinical indications*. 2016. DOI: 10.2176/nmc.ra.2015-0344 (cit. on p. 3).

[7] Churl-Su Kwon, Ann Jacoby, Amza Ali, Joan Austin, Gretchen L. Birbeck, et al. "Systematic review of frequency of felt and enacted stigma in epilepsy and determining factors and attitudes toward persons living with epilepsy—Report from the International League Against Epilepsy Task Force on Stigma in Epilepsy". In: *Epilepsia* 63.3 (Mar. 2022), pp. 573–597. DOI: 10.1111/epi.17135 (cit. on p. 3).

[8] Robert S. Fisher, Carlos Acevedo, Alexis Arzimanoglou, Alicia Bogacz, J. Helen Cross, et al. "ILAE Official Report: A practical clinical definition of epilepsy". In: *Epilepsia* 55.4 (2014), pp. 475–482. DOI: 10.1111/epi.12550 (cit. on p. 3).

[9] Samuel F. Berkovic, John C. Mulley, Ingrid E. Scheffer, and Steven Petrou. "Human epilepsies: interaction of genetic and acquired factors". In: *Trends in Neurosciences* 29.7 (July 2006), pp. 391–397. DOI: 10.1016/j.tins.2006.05.009 (cit. on p. 3).

[10] John F. Annegers, Walter A. Rocca, and W. Allen Hauser. "Causes of Epilepsy: Contributions of the Rochester Epidemiology Project". In: *Mayo Clinic Proceedings* 71.6 (June 1996), pp. 570–575. DOI: 10.4065/71.6.570 (cit. on p. 3).

[11] Alejandro Scaramelli, Patricia Braga, Andrea Avellanal, Alicia Bogacz, Claudia Camejo, et al. "Prodromal symptoms in epileptic patients: clinical characterization of the pre-ictal phase." In: *Seizure* 18.4 (May 2009), pp. 246–50. DOI: 10.1016/j.seizure.2008.10.007 (cit. on p. 3).

[12] Elisa Bruno, Andrea Biondi, and Mark P. Richardson. "Pre-ictal heart rate changes: A systematic review and meta-analysis". In: *Seizure* 55 (Feb. 2018), pp. 48–56. DOI: 10.1016/j.seizure.2018.01.003 (cit. on pp. 3, 6, 37).

[13] Marco Mula and Francesco Monaco. "Ictal and peri-ictal psychopathology." In: *Behavioural neurology* 24.1 (2011), pp. 21–5. DOI: 10.3233/BEN-2011-0314 (cit. on p. 4).

[14] Roland D Thijs, Rainer Surges, Terence J O'Brien, and Josemir W Sander. "Epilepsy in adults". In: *The Lancet* 393.10172 (Feb. 2019), pp. 689–701. DOI: 10.1016/S0140-6736(18)32596-0 (cit. on pp. 4, 8, 126).

[15] Ettore Beghi. "The Epidemiology of Epilepsy". In: *Neuroepidemiology* 54.Suppl. 2 (2020), pp. 185–191. DOI: 10.1159/000503831 (cit. on pp. 4, 6, 36, 65).

[16] World Health Organization. *Epilepsy*. Feb. 2022. URL: https://www.who.int/en/news-room/fact-sheets/detail/epilepsy (visited on June 11, 2023) (cit. on p. 4).

[17] Kirsten M. Fiest, Khara M. Sauro, Samuel Wiebe, Scott B. Patten, Churl-Su Kwon, et al. "Prevalence and incidence of epilepsy". In: *Neurology* 88.3 (Jan. 2017), pp. 296–303. DOI: 10.1212/WNL.0000000000003509 (cit. on p. 4).

[18] Olafur Sveinsson, Tomas Andersson, Sofia Carlsson, and Torbjörn Tomson. "The incidence of SUDEP". In: *Neurology* 89.2 (July 2017), pp. 170–177. DOI: 10.1212/WNL.0000000000004094 (cit. on pp. 4, 34).

[19] Cynthia Harden, Torbjörn Tomson, David Gloss, Jeffrey Buchhalter, J. Helen Cross, et al. "Practice guideline summary: Sudden unexpected death in epilepsy incidence rates and risk factors: Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology and the American Epilepsy Society". In: *Neurology* 88.17 (Apr. 2017), pp. 1674–1680. DOI: 10.1212/WNL.0000000000003685 (cit. on pp. 4, 6, 34, 87).

[20] Leone Ridsdale, Judith Charlton, Mark Ashworth, Mark P Richardson, and Martin C Gulliford. "Epilepsy mortality and risk factors for death in epilepsy: a population-based study". In: *British Journal of General Practice* 61.586 (May 2011), e271–e278. DOI: 10.3399/bjgp11X572463 (cit. on p. 4).

[21] Robert S. Fisher, Walter van Emde Boas, Warren Blume, Christian Elger, Pierre Genton, et al. "Epileptic seizures and epilepsy: definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE)." In: *Epilepsia* 46.4 (Apr. 2005), pp. 470–2. DOI: 10.1111/j.0013-9580.2005.66104.x (cit. on p. 4).

[22] Soheyl Noachtar and Astrid S. Peters. "Semiology of epileptic seizures: a critical review." In: *Epilepsy & behavior : E&B* 15.1 (May 2009), pp. 2–9. DOI: 10.1016/j.yebeh.2009.02.029 (cit. on pp. 4–6).

[23] Guadalupe Fernandez-Baca Vaca, Carlos L. Mayor, Naira García Losarcos, Jun T. Park, and Hans O. Lüders. "Epileptic seizure semiology in different age groups". In: *Epileptic Disorders* 20.3 (June 2018), pp. 179–188. DOI: 10.1684/epd.2018.0970 (cit. on pp. 4, 115, 117).

[24] H. Gastaut, O. Magnus, W. F. Caveness, J. K. Merlis, H. Landolt, et al. "A Proposed International Classification of Epileptic Seizures". In: *Epilepsia* 5.4 (Dec. 1964), pp. 297–306. DOI: 10.1111/j.1528-1157.1964.tb03337.x (cit. on p. 4).

[25] Commission on Classification and Terminology of the International League Against Epilepsy. "Proposal for Revised Clinical and Electroencephalographic Classification of Epileptic Seizures". In: *Epilepsia* 22.4 (Aug. 1981), pp. 489–501. DOI: 10.1111/j.1528-1157.1981.tb06159.x (cit. on pp. 4, 6).

[26] Anne T. Berg, Samuel F. Berkovic, Martin J. Brodie, Jeffrey Buchhalter, J. Helen Cross, et al. "Revised terminology and concepts for organization of seizures and epilepsies: Report of the ILAE Commission on Classification and Terminology, 2005-2009". In: *Epilepsia* 51.4 (Apr. 2010), pp. 676–685. DOI: 10.1111/j.1528-1167.2010.02522.x (cit. on p. 4).

[27] Robert S. Fisher, J. Helen Cross, Jacqueline A. French, Norimichi Higurashi, Edouard Hirsch, et al. "Operational classification of seizure types by the International League Against Epilepsy: Position Paper of the ILAE Commission for Classification and Terminology". In: *Epilepsia* 58.4 (Apr. 2017), pp. 522–530. DOI: 10.1111/epi.13670 (cit. on pp. 4–7, 38, 65, 104, 117, 128).

[28] Ann Jacoby, Dee Snape, and Gus A Baker. "Epilepsy and social identity: the stigma of a chronic neurological disorder." In: *The Lancet. Neurology* 4.3 (Mar. 2005), pp. 171–8. DOI: 10.1016/S1474-4422(05)01014-8 (cit. on pp. 5, 6).

[29] Katherine S. Eggleston, Bryan D. Olin, and Robert S. Fisher. *Ictal tachycardia: The head-heart connection.* 2014. DOI: 10.1016/j.seizure.2014.02.012 (cit. on pp. 5, 6, 37).

[30] Harry Christie, Wendyl D'Souza, Mark Cook, and Udaya Seneviratne. "Can semiology differentiate between bilateral tonic-clonic seizures of focal-onset and generalized-onset? A systematic review." In: *Epilepsy & behavior : E&B* 116 (Mar. 2021), p. 107769. DOI: 10.1016/j.yebeh.2021.107769 (cit. on p. 6).

[31] Krikor Tufenkjian and Hans O. Lüders. "Seizure semiology: its value and limitations in localizing the epileptogenic zone." In: *Journal of clinical neurology (Seoul, Korea)* 8.4 (Dec. 2012), pp. 243–50. DOI: 10.3988/jcn.2012.8.4.243 (cit. on p. 6).

[32] Robert S. Fisher, J. Helen Cross, Carol D'Souza, Jacqueline A. French, Sheryl R. Haut, et al. "Instruction manual for the ILAE 2017 operational classification of seizure types". In: *Epilepsia* 58.4 (Apr. 2017), pp. 531–542. DOI: 10.1111/epi.13671 (cit. on p. 6).

[33] Khalid Alqadi, Ram Sankaraneni, Ursula Thome, and Prakash Kotagal. "Semiology of hypermotor (hyperkinetic) seizures." In: *Epilepsy & behavior : E&B* 54 (Jan. 2016), pp. 137–41. DOI: 10.1016/j.yebeh.2015.11.017 (cit. on p. 6).

[34] Belén Abarrategui, Roberto Mai, Ivana Sartori, Stefano Francione, Veronica Pelliccia, et al. "Temporal lobe epilepsy: A never-ending story." In: *Epilepsy & behavior : E&B* 122 (Sept. 2021), p. 108122. DOI: 10.1016/j.yebeh.2021.108122 (cit. on p. 6).

[35] Pierre Genton, Pierre Thomas, Dorothee G A Kasteleijn-Nolst Trenité, Marco Tulio Medina, and Javier Salas-Puig. "Clinical aspects of juvenile myoclonic epilepsy." In: *Epilepsy & behavior : E&B* 28 Suppl 1 (July 2013), S8–14. DOI: 10.1016/j.yebeh.2012.10.034 (cit. on p. 6).

[36] M. Bellon, C. Walker, and C. Peterson. "Seizure-related injuries and hospitalizations: self-report data from the 2010 Australian Epilepsy Longitudinal Survey." In: *Epilepsy & behavior : E&B* 26.1 (Jan. 2013), pp. 7–10. DOI: 10.1016/j.yebeh.2012.10.019 (cit. on p. 6).

[37] Martin Hirsch, Dirk-Matthias Altenmüller, and Andreas Schulze-Bonhage. "Latencies from intracranial seizure onset to ictal tachycardia: A comparison to surface EEG patterns and other clinical signs". In: *Epilepsia* 56.10 (Oct. 2015), pp. 1639–1647. DOI: 10.1111/epi.13117 (cit. on pp. 7, 37).

[38] Giovanna Calandra-Buonaura, Nicola Toschi, Federica Provini, Ivan Corazza, Francesca Bisulli, et al. "Physiologic autonomic arousal heralds motor manifestations of seizures in nocturnal frontal lobe epilepsy: Implications for pathophysiology". In: *Sleep Medicine* 13.3 (Mar. 2012), pp. 252–262. DOI: 10.1016/j.sleep.2011.11.007 (cit. on p. 7).

[39] Elisa Bruno, Andrea Biondi, Mark P. Richardson, and on behalf of the RADAR-CNS CNS Consortium. "Digital semiology and time-evolution pattern of bio-signals in focal onset motor seizures". In: *Seizure* 87.March (2021), pp. 114–120. DOI: 10.1016/j.seizure.2021.03.013 (cit. on p. 7).

[40] Rima El Atrache, Eleonora Tamilia, Fatemeh Mohammadpour Touserkani, Sarah Hammond, Christos Papadelis, et al. "Photoplethysmography: A measure for the function of the autonomic nervous system in focal impaired awareness seizures". In: *Epilepsia* (2020). DOI: 10.1111/epi.16621 (cit. on p. 7).

[41] Anouk van Westrhenen, Thomas De Cooman, Richard H.C. Lazeron, Sabine Van Huffel, and Roland D. Thijs. *Ictal autonomic changes as a tool for seizure detection: a systematic review*. 2019. DOI: 10.1007/s10286-018-0568-1 (cit. on pp. 7, 37).

[42] Solveig Vieluf, Claus Reinsberger, Rima El Atrache, Michele Jackson, Sarah Schubach, et al. "Autonomic nervous system changes detected with peripheral sensors in the setting of epileptic seizures". In: *Scientific Reports* 10.1 (2020), pp. 1–8. DOI: 10.1038/s41598-020-68434-z (cit. on pp. 7, 29, 39, 107, 122).

[43] Rani A Sarkis, Sigride Thome-Souza, Ming Zher Poh, Nichelle Llewellyn, Jacquelyn Klehm, et al. "Autonomic changes following generalized tonic clonic seizures: An analysis of adult and pediatric patients with epilepsy". In: *Epilepsy Research* 115 (Sept. 2015), pp. 113–118. DOI: 10.1016/j.eplepsyres.2015.06.005 (cit. on pp. 7, 29, 87, 122).

[44] Anne T. Berg. "Epilepsy, cognition, and behavior: The clinical picture." In: *Epilepsia* 52 Suppl 1 (Jan. 2011), pp. 7–12. DOI: 10.1111/j.1528-1167.2010.02905.x (cit. on p. 7).

[45] Christoph Helmstaedter and Juri-Alexander Witt.
"Epilepsy and cognition - A bidirectional relationship?" In: *Seizure* 49 (July 2017), pp. 83–89.
DOI: 10.1016/j.seizure.2017.02.017 (cit. on p. 7).

[46] Fangzhao Yin, Duanyu Ni, Cuiping Xu, Xiaoming Yan, Kai Ma, et al.
"Auras in intractable frontal lobe epilepsy: Clinical characteristics, values, and limitations."
In: *Epilepsy & behavior : E&B* 115 (Feb. 2021), p. 107724. DOI: 10.1016/j.yebeh.2020.107724
(cit. on p. 7).

[47] Tobias Loddenkemper and Prakash Kotagal. "Lateralizing signs during seizures in focal epilepsy."
In: *Epilepsy & behavior : E&B* 7.1 (Aug. 2005), pp. 1–17. DOI: 10.1016/j.yebeh.2005.04.004
(cit. on p. 7).

[48] Helena Mielke, Sonja Meissner, Kathrin Wagner, Andreas Joos, and Andreas Schulze-Bonhage.
"Which seizure elements do patients memorize? A comparison of history and seizure documentation."
In: *Epilepsia* 61.7 (July 2020), pp. 1365–1375. DOI: 10.1111/epi.16550 (cit. on p. 7).

[49] Kiran M. Kanth, Christopher S. Zimmerman, Sheela C. Toprani, and Masud Seyal. "Duration of postictal impaired awareness after bilateral tonic-clonic seizures: EEG and patient characteristics."
In: *Epilepsy & behavior : E&B* 128 (Mar. 2022), p. 108576. DOI: 10.1016/j.yebeh.2022.108576
(cit. on p. 7).

[50] Andreas Schulze-Bonhage, Elisa Bruno, Armin Brandt, Anthony Shek, Pedro Viana, et al.
"Diagnostic yield and limitations of in-hospital documentation in patients with epilepsy."
In: *Epilepsia* (May 2022). DOI: 10.1111/epi.17307 (cit. on pp. 7, 26, 33).

[51] Vernon L. Towle, José Bolaños, Diane Suarez, Kim Tan, Robert Grzeszczuk, et al.
"The spatial location of EEG electrodes: locating the best-fitting sphere relative to cortical anatomy."
In: *Electroencephalography and clinical neurophysiology* 86.1 (Jan. 1993), pp. 1–6.
DOI: 10.1016/0013-4694(93)90061-y (cit. on pp. 7, 32).

[52] Committee on methods of clinical examination in electroencephalography.
"Report of the committee on methods of clinical examination in electroencephalography".
In: *Electroencephalography and Clinical Neurophysiology* 10.2 (May 1958), pp. 370–375.
DOI: 10.1016/0013-4694(58)90053-1 (cit. on pp. 7, 32).

[53] Dean R. Freestone, Philippa J. Karoly, and Mark J. Cook.
"A forward-looking review of seizure prediction."
In: *Current opinion in neurology* 30.2 (Apr. 2017), pp. 167–173.
DOI: 10.1097/WCO.0000000000000429 (cit. on pp. 7, 39).

[54] Mark J. Cook, Terence J. O'Brien, Samuel F. Berkovic, Michael Murphy, Andrew Morokoff, et al.
"Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: A first-in-man study". In: *The Lancet Neurology* 12.6 (2013), pp. 563–571.
DOI: 10.1016/S1474-4422(13)70075-9 (cit. on pp. 7, 19, 20, 39, 45, 65, 104).

[55] Marcel Heers, Sebastian Böttcher, Adam Kalina, Stefan Katletz, Dirk-Matthias Altenmüller, et al.
"Detection of interictal epileptiform discharges in an extended scalp EEG array and high-density EEG – A prospective multicenter study". In: *Epilepsia* December 2021 (Mar. 2022), pp. 1–11.
DOI: 10.1111/epi.17246 (cit. on p. 8).

[56] Shirin Jamal Omidi, Johnson P. Hampson, and Samden D. Lhatoo.
"Long-term Home Video EEG for Recording Clinical Events." In: *Journal of clinical neurophysiology : official publication of the American Electroencephalographic Society* 38.2 (Mar. 2021), pp. 92–100.
DOI: 10.1097/WNP.0000000000000746 (cit. on pp. 8, 19, 26).

[57] Pedro F. Viana, Tal Pal Attia, Mona Nasseri, Jonas Duun-Henriksen, Andrea Biondi, et al.
"Seizure forecasting using minimally invasive, ultra-long-term subcutaneous electroencephalography: Individualized intrapatient models." In: *Epilepsia* April (Apr. 2022), pp. 1–10.
DOI: 10.1111/epi.17252 (cit. on p. 8).

[58] Farrokh Manzouri, Lakshay Khurana, Kristina Kravalis, Thomas Stieglitz, Andreas Schulze-Bonhage, et al. "An Optimized EEG-Based Seizure Detection Algorithm for Implantable Devices".
In: *2021 10th International IEEE/EMBS Conference on Neural Engineering (NER)*. Vol. 2021-May.
IEEE, May 2021, pp. 995–998. ISBN: 978-1-7281-4337-8.
DOI: 10.1109/NER49283.2021.9441348 (cit. on p. 8).

[59] Jonathan Goldstein, Churl-Su Kwon, Michael Harmon, Jeffrey Buchhalter, Alison Kukla, et al.
"Seizure documentation in people living with epilepsy."
In: *Epilepsy & behavior : E&B* 125 (Dec. 2021), p. 108383. DOI: 10.1016/j.yebeh.2021.108383
(cit. on pp. 8, 19).

[60] Benjamin H. Brinkmann, Philippa J. Karoly, Ewan S. Nurse, Sonya B. Dumanis, Mona Nasseri, et al. "Seizure Diaries and Forecasting With Wearables: Epilepsy Monitoring Outside the Clinic". In: *Frontiers in Neurology* 12.July (July 2021), pp. 1–14. DOI: 10.3389/fneur.2021.690404 (cit. on pp. 8, 26, 39).

[61] Jerry J. Shih, Nathan B. Fountain, Susan T. Herman, Anto Bagic, Fred Lado, et al. "Indications and methodology for video-electroencephalographic studies in the epilepsy monitoring unit." In: *Epilepsia* 59.1 (Jan. 2018), pp. 27–36. DOI: 10.1111/epi.13938 (cit. on pp. 8, 25).

[62] Barbara C. Jobst and Gregory D. Cascino. "Resective epilepsy surgery for drug-resistant focal epilepsy: a review." In: *JAMA* 313.3 (Jan. 2015), pp. 285–93. DOI: 10.1001/jama.2014.17426 (cit. on p. 8).

[63] Herm J Lamberink, Willem M Otte, Ingmar Blümcke, Kees P J Braun, European Epilepsy Brain Bank writing group, et al. "Seizure outcome and use of antiepileptic drugs after epilepsy surgery according to histopathological diagnosis: a retrospective multicentre cohort study." In: *The Lancet. Neurology* 19.9 (Sept. 2020), pp. 748–757. DOI: 10.1016/S1474-4422(20)30220-9 (cit. on p. 8).

[64] Evren Burakgazi Dalkilic. "Neurostimulation Devices Used in Treatment of Epilepsy." In: *Current treatment options in neurology* 19.2 (Feb. 2017), p. 7. DOI: 10.1007/s11940-017-0442-9 (cit. on p. 8).

[65] Rhaya L Johnson and Christopher G Wilson. "A review of vagus nerve stimulation as a therapeutic intervention." In: *Journal of inflammation research* 11 (May 2018), pp. 203–213. DOI: 10.2147/JIR.S163248 (cit. on p. 8).

[66] Nasser Zangiabadi, Lady Diana Ladino, Farzad Sina, Juan Pablo Orozco-Hernández, Alexandra Carter, et al. "Deep Brain Stimulation and Drug-Resistant Epilepsy: A Review of the Literature." In: *Frontiers in neurology* 10 (June 2019), p. 601. DOI: 10.3389/fneur.2019.00601 (cit. on p. 8).

[67] Tobias Loddenkemper, Andrew Pan, Silvia Neme, Kenneth B. Baker, Ali R. Rezai, et al. "Deep brain stimulation in epilepsy." In: *Journal of clinical neurophysiology : official publication of the American Electroencephalographic Society* 18.6 (Nov. 2001), pp. 514–32. DOI: 10.1097/00004691-200111000-00002 (cit. on p. 8).

[68] Farrokh Manzouri, Simon Heller, Matthias Dümpelmann, Peter Woias, and Andreas Schulze-Bonhage. "A Comparison of Machine Learning Classifiers for Energy-Efficient Implementation of Seizure Detection." In: *Frontiers in systems neuroscience* 12.September (Sept. 2018), p. 43. DOI: 10.3389/fnsys.2018.00043 (cit. on p. 9).

[69] Sofie Carrette, Paul Boon, Mathieu Sprengers, Robrecht Raedt, and Kristl Vonck. "Responsive neurostimulation in epilepsy." In: *Expert review of neurotherapeutics* 15.12 (Dec. 2015), pp. 1445–54. DOI: 10.1586/14737175.2015.1113875 (cit. on p. 9).

[70] Steve Mann. "Wearable computing: a first step toward personal imaging". In: *Computer* 30.2 (1997), pp. 25–32. DOI: 10.1109/2.566147 (cit. on p. 10).

[71] S. Mann. "An historical account of the 'WearComp' and 'WearCam' inventions developed for applications in 'personal imaging'". In: *Digest of Papers. First International Symposium on Wearable Computers.* Cambridge, MA, USA: IEEE Comput. Soc, 1997, pp. 66–73. ISBN: 0-8186-8192-6. DOI: 10.1109/ISWC.1997.629921 (cit. on p. 10).

[72] Sara Katherine Simblett, Andrea Biondi, Elisa Bruno, Dominic Ballard, Amanda Stoneman, et al. "Patients' experience of wearing multimodal sensor devices intended to detect epileptic seizures: A qualitative analysis". In: *Epilepsy & Behavior* 102 (Jan. 2020), p. 106717. DOI: 10.1016/j.yebeh.2019.106717 (cit. on pp. 10, 26, 52, 119).

[73] Steven Boeckx, Wim van Paesschen, Brecht Bonte, and Jonathan Dan. "Live Demonstration: SeizeIT - A wearable multimodal epileptic seizure detection device". In: *2018 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. IEEE, Oct. 2018, pp. 1–1. ISBN: 978-1-5386-3603-9. DOI: 10.1109/BIOCAS.2018.8584738 (cit. on pp. 10, 32).

[74] Kaat Vandecasteele, Thomas De Cooman, Ying Gu, Evy Cleeren, Kasper Claes, et al. "Automated epileptic seizure detection based on wearable ECG and PPG in a hospital environment". In: *Sensors (Switzerland)* 17.10 (Oct. 2017), p. 2338. DOI: 10.3390/s17102338 (cit. on pp. 10, 32, 36, 38, 97, 104, 120, 121).

[75] Thomas L. Babb, Elmo Mariani, and Paul H. Crandall.
"An electronic circuit for detection of EEG seizures recorded with implanted electrodes".
In: *Electroencephalography and Clinical Neurophysiology* 37.3 (1974), pp. 305–308.
DOI: 10.1016/0013-4694(74)90036-4 (cit. on p. 13).

[76] J Gotman and P Gloor.
"Automatic recognition and quantification of interictal epileptic activity in the human scalp EEG."
In: *Electroencephalography and clinical neurophysiology* 41.5 (Nov. 1976), pp. 513–29.
DOI: 10.1016/0013-4694(76)90063-8 (cit. on p. 13).

[77] Tammy Jiang, Jaimie L. Gradus, and Anthony J. Rosellini.
"Supervised Machine Learning: A Brief Primer." In: *Behavior therapy* 51.5 (Sept. 2020), pp. 675–687.
DOI: 10.1016/j.beth.2020.05.002 (cit. on pp. 13, 14).

[78] A. K. Jain, M. N. Murty, and P. J. Flynn. "Data clustering".
In: *ACM Computing Surveys* 31.3 (Sept. 1999), pp. 264–323. DOI: 10.1145/331499.331504
(cit. on p. 13).

[79] Jesper E. van Engelen and Holger H. Hoos. "A survey on semi-supervised learning".
In: *Machine Learning* 109.2 (Feb. 2020), pp. 373–440. DOI: 10.1007/s10994-019-05855-6
(cit. on p. 13).

[80] Toshitaka Yamakawa, Miho Miyajima, Koichi Fujiwara, Manabu Kano, Yoko Suzuki, et al.
"Wearable Epileptic Seizure Prediction System with Machine-Learning-Based Anomaly Detection of Heart Rate Variability". In: *Sensors* 20.14 (July 2020), p. 3987. DOI: 10.3390/s20143987
(cit. on p. 13).

[81] Arthur Guez, Robert D. Vincent, Joelle Pineau, and Massimo Avoli.
"Adaptive treatment of epilepsy via batch-mode reinforcement learning".
In: *Proceedings of the National Conference on Artificial Intelligence*. Vol. 3. 2008, pp. 1671–1678.
ISBN: 9781577353683 (cit. on p. 14).

[82] Omer Gottesman, Fredrik Johansson, Matthieu Komorowski, Aldo Faisal, David Sontag, et al.
"Guidelines for reinforcement learning in healthcare." In: *Nature medicine* 25.1 (Jan. 2019), pp. 16–18.
DOI: 10.1038/s41591-018-0310-5 (cit. on p. 14).

[83] Georg Heinze, Christine Wallisch, and Daniela Dunkler.
"Variable selection - A review and recommendations for the practicing statistician."
In: *Biometrical journal* 60.3 (May 2018), pp. 431–449. DOI: 10.1002/bimj.201700067
(cit. on p. 14).

[84] Nadia Burkart and Marco F. Huber. "A Survey on the Explainability of Supervised Machine Learning".
In: *Journal of Artificial Intelligence Research* 70 (Jan. 2021), pp. 245–317.
DOI: 10.1613/jair.1.12228. arXiv: 2011.07876 (cit. on pp. 14, 16).

[85] Osisanwo F.Y, Akinsola J.E.T, Awodele O, Hinmikaiye J. O, Olakanmi O, et al.
"Supervised Machine Learning Algorithms: Classification and Comparison".
In: *International Journal of Computer Trends and Technology* 48.3 (June 2017), pp. 128–138.
DOI: 10.14445/22312803/IJCTT-V48P126 (cit. on p. 14).

[86] Amanpreet Singh, Narina Thakur, and Aakanksha Sharma.
"A review of supervised machine learning algorithms".
In: *Proceedings of the 10th INDIACom; 2016 3rd International Conference on Computing for Sustainable Global Development, INDIACom 2016* (2016), pp. 1310–1315 (cit. on p. 14).

[87] S. B. Kotsiantis. "Decision trees: a recent overview".
In: *Artificial Intelligence Review* 39.4 (Apr. 2013), pp. 261–283.
DOI: 10.1007/s10462-011-9272-4 (cit. on p. 14).

[88] Klaus-Robert Müller, Sebastian Mika, Gunnar Rätsch, Koji Tsuda, and Bernhard Schölkopf.
"An Introduction to Kernel-Based Learning Algorithms".
In: *Handbook of Neural Network Signal Processing*. Vol. 12. 2. Sept. 2001, pp. 4–1–4–40.
ISBN: 9781420038613. DOI: 10.1201/9781420038613.ch4 (cit. on p. 15).

[89] Leif Peterson. "K-nearest neighbor". In: *Scholarpedia* 4.2 (2009).
DOI: 10.4249/scholarpedia.1883 (cit. on p. 15).

[90] David J. Hand and Keming Yu. "Idiot's Bayes: Not So Stupid after All?"
In: *International Statistical Review / Revue Internationale de Statistique* 69.3 (Dec. 2001), p. 385.
DOI: 10.2307/1403452 (cit. on p. 15).

[91] A.K. Jain, Jianchang Mao, and K.M. Mohiuddin. "Artificial neural networks: a tutorial". In: *Computer* 29.3 (Mar. 1996), pp. 31–44. DOI: 10.1109/2.485891 (cit. on p. 15).

[92] Guangyu Robert Yang and Xiao-Jing Wang. "Artificial Neural Networks for Neuroscientists: A Primer." In: *Neuron* 107.6 (Sept. 2020), pp. 1048–1070. DOI: 10.1016/j.neuron.2020.09.005. arXiv: 2006.01001 (cit. on p. 15).

[93] Gérard Biau and Erwan Scornet. "A random forest guided tour". In: *Test* 25.2 (June 2016), pp. 197–227. DOI: 10.1007/s11749-016-0481-7. arXiv: 1511.05741 (cit. on p. 15).

[94] Leo Breiman. "Bagging predictors". In: *Machine Learning* 24.2 (Aug. 1996), pp. 123–140. DOI: 10.1007/BF00058655 (cit. on p. 15).

[95] Alexey Natekin and Alois Knoll. "Gradient boosting machines, a tutorial". In: *Frontiers in Neurorobotics* 7.DEC (2013). DOI: 10.3389/fnbot.2013.00021 (cit. on pp. 15, 69, 109).

[96] J. R. Quinlan. "Bagging, boosting, and C4.5". In: *Proceedings of the National Conference on Artificial Intelligence*. Vol. 1. 1996, pp. 725–730 (cit. on p. 15).

[97] Andreas Bulling, Ulf Blanke, and Bernt Schiele. "A tutorial on human activity recognition using body-worn inertial sensors". In: *ACM Computing Surveys* 46.3 (Jan. 2014), pp. 1–33. DOI: 10.1145/2499621 (cit. on pp. 16, 28).

[98] Christian Janiesch, Patrick Zschech, and Kai Heinrich. "Machine learning and deep learning". In: *Electronic Markets* 31.3 (Sept. 2021), pp. 685–695. DOI: 10.1007/s12525-021-00475-2 (cit. on p. 16).

[99] Sándor Beniczky, Philippa Karoly, Ewan Nurse, Philippe Ryvlin, and Mark Cook. "Machine learning and wearable devices of the future". In: *Epilepsia* (July 2020), epi.16555. DOI: 10.1111/epi.16555 (cit. on p. 19).

[100] Franz Brunnhuber, Jeremy Slater, Sushma Goyal, Devyani Amin, Gardar Thorvardsson, et al. "Past, Present and Future of Home video-electroencephalographic telemetry: A review of the development of in-home video-electroencephalographic recordings." In: *Epilepsia* 61 Suppl 1.S1 (Nov. 2020), S3–S10. DOI: 10.1111/epi.16578 (cit. on pp. 19, 26, 33).

[101] Rosalind Kandler, Athi Ponnusamy, and Claire Wragg. "Video ambulatory EEG: A good alternative to inpatient video telemetry?" In: *Seizure* 47 (Apr. 2017), pp. 66–70. DOI: 10.1016/j.seizure.2017.02.010 (cit. on p. 19).

[102] Christian E. Elger and Christian Hoppe. "Diagnostic challenges in epilepsy: seizure under-reporting and seizure detection". In: *The Lancet Neurology* 17.3 (Mar. 2018), pp. 279–288. DOI: 10.1016/S1474-4422(18)30038-3 (cit. on pp. 19, 25, 33, 45, 104).

[103] Elisa Bruno, Sebastian Böttcher, Pedro F. Viana, Marta Amengual-Gual, Boney Joseph, et al. "Wearable devices for seizure detection: Practical experiences and recommendations from the Wearables for Epilepsy And Research (WEAR) International Study Group". In: *Epilepsia* 62.10 (Oct. 2021), pp. 2307–2321. DOI: 10.1111/epi.17044 (cit. on pp. 21, 44, 45, 49, 54, 106, 111).

[104] Sebastian Böttcher, Elisa Bruno, Nikolay V Manyakov, Nino Epitashvili, Kasper Claes, et al. "Detecting Tonic-Clonic Seizures in Multimodal Biosignal Data From Wearables: Methodology Design and Validation". In: *JMIR mHealth and uHealth* 9.11 (Nov. 2021), e27674. DOI: 10.2196/27674 (cit. on pp. 21, 65, 68, 104, 106, 109).

[105] Elisa Bruno, Sebastian Böttcher, Andrea Biondi, Nino Epitashvili, Nikolay V. Manyakov, et al. "Post-ictal accelerometer silence as a marker of post-ictal immobility". In: *Epilepsia* 61.7 (July 2020), pp. 1397–1405. DOI: 10.1111/epi.16552 (cit. on pp. 21, 45, 65, 80, 83, 85, 122).

[106] Sebastian Böttcher, Nikolay V. Manyakov, Nino Epitashvili, Amos Folarin, Mark P. Richardson, et al. "Using multimodal biosignal data from wearables to detect focal motor seizures in individual epilepsy patients". In: *Proceedings of the 6th international Workshop on Sensor-based Activity Recognition and Interaction*. New York, NY, USA: ACM, Sept. 2019, pp. 1–6. ISBN: 9781450377140. DOI: 10.1145/3361684.3361687 (cit. on pp. 21, 66, 69, 95, 101, 104).

[107] Sebastian Böttcher, Elisa Bruno, Nino Epitashvili, Matthias Dümpelmann, Nicolas Zabler, et al. "Intra- and Inter-Subject Perspectives on the Detection of Focal Onset Motor Seizures in Epilepsy Patients". In: *Sensors* 22.9 (Apr. 2022), p. 3318. DOI: 10.3390/s22093318 (cit. on pp. 21, 104).

[108] Vaidehi Naganur, Shobi Sivathamboo, Zhibin Chen, Shitanshu Kusmakar, Ana Antonic-Baker, et al. "Automated seizure detection with noninvasive wearable devices: A systematic review and meta-analysis". In: *Epilepsia* May (May 2022), pp. 1–12. DOI: 10.1111/epi.17297 (cit. on pp. 24, 34, 36).

[109] Sándor Beniczky, Samuel Wiebe, Jesper Jeppesen, William O. Tatum, Milan Brazdil, et al. "Automated seizure detection using wearable devices: A clinical practice guideline of the International League Against Epilepsy and the International Federation of Clinical Neurophysiology". In: *Epilepsia* 62.3 (Mar. 2021), pp. 632–646. DOI: 10.1111/epi.16818 (cit. on pp. 24, 34, 35, 120, 121, 130).

[110] Philippe Ryvlin, Leila Cammoun, Ilona Hubbard, France Ravey, Sandor Beniczky, et al. "Noninvasive detection of focal seizures in ambulatory patients". In: *Epilepsia* 61.S1 (Nov. 2020), pp. 1–8. DOI: 10.1111/epi.16538 (cit. on pp. 24, 26, 34, 104).

[111] Jayoung Kim, Alan S. Campbell, Berta Esteban-Fernández de Ávila, and Joseph Wang. "Wearable biosensors for healthcare monitoring." In: *Nature biotechnology* 37.4 (Apr. 2019), pp. 389–406. DOI: 10.1038/s41587-019-0045-y (cit. on p. 25).

[112] Alfredo J. Perez and Sherali Zeadally. "Recent Advances in Wearable Sensing Technologies." In: *Sensors (Basel, Switzerland)* 21.20 (Oct. 2021), p. 6828. DOI: 10.3390/s21206828 (cit. on p. 25).

[113] Catherine Dinh-Le, Rachel Chuang, Sara Chokshi, and Devin Mann. "Wearable Health Technology and Electronic Health Record Integration: Scoping Review and Future Directions." In: *JMIR mHealth and uHealth* 7.9 (Feb. 2019), e12861. DOI: 10.2196/12861 (cit. on p. 25).

[114] Kristof Van Laerhoven, Benny P.L. Lo, Jason W.P. Ng, Surapa Thiemjarus, Rachel King, et al. "Medical healthcare monitoring with wearable and implantable sensors". In: *Pervasive Healthcare* (2004) (cit. on p. 25).

[115] Joshua M. Pevnick, Kade Birkeland, Raymond Zimmer, Yaron Elad, and Ilan Kedan. "Wearable technology for cardiology: An update and framework for the future". In: *Trends in Cardiovascular Medicine* 28.2 (Feb. 2018), pp. 144–150. DOI: 10.1016/j.tcm.2017.08.003 (cit. on p. 25).

[116] Philip Schmidt, Attila Reiss, Robert Dürichen, and Kristof Van Laerhoven. "Wearable-Based Affect Recognition-A Review." In: *Sensors (Basel, Switzerland)* 19.19 (Sept. 2019), p. 4079. DOI: 10.3390/s19194079 (cit. on pp. 25, 40).

[117] Selim R. Benbadis, Edward O'Neill, William O. Tatum, and Leanne Heriaud. "Outcome of prolonged video-EEG monitoring at a typical referral epilepsy center." In: *Epilepsia* 45.9 (Sept. 2004), pp. 1150–3. DOI: 10.1111/j.0013-9580.2004.14504.x (cit. on p. 25).

[118] Daniel F. Ghougassian, Wendyl D'Souza, Mark J. Cook, and Terence J. O'Brien. "Evaluating the utility of inpatient video-EEG monitoring." In: *Epilepsia* 45.8 (Aug. 2004), pp. 928–32. DOI: 10.1111/j.0013-9580.2004.51003.x (cit. on pp. 25, 33).

[119] Jonas Duun-Henriksen, Maxime Baud, Mark P. Richardson, Mark Cook, George Kouvas, et al. "A new era in electroencephalographic monitoring? Subscalp devices for ultra–long-term recordings". In: *Epilepsia* 61.9 (Sept. 2020), pp. 1805–1817. DOI: 10.1111/epi.16630 (cit. on pp. 26, 65).

[120] Ushtar Amin, Christopher T Primiani, Stephanie MacIver, Angélica Rivera-Cruz, Alfred T Frontera, et al. "Value of smartphone videos for diagnosis of seizures: Everyone owns half an epilepsy monitoring unit". In: *Epilepsia* 62.9 (Sept. 2021), epi.17001. DOI: 10.1111/epi.17001 (cit. on p. 26).

[121] Andrea Biondi, Viviana Santoro, Pedro F. Viana, Petroula Laiou, Deb K. Pal, et al. "Noninvasive mobile EEG as a tool for seizure monitoring and management: A systematic review". In: *Epilepsia* 63.5 (May 2022), pp. 1041–1063. DOI: 10.1111/epi.17220 (cit. on pp. 26, 34, 35).

[122] Jennifer Shum and Daniel Friedman. "Commercially available seizure detection devices: A systematic review". In: *Journal of the Neurological Sciences* 428.May (Sept. 2021), p. 117611. DOI: 10.1016/j.jns.2021.117611 (cit. on pp. 26, 34, 129).

[123] Judith van Andel, Roland D. Thijs, Al de Weerd, Johan Arends, and Frans Leijten.
Non-EEG based ambulatory seizure detection designed for home use: What is available and how will it influence epilepsy care? Apr. 2016. DOI: 10.1016/j.yebeh.2016.01.003 (cit. on p. 26).

[124] Jeroen H M Bergmann, Vikesh Chandaria, and Alison McGregor.
"Wearable and implantable sensors: the patient's perspective."
In: *Sensors (Basel, Switzerland)* 12.12 (Dec. 2012), pp. 16695–709. DOI: 10.3390/s121216695 (cit. on pp. 26, 32).

[125] Sara K. Simblett, Elisa Bruno, Sara Siddi, Faith Matcham, Loretta Giuliano, et al.
"Patient perspectives on the acceptability of mHealth technology for remote measurement and management of epilepsy: A qualitative analysis". In: *Epilepsy and Behavior* 97 (Aug. 2019), pp. 123–129. DOI: 10.1016/j.yebeh.2019.05.035 (cit. on pp. 26, 52).

[126] E. Bruno, A. Biondi, S. Thorpe, and M.P. Richardson.
"Patients self-mastery of wearable devices for seizure detection: A direct user-experience".
In: *Seizure* 81.May (Oct. 2020), pp. 236–240. DOI: 10.1016/j.seizure.2020.08.023 (cit. on p. 26).

[127] Levente Hadady, Péter Klivényi, Dániel Fabó, and Sándor Beniczky.
"Real-world user experience with seizure detection wearable devices in the home environment".
In: *Epilepsia* December 2021 (Feb. 2022), pp. 1–6. DOI: 10.1111/epi.17189 (cit. on p. 26).

[128] Lotte Schelde Olsen, Jonas Munch Nielsen, Charlotte Simonÿ, Troels Wesenberg Kjær, and Malene Beck.
"Wearables in real life: A qualitative study of experiences of people with epilepsy who use home seizure monitoring devices". In: *Epilepsy & Behavior* 125 (Dec. 2021), p. 108398.
DOI: 10.1016/j.yebeh.2021.108398 (cit. on p. 26).

[129] Anouk Engelgeer, Anouk van Westrhenen, Roland D Thijs, and Silvia M A A Evers.
"An economic evaluation of the NightWatch for children with refractory epilepsy: Insight into the cost-effectiveness and cost-utility." In: *Seizure* 101.March (Aug. 2022), pp. 156–161.
DOI: 10.1016/j.seizure.2022.08.003 (cit. on p. 26).

[130] Elisa Bruno, Pedro F. Viana, Michael R. Sperling, and Mark P. Richardson. "Seizure detection at home: Do devices on the market match the needs of people living with epilepsy and their caregivers?"
In: *Epilepsia* 61.S1 (Nov. 2020), S11–S24. DOI: 10.1111/epi.16521 (cit. on pp. 26, 46, 119, 123).

[131] Anouk van Westrhenen, Tessa Souhoka, Maaike E. Ballieux, and Roland D. Thijs.
"Seizure detection devices: Exploring caregivers' needs and wishes."
In: *Epilepsy & behavior : E&B* 116 (Mar. 2021), p. 107723. DOI: 10.1016/j.yebeh.2020.107723 (cit. on pp. 26, 119).

[132] Anouk van Westrhenen, Ben F. M. Wijnen, and Roland D. Thijs.
"Parental preferences for seizure detection devices: A discrete choice experiment."
In: *Epilepsia* 63.5 (May 2022), pp. 1152–1163. DOI: 10.1111/epi.17202 (cit. on p. 26).

[133] Apoorva Dwivedi and Gargi Khanna. "Sensitivity enhancement of a folded beam MEMS capacitive accelerometer-based microphone for fully implantable hearing application."
In: *Biomedizinische Technik. Biomedical engineering* 63.6 (Nov. 2018), pp. 699–708.
DOI: 10.1515/bmt-2016-0181 (cit. on p. 27).

[134] Johan B. A. M. Arends. "Movement-based seizure detection". In: *Epilepsia* 59 (June 2018), pp. 30–35.
DOI: 10.1111/epi.14053 (cit. on pp. 27, 35, 86).

[135] Hugo F. Posada-Quintero and Ki H. Chon.
"Innovations in Electrodermal Activity Data Collection and Signal Processing: A Systematic Review".
In: *Sensors* 20.2 (Jan. 2020), p. 479. DOI: 10.3390/s20020479 (cit. on pp. 27, 29, 107).

[136] Peter H. Charlton, Panicos A. Kyriacou, Jonathan Mant, Vaidotas Marozas, Phil Chowienczyk, et al.
"Wearable Photoplethysmography for Cardiovascular Monitoring".
In: *Proceedings of the IEEE* 110.3 (Mar. 2022), pp. 355–381. DOI: 10.1109/JPROC.2022.3149785 (cit. on pp. 27, 30).

[137] Toshiyo Tamura, Ming Huang, and Tatsuo Togawa.
"Current Developments in Wearable Thermometers".
In: *Advanced Biomedical Engineering* 7 (2018), pp. 88–99. DOI: 10.14326/abe.7.88 (cit. on pp. 27, 32).

[138] Sumit Majumder and M. Jamal Deen. "Smartphone Sensors for Health Monitoring and Diagnosis."
In: *Sensors (Basel, Switzerland)* 19.9 (May 2019), p. 2164. DOI: 10.3390/s19092164 (cit. on p. 27).

[139] A. Albarbar, A. Badri, Jyoti K. Sinha, and A. Starr. "Performance evaluation of MEMS accelerometers". In: *Measurement* 42.5 (June 2009), pp. 790–795. DOI: 10.1016/j.measurement.2008.12.002 (cit. on p. 27).

[140] D. Hollocher, X. Zhang, A. Sparks, S. Bart, W. Sawyer, et al. "A very low cost, 3-axis, MEMS accelerometer for consumer applications". In: *2009 IEEE Sensors*. IEEE, Oct. 2009, pp. 953–957. ISBN: 978-1-4244-4548-6. DOI: 10.1109/ICSENS.2009.5398189 (cit. on p. 28).

[141] Benoit Huyghe, Jan Doutreloigne, and Jan Vanfleteren. "3D orientation tracking based on unscented Kalman filtering of accelerometer and magnetometer data". In: *2009 IEEE Sensors Applications Symposium*. IEEE, Feb. 2009, pp. 148–152. ISBN: 978-1-4244-2786-4. DOI: 10.1109/SAS.2009.4801796 (cit. on p. 28).

[142] Jade Bayly, John Carino, Slavé Petrovski, Michelle Smit, Dilini A. Fernando, et al. "Time-frequency mapping of the rhythmic limb movements distinguishes convulsive epileptic from psychogenic nonepileptic seizures". In: *Epilepsia* 54.8 (Aug. 2013), pp. 1402–1408. DOI: 10.1111/epi.12207 (cit. on pp. 28, 40, 45).

[143] Tamara M.E. Nijsen, Ronald M. Aarts, Johan B.A.M. Arends, and Pierre J.M. Cluitmans. "Automated detection of tonic seizures using 3-D accelerometry". In: *IFMBE Proceedings*. Vol. 22. Springer Berlin Heidelberg, 2008, pp. 188–191. ISBN: 9783540892076. DOI: 10.1007/978-3-540-89208-3_47 (cit. on p. 28).

[144] Marcus Varanis, Anderson Silva, Arthur Mereles, and Robson Pederiva. "MEMS accelerometers for mechanical vibrations analysis: a comprehensive review with applications". In: *Journal of the Brazilian Society of Mechanical Sciences and Engineering* 40.11 (Nov. 2018), p. 527. DOI: 10.1007/s40430-018-1445-5 (cit. on p. 28).

[145] Fredrik Olsson, Manon Kok, Kjartan Halvorsen, and Thomas B. Schon. "Accelerometer calibration using sensor fusion with a gyroscope". In: *2016 IEEE Statistical Signal Processing Workshop (SSP)*. IEEE, June 2016, pp. 1–5. ISBN: 978-1-4673-7803-1. DOI: 10.1109/SSP.2016.7551836 (cit. on p. 28).

[146] Norhafizan Ahmad, Raja Ariffin Raja Ghazilla, Nazirah M. Khairi, and Vijayabaskar Kasi. "Reviews on Various Inertial Measurement Unit (IMU) Sensor Applications". In: *International Journal of Signal Processing Systems* 1.2 (2013), pp. 256–262. DOI: 10.12720/ijsps.1.2.256-262 (cit. on p. 28).

[147] Andreas Schulze-Bonhage, Sebastian Böttcher, Nicolas Zabler, Martin Glasstetter, and Matthias Dümpelmann. "Automatisierte Anfallsdetektion mit Wearables: Welche Technologien für welche Biosignale?" In: *Zeitschrift für Epileptologie* 34.3 (Aug. 2021), pp. 257–266. DOI: 10.1007/s10309-021-00421-4 (cit. on p. 29).

[148] L W Freedman, A S Scerbo, M E Dawson, A Raine, W O McClure, et al. "The relationship of sweat gland count to electrodermal activity." In: *Psychophysiology* 31.2 (Mar. 1994), pp. 196–200. DOI: 10.1111/j.1469-8986.1994.tb01040.x (cit. on p. 28).

[149] Mathias Benedek and Christian Kaernbach. "Physiological correlates and emotional specificity of human piloerection". In: *Biological Psychology* 86.3 (Mar. 2011), pp. 320–329. DOI: 10.1016/j.biopsycho.2010.12.012 (cit. on p. 29).

[150] Wolfram Boucsein. *Electrodermal Activity*. New York, New York, USA: Springer Science & Business Media, 2012, p. 618. ISBN: 978-1-4614-1125-3 (cit. on pp. 29, 87, 107).

[151] Hugo D. Critchley and Neil A. Harrison. "Visceral Influences on Brain and Behavior". In: *Neuron* 77.4 (Feb. 2013), pp. 624–638. DOI: 10.1016/j.neuron.2013.02.008 (cit. on p. 29).

[152] Marta Casanovas Ortega, Elisa Bruno, and Mark P Richardson. "Electrodermal activity response during seizures: A systematic review and meta-analysis." In: *Epilepsy & behavior : E&B* 134 (Aug. 2022), p. 108864. DOI: 10.1016/j.yebeh.2022.108864 (cit. on p. 29).

[153] M.- Z Poh, T Loddenkemper, C Reinsberger, N C Swenson, S Goyal, et al. "Autonomic changes with seizures correlate with postictal EEG suppression". In: *Neurology* 78.23 (Apr. 2012), pp. 1868–1876. DOI: 10.1212/WNL.0b013e318258f7f1 (cit. on pp. 29, 45, 67, 82, 87, 96, 97, 99).

[154] A. S. Anusha, S. P. Preejith, Tony J. Akl, Jayaraj Joseph, and Mohanasankar Sivaprakasam. "Dry Electrode Optimization for Wrist-based Electrodermal Activity Monitoring". In: *2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. IEEE, June 2018, pp. 1–6. ISBN: 978-1-5386-3392-2. DOI: 10.1109/MeMeA.2018.8438595 (cit. on p. 29).

[155] Wolfram Boucsein, Don C. Fowles, Sverre Grimnes, Gershon Ben-Shakhar, Walton T Roth, et al. "Publication recommendations for electrodermal measurements." In: *Psychophysiology* 49.8 (Aug. 2012), pp. 1017–34. DOI: 10.1111/j.1469-8986.2012.01384.x (cit. on p. 29).

[156] Oliver Pabst, Christian Tronstad, Sverre Grimnes, Don Fowles, and Ørjan G. Martinsen. "Comparison between the AC and DC measurement of electrodermal activity." In: *Psychophysiology* 54.3 (Mar. 2017), pp. 374–385. DOI: 10.1111/psyp.12803 (cit. on p. 29).

[157] Hugo F. Posada-Quintero, John P. Florian, Alvaro D. Orjuela-Cañón, Tomas Aljama-Corrales, Sonia Charleston-Villalobos, et al. "Power Spectral Density Analysis of Electrodermal Activity for Sympathetic Function Assessment". In: *Annals of Biomedical Engineering* 44.10 (Oct. 2016), pp. 3124–3135. DOI: 10.1007/s10439-016-1606-6 (cit. on p. 30).

[158] Sara Taylor, Natasha Jaques, Weixuan Chen, Szymon Fedor, Akane Sano, et al. "Automatic identification of artifacts in electrodermal activity data." In: *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference* 2015.August (Aug. 2015), pp. 1934–7. DOI: 10.1109/EMBC.2015.7318762 (cit. on p. 30).

[159] Mona Nasseri, Ewan Nurse, Martin Glasstetter, Sebastian Böttcher, Nicholas M. Gregg, et al. "Signal quality and patient experience with wearable devices for epilepsy management". In: *Epilepsia* 61.S1 (Nov. 2020), S25–S35. DOI: 10.1111/epi.16527 (cit. on pp. 30, 32, 50, 54, 107, 131).

[160] John Allen. *Photoplethysmography and its application in clinical physiological measurement*. 2007. DOI: 10.1088/0967-3334/28/3/R01 (cit. on p. 30).

[161] Aymen A. Alian and Kirk H. Shelley. "Photoplethysmography." In: *Best practice & research. Clinical anaesthesiology* 28.4 (Dec. 2014), pp. 395–406. DOI: 10.1016/j.bpa.2014.08.006 (cit. on p. 30).

[162] R. Rox Anderson and John A. Parrish. "The optics of human skin." In: *The Journal of investigative dermatology* 77.1 (July 1981), pp. 13–9. DOI: 10.1111/1523-1747.ep12479191 (cit. on p. 30).

[163] Yu Sun and Nitish Thakor. "Photoplethysmography Revisited: From Contact to Noncontact, From Point to Imaging." In: *IEEE transactions on bio-medical engineering* 63.3 (Mar. 2016), pp. 463–77. DOI: 10.1109/TBME.2015.2476337 (cit. on p. 30).

[164] A. N. Bashkatov, E. A. Genina, V. I. Kochubey, and V. V. Tuchin. "Optical properties of human skin, subcutaneous and mucous tissues in the wavelength range from 400 to 2000 nm". In: *Journal of Physics D: Applied Physics* 38.15 (Aug. 2005), pp. 2543–2555. DOI: 10.1088/0022-3727/38/15/004 (cit. on p. 30).

[165] Dirk J. Faber, Maurice C. G. Aalders, Egbert G. Mik, Brett A. Hooper, Martin J. C. van Gemert, et al. "Oxygen Saturation-Dependent Absorption and Scattering of Blood". In: *Physical Review Letters* 93.2 (July 2004), p. 028102. DOI: 10.1103/PhysRevLett.93.028102 (cit. on p. 30).

[166] Jihyoung Lee, Kenta Matsumura, Ken-ichi Yamakoshi, Peter Rolfe, Shinobu Tanaka, et al. "Comparison between red, green and blue light reflection photoplethysmography for heart rate monitoring during motion." In: *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference* 2013 (July 2013), pp. 1724–7. DOI: 10.1109/EMBC.2013.6609852 (cit. on p. 30).

[167] Yuka Maeda, Masaki Sekine, and Toshiyo Tamura. "The advantages of wearable green reflected photoplethysmography." In: *Journal of medical systems* 35.5 (Oct. 2011), pp. 829–34. DOI: 10.1007/s10916-010-9506-z (cit. on p. 30).

[168] Toshiyo Tamura. "Current progress of photoplethysmography and SPO2 for health monitoring." In: *Biomedical engineering letters* 9.1 (Feb. 2019), pp. 21–36. DOI: 10.1007/s13534-019-00097-w (cit. on p. 30).

[169] T. Y. Abay and P. A. Kyriacou. "Photoplethysmography for blood volumes and oxygenation changes during intermittent vascular occlusions."
In: *Journal of clinical monitoring and computing* 32.3 (June 2018), pp. 447–455.
DOI: 10.1007/s10877-017-0030-2 (cit. on p. 30).

[170] John Allen and Alan Murray. "Similarity in bilateral photoplethysmographic peripheral pulse wave characteristics at the ears, thumbs and toes." In: *Physiological measurement* 21.3 (Aug. 2000), pp. 369–77.
DOI: 10.1088/0967-3334/21/3/303 (cit. on p. 30).

[171] Mohamed Elgendi. "On the Analysis of Fingertip Photoplethysmogram Signals".
In: *Current Cardiology Reviews* 8.1 (June 2012), pp. 14–25. DOI: 10.2174/157340312801215782 (cit. on p. 30).

[172] Peter H. Charlton, Drew A. Birrenkott, Timothy Bonnici, Marco A. F. Pimentel, Alistair E. W. Johnson, et al. "Breathing Rate Estimation From the Electrocardiogram and Photoplethysmogram: A Review."
In: *IEEE reviews in biomedical engineering* 11 (2018), pp. 2–20.
DOI: 10.1109/RBME.2017.2763681 (cit. on p. 30).

[173] Brinnae Bent, Benjamin A. Goldstein, Warren A. Kibbe, and Jessilyn P. Dunn.
"Investigating sources of inaccuracy in wearable optical heart rate sensors".
In: *npj Digital Medicine* 3.1 (Dec. 2020), p. 18. DOI: 10.1038/s41746-020-0226-6 (cit. on pp. 30, 50, 107, 130).

[174] Sebastian Böttcher, Solveig Vieluf, Elisa Bruno, Boney Joseph, Nino Epitashvili, et al.
"Data quality evaluation in wearable monitoring." In: *Scientific reports* 12.1 (Dec. 2022), p. 21412.
DOI: 10.1038/s41598-022-25949-x (cit. on pp. 31, 47, 53, 131).

[175] Byung S Kim and Sun K Yoo.
"Motion artifact reduction in photoplethysmography using independent component analysis".
In: *IEEE Transactions on Biomedical Engineering* 53.3 (Mar. 2006), pp. 566–568.
DOI: 10.1109/TBME.2005.869784 (cit. on p. 31).

[176] Kristen M. Warren, Joshua R. Harvey, Ki H. Chon, and Yitzhak Mendelson.
"Improving pulse rate measurements during random motion using a wearable multichannel reflectance photoplethysmograph". In: *Sensors (Switzerland)* 16.3 (Mar. 2016), p. 342.
DOI: 10.3390/s16030342 (cit. on p. 31).

[177] Duy Dao, S. M.A. Salehizadeh, Yeonsik Noh, Jo Woon Chong, Chae Ho Cho, et al.
"A Robust Motion Artifact Detection Algorithm for Accurate Detection of Heart Rates from Photoplethysmographic Signals Using Time-Frequency Spectral Features".
In: *IEEE Journal of Biomedical and Health Informatics* 21.5 (Sept. 2017), pp. 1242–1253.
DOI: 10.1109/JBHI.2016.2612059 (cit. on p. 31).

[178] Hyonyoung Han, Min-Joon Kim, and Jung Kim.
"Development of real-time motion artifact reduction algorithm for a wearable photoplethysmography."
In: *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference* 2007 (Aug. 2007), pp. 1538–41. DOI: 10.1109/IEMBS.2007.4352596 (cit. on p. 31).

[179] Boreom Lee, Jonghee Han, Hyun Jae Baek, Jae Hyuk Shin, Kwang Suk Park, et al.
"Improved elimination of motion artifacts from a photoplethysmographic signal using a Kalman smoother with simultaneous accelerometry."
In: *Physiological measurement* 31.12 (Dec. 2010), pp. 1585–603.
DOI: 10.1088/0967-3334/31/12/003 (cit. on p. 31).

[180] Seyedfakhreddin Nabavi and Sharmistha Bhadra.
"A Robust Fusion Method for Motion Artifacts Reduction in Photoplethysmography Signal".
In: *IEEE Transactions on Instrumentation and Measurement* 69.12 (Dec. 2020), pp. 9599–9608.
DOI: 10.1109/TIM.2020.3006636 (cit. on p. 31).

[181] David Pollreisz and Nima TaheriNejad. "Detection and Removal of Motion Artifacts in PPG Signals".
In: *Mobile Networks and Applications* 27.2 (Apr. 2022), pp. 728–738.
DOI: 10.1007/s11036-019-01323-6 (cit. on p. 31).

[182] Barbara Cravello and Ada Ferri.
"Relationships between skin properties and environmental parameters."
In: *Skin research and technology : official journal of International Society for Bioengineering and the Skin (ISBS) [and] International Society for Digital Imaging of Skin (ISDIS) [and] International Society for Skin Imaging (ISSI)* 14.2 (May 2008), pp. 180–6. DOI: 10.1111/j.1600-0846.2007.00275.x (cit. on p. 31).

[183] A. A. Romanovsky. "Skin temperature: its role in thermoregulation". In: *Acta Physiologica* 210.3 (Mar. 2014), pp. 498–507. DOI: 10.1111/apha.12231 (cit. on p. 31).

[184] Jan A. J. Stolwijk and James D. Hardy. "Control of Body Temperature". In: *Comprehensive Physiology*. Wiley, Dec. 1977, pp. 45–68. DOI: 10.1002/cphy.cp090104 (cit. on p. 32).

[185] Alan Davidson, Arjan Buis, and Ivan Glesk. "Toward Novel Wearable Pyroelectric Temperature Sensor for Medical Applications". In: *IEEE Sensors Journal* 17.20 (Oct. 2017), pp. 6682–6689. DOI: 10.1109/JSEN.2017.2744181 (cit. on p. 32).

[186] Sebastian Pollandt and Thomas P. Bleck. "Thermoregulation in epilepsy". In: *Handbook of Clinical Neurology*. Vol. 157. 2018, pp. 737–747. DOI: 10.1016/B978-0-444-64074-1.00045-8 (cit. on p. 32).

[187] Hermann Stefan, Michael Feichtinger, and Andrew Black. "Autonomic phenomena of temperature regulation in temporal lobe epilepsy". In: *Epilepsy & Behavior* 4.1 (Feb. 2003), pp. 65–69. DOI: 10.1016/S1525-5050(02)00644-3 (cit. on p. 32).

[188] Furrukh Sana, Eric M. Isselbacher, Jagmeet P. Singh, E. Kevin Heist, Bhupesh Pathik, et al. "Wearable Devices for Ambulatory Cardiac Monitoring: JACC State-of-the-Art Review." In: *Journal of the American College of Cardiology* 75.13 (Apr. 2020), pp. 1582–1592. DOI: 10.1016/j.jacc.2020.01.046 (cit. on p. 32).

[189] Alexander J. Casson, Shelagh Smith, John S. Duncan, and Esther Rodriguez-Villegas. "Wearable EEG: what is it, why is it needed and what does it entail?" In: *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference* 2008.June (Aug. 2008), pp. 5867–70. DOI: 10.1109/IEMBS.2008.4650549 (cit. on p. 32).

[190] Alexander J. Casson. "Wearable EEG and beyond." In: *Biomedical engineering letters* 9.1 (Feb. 2019), pp. 53–71. DOI: 10.1007/s13534-018-00093-6 (cit. on p. 32).

[191] Kaat Vandecasteele, Thomas De Cooman, Jonathan Dan, Evy Cleeren, Sabine Van Huffel, et al. "Visual seizure annotation and automated seizure detection using behind-the-ear electroencephalographic channels". In: *Epilepsia* 61.4 (2020), pp. 766–775. DOI: 10.1111/epi.16470 (cit. on pp. 32, 37).

[192] Sigge Weisdorf, Jonas Duun-Henriksen, Marianne J. Kjeldsen, Frantz R. Poulsen, Sirin W. Gangstad, et al. "Ultra-long-term subcutaneous home monitoring of epilepsy-490 days of EEG from nine patients." In: *Epilepsia* 60.11 (Nov. 2019), pp. 2204–2214. DOI: 10.1111/epi.16360 (cit. on p. 32).

[193] Sudha Ramasamy and Archana Balan. "Wearable sensors for ECG measurement: a review". In: *Sensor Review* 38.4 (Sept. 2018), pp. 412–419. DOI: 10.1108/SR-06-2017-0110 (cit. on p. 32).

[194] Kirti Khunti. "Accurate interpretation of the 12-lead ECG electrode placement: A systematic review". In: *Health Education Journal* 73.5 (Sept. 2014), pp. 610–623. DOI: 10.1177/0017896912472328 (cit. on p. 32).

[195] Mohamed Elgendi, Björn Eskofier, Socrates Dokos, and Derek Abbott. "Revisiting QRS detection methodologies for portable, wearable, battery-operated, and wireless ECG systems." In: *PloS one* 9.1 (Jan. 2014). Ed. by Luís A. Nunes. Amaral, e84018. DOI: 10.1371/journal.pone.0084018 (cit. on p. 32).

[196] Michele Brignole, Panos Vardas, Ellen Hoffman, Heikki Huikuri, Angel Moya, et al. "Indications for the use of diagnostic implantable and external ECG loop recorders." In: *Europace : European pacing, arrhythmias, and cardiac electrophysiology : journal of the working groups on cardiac pacing, arrhythmias, and cardiac cellular electrophysiology of the European Society of Cardiology* 11.5 (May 2009), pp. 671–87. DOI: 10.1093/europace/eup097 (cit. on p. 33).

[197] Nino Isakadze and Seth S. Martin. "How useful is the smartwatch ECG?" In: *Trends in cardiovascular medicine* 30.7 (Oct. 2020), pp. 442–448. DOI: 10.1016/j.tcm.2019.10.010 (cit. on p. 33).

[198] Angkoon Phinyomark, Evan Campbell, and Erik Scheme. "Surface Electromyography (EMG) Signal Processing, Classification, and Practical Considerations". In: 2020, pp. 3–29. DOI: 10.1007/978-981-13-9097-5_1 (cit. on p. 33).

[199] I Mendez, B W Hansen, C M Grabow, E J L Smedegaard, N B Skogberg, et al.
"Evaluation of the Myo armband for the classification of hand motions".
In: *2017 International Conference on Rehabilitation Robotics (ICORR)*. Vol. 2017. IEEE, July 2017, pp. 1211–1214. ISBN: 978-1-5386-2296-4. DOI: 10.1109/ICORR.2017.8009414 (cit. on p. 33).

[200] Jhonatan Kobylarz, Jordan J. Bird, Diego R. Faria, Eduardo Parente Ribeiro, and Anikó Ekárt.
"Thumbs up, thumbs down: non-verbal human-robot interaction through real-time EMG classification via inductive and supervised transductive transfer learning".
In: *Journal of Ambient Intelligence and Humanized Computing* 11.12 (Dec. 2020), pp. 6021–6031.
DOI: 10.1007/s12652-020-01852-z (cit. on p. 33).

[201] Aatif M. Husain, Alan R. Towne, David K. Chen, Luke E. Whitmire, Shannon R. Voyles, et al.
"Differentiation of Epileptic and Psychogenic Nonepileptic Seizures Using Single-Channel Surface Electromyography". In: *Journal of Clinical Neurophysiology* 38.5 (Sept. 2021), pp. 432–438.
DOI: 10.1097/WNP.0000000000000703 (cit. on pp. 33, 40, 45).

[202] P. Visconti, F. Gaetani, G.A. Zappatore, and P. Primiceri.
"Technical Features and Functionalities of Myo Armband: An Overview on Related Literature and Advanced Applications of Myoelectric Armbands Mainly Focused on Arm Prostheses".
In: *International Journal on Smart Sensing and Intelligent Systems* 11.1 (Jan. 2018), pp. 1–25.
DOI: 10.21307/ijssis-2018-005 (cit. on p. 33).

[203] Sándor Beniczky, Isa Conradsen, Oliver Henning, Martin Fabricius, and Peter Wolf.
"Automated real-time detection of tonic-clonic seizures using a wearable EMG device".
In: *Neurology* 90.5 (Jan. 2018), e428–e434. DOI: 10.1212/WNL.0000000000004893 (cit. on pp. 33, 35).

[204] Evelien E. Geertsema, Roland D. Thijs, Therese Gutter, Ben Vledder, Johan B. Arends, et al.
"Automated video-based detection of nocturnal convulsive seizures in a residential care setting".
In: *Epilepsia* 59 (June 2018), pp. 53–60. DOI: 10.1111/epi.14050 (cit. on pp. 33, 35, 46).

[205] Jukka Peltola, Pabitra Basnyat, Sidsel Armand Larsen, Tim Østerkjaerhuus, Torsten Vinding Merinder, et al. "Semiautomated classification of nocturnal seizures using video recordings."
In: *Epilepsia* (Feb. 2022). DOI: 10.1111/epi.17207 (cit. on p. 33).

[206] Sidsel Armand Larsen, Daniella Terney, Tim Østerkjerhuus, Torsten Vinding Merinder, Kaapo Annala, et al. "Automated detection of nocturnal motor seizures using an audio-video system."
In: *Brain and behavior* July (Aug. 2022), e2737. DOI: 10.1002/brb3.2737 (cit. on p. 33).

[207] Pabitra Basnyat, Jussi Mäkinen, Jukka T. Saarinen, and Jukka Peltola.
"Clinical utility of a video/audio-based epilepsy monitoring system Nelli."
In: *Epilepsy & behavior : E&B* 133 (Aug. 2022), p. 108804. DOI: 10.1016/j.yebeh.2022.108804 (cit. on p. 33).

[208] Aditi P. Narechania, Irena I. Garić, Indranil Sen-Gupta, Mícheál P. Macken, Elizabeth E. Gerard, et al.
"Assessment of a quasi-piezoelectric mattress monitor as a detection system for generalized convulsions". In: *Epilepsy & Behavior* 28.2 (Aug. 2013), pp. 172–176.
DOI: 10.1016/j.yebeh.2013.04.017 (cit. on pp. 33, 46).

[209] Kate Van Poppel, Stephen P. Fulton, Amy McGregor, Michelle Ellis, Andrea Patters, et al.
"Prospective Study of the Emfit Movement Monitor."
In: *Journal of child neurology* 28.11 (Nov. 2013), pp. 1434–1436.
DOI: 10.1177/0883073812471858 (cit. on p. 33).

[210] Stephen Fulton, Kate Van Poppel, Amy McGregor, Michelle Ellis, Andrea Patters, et al.
"Prospective Study of 2 Bed Alarms for Detection of Nocturnal Seizures."
In: *Journal of child neurology* 28.11 (Nov. 2013), pp. 1430–1433.
DOI: 10.1177/0883073812462064 (cit. on p. 33).

[211] Marta Amengual-Gual, Adriana Ulate-Campos, and Tobias Loddenkemper. "Status epilepticus prevention, ambulatory monitoring, early seizure detection and prediction in at-risk patients."
In: *Seizure* 68 (May 2019), pp. 31–37. DOI: 10.1016/j.seizure.2018.09.013 (cit. on p. 33).

[212] Sándor Beniczky and Philippe Ryvlin.
"Standards for testing and clinical validation of seizure detection devices".
In: *Epilepsia* 59.October 2017 (June 2018), pp. 9–13. DOI: 10.1111/epi.14049 (cit. on pp. 34, 35, 54, 55, 105).

[213] Ilona Hubbard, Sandor Beniczky, and Philippe Ryvlin. "The Challenging Path to Developing a Mobile Health Device for Epilepsy: The Current Landscape and Where We Go From Here". In: *Frontiers in Neurology* 12.October (Oct. 2021), pp. 1–10. DOI: 10.3389/fneur.2021.740743 (cit. on p. 34).

[214] Dongni Johansson, Fredrik Ohlsson, David Krýsl, Bertil Rydenhag, Madeleine Czarnecki, et al. "Tonic-clonic seizure detection using accelerometry-based wearable sensors: A prospective, video-EEG controlled study". In: *Seizure* 65.September 2018 (Feb. 2019), pp. 48–54. DOI: 10.1016/j.seizure.2018.12.024 (cit. on pp. 35, 67, 78).

[215] Shitanshu Kusmakar, Chandan K Karmakar, Bernard Yan, Terence J O'Brien, Ramanathan Muthuganapathy, et al. "Detection of generalized tonic-clonic seizures using short length accelerometry signal". In: *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*. July 2017, pp. 4566–4569. ISBN: 9781509028092. DOI: 10.1109/EMBC.2017.8037872 (cit. on pp. 35, 67, 78, 96).

[216] Maromi Nei. "Cardiac Effects of Seizures". In: *Epilepsy Currents* 9.4 (July 2009), pp. 91–95. DOI: 10.1111/j.1535-7511.2009.01303.x (cit. on p. 35).

[217] Rainer Surges and Josemir W. Sander. "Sudden unexpected death in epilepsy: mechanisms, prevalence, and prevention." In: *Current opinion in neurology* 25.2 (Apr. 2012), pp. 201–7. DOI: 10.1097/WCO.0b013e3283506714 (cit. on p. 35).

[218] Fatemeh Mohammadpour Touserkani, Eleonora Tamilia, Francesca Coughlin, Sarah Hammond, Rima El Atrache, et al. "Photoplethysmographic evaluation of generalized tonic-clonic seizures". In: *Epilepsia* 61.8 (Aug. 2020), pp. 1606–1616. DOI: 10.1111/epi.16590 (cit. on p. 35).

[219] Jesper Jeppesen, Anders Fuglsang-Frederiksen, Peter Johansen, Jakob Christensen, Stephan Wüstenhagen, et al. "Seizure detection based on heart rate variability using a wearable electrocardiography device". In: *Epilepsia* 60.10 (Sept. 2019), pp. 2105–2113. DOI: 10.1111/epi.16343 (cit. on pp. 35, 37, 104).

[220] Anca A. Arbune, Jesper Jeppesen, Isa Conradsen, Philippe Ryvlin, and Sándor Beniczky. "Peri-ictal heart rate variability parameters as surrogate markers of seizure severity". In: *Epilepsia* 61.S1 (2020), S55–S60. DOI: 10.1111/epi.16491 (cit. on p. 35).

[221] I C Zibrandtsen, P Kidmose, and T W Kjaer. "Detection of generalized tonic-clonic seizures from ear-EEG based on EMG analysis". In: *Seizure* 59 (July 2018), pp. 54–59. DOI: 10.1016/j.seizure.2018.05.001 (cit. on p. 35).

[222] Petri Ojanen, Andrew Knight, Anna Hakala, Julia Bondarchik, Soheyl Noachtar, et al. "An integrative method to quantitatively detect nocturnal motor seizures". In: *Epilepsy Research* 169.October 2020 (Jan. 2021), p. 106486. DOI: 10.1016/j.eplepsyres.2020.106486 (cit. on p. 35).

[223] Francesco Onorati, Giulia Regalia, Chiara Caborni, W. Curt LaFrance, Andrew S. Blum, et al. "Prospective Study of a Multimodal Convulsive Seizure Detection Wearable System on Pediatric and Adult Patients in the Epilepsy Monitoring Unit". In: *Frontiers in Neurology* 12.August (Aug. 2021), pp. 1–15. DOI: 10.3389/fneur.2021.724904 (cit. on pp. 35, 36, 122, 130).

[224] Francesco Onorati, Giulia Regalia, Chiara Caborni, Matteo Migliorini, Daniel Bender, et al. "Multicenter clinical assessment of improved wearable multimodal convulsive seizure detectors". In: *Epilepsia* 58.11 (Nov. 2017), pp. 1870–1879. DOI: 10.1111/epi.13899 (cit. on pp. 35, 36, 67, 77, 86, 90, 97, 121).

[225] Jianbin Tang, Rima El Atrache, Shuang Yu, Umar Asif, Michele Jackson, et al. "Seizure detection using wearable sensors and machine learning: Setting a benchmark". In: *Epilepsia* 62.8 (Aug. 2021), pp. 1807–1819. DOI: 10.1111/epi.16967 (cit. on pp. 36, 37).

[226] Johan Arends, Roland D. Thijs, Thea Gutter, Constantin Ungureanu, Pierre Cluitmans, et al. "Multimodal nocturnal seizure detection in a residential care setting A long-term prospective trial". In: *Neurology* 91.21 (Oct. 2018), E2010–E2019. DOI: 10.1212/WNL.0000000000006545 (cit. on pp. 36, 46, 52, 78, 104, 121, 129).

[227] Diana Cogan, Javad Birjandtalab, Mehrdad Nourani, Jay Harvey, and Venkatesh Nagaraddi. "Multi-Biosignal Analysis for Epileptic Seizure Monitoring". In: *International Journal of Neural Systems* 27.01 (Feb. 2017), p. 1650031. DOI: 10.1142/S0129065716500313 (cit. on pp. 36, 38, 97, 104, 120, 121).

[228] Elaine C. Wirrell, Brandon R. Grossardt, Lily C L Wong-Kisiel, and Katherine C. Nickels. "Incidence and classification of new-onset epilepsy and epilepsy syndromes in children in Olmsted County, Minnesota from 1980 to 2004: a population-based study." In: *Epilepsy research* 95.1-2 (June 2011), pp. 110–8. DOI: 10.1016/j.eplepsyres.2011.03.009 (cit. on p. 36).

[229] W. Allen Hauser, John F. Annegers, and Leonard T. Kurland. "Incidence of epilepsy and unprovoked seizures in Rochester, Minnesota: 1935-1984." In: *Epilepsia* 34.3 (May 1993), pp. 453–68. DOI: 10.1111/j.1528-1157.1993.tb02586.x (cit. on p. 36).

[230] Jonas Munch Nielsen, Ivan C. Zibrandtsen, Paolo Masulli, Torben Lykke Sørensen, Tobias S. Andersen, et al. "Towards a wearable multi-modal seizure detection system in epilepsy: A pilot study". In: *Clinical Neurophysiology* 136 (Apr. 2022), pp. 40–48. DOI: 10.1016/j.clinph.2022.01.005 (cit. on pp. 36, 38, 121).

[231] Christoph Baumgartner, Luke E. Whitmire, Shannon R. Voyles, and Damon P. Cardenas. "Using sEMG to identify seizure semiology of motor seizures". In: *Seizure* 86 (Mar. 2021), pp. 52–59. DOI: 10.1016/j.seizure.2020.11.013 (cit. on pp. 37, 121).

[232] Jelena Hyppönen, Anna Hakala, Kaapo Annala, Honglei Zhang, Jukka Peltola, et al. "Automatic assessment of the myoclonus severity from videos recorded according to standardized Unified Myoclonus Rating Scale protocol and using human pose and body movement analysis". In: *Seizure* 76.December 2019 (Mar. 2020), pp. 72–78. DOI: 10.1016/j.seizure.2020.01.014 (cit. on p. 37).

[233] Jesper Jeppesen, Anders Fuglsang-Frederiksen, Peter Johansen, Jakob Christensen, Stephan Wüstenhagen, et al. "Seizure detection using heart rate variability: A prospective validation study". In: *Epilepsia* 61.S1 (2020), S41–S46. DOI: 10.1111/epi.16511 (cit. on p. 37).

[234] Thomas De Cooman, Troels W Kjær, Sabine Van Huffel, and Helge B Sorensen. "Adaptive heart rate-based epileptic seizure detection using real-time user feedback". In: *Physiological Measurement* 39.1 (Jan. 2018), p. 14005. DOI: 10.1088/1361-6579/aaa216 (cit. on p. 38).

[235] Thomas De Cooman, Kaat Vandecasteele, Carolina Varon, Borbála Hunyadi, Evy Cleeren, et al. "Personalizing Heart Rate-Based Seizure Detection Using Supervised SVM Transfer Learning". In: *Frontiers in Neurology* 11.February (Feb. 2020), pp. 1–13. DOI: 10.3389/fneur.2020.00145 (cit. on pp. 38, 104).

[236] Kaat Vandecasteele, Thomas De Cooman, Christos Chatzichristos, Evy Cleeren, Lauren Swinnen, et al. "The power of ECG in multimodal patient-specific seizure monitoring: Added value to an EEG-based detector using limited channels". In: *Epilepsia* 62.10 (Oct. 2021), pp. 2333–2343. DOI: 10.1111/epi.16990 (cit. on p. 38).

[237] Jesper Jeppesen, Sándor Beniczky, Peter Johansen, Per Sidenius, and Anders Fuglsang-Frederiksen. "Detection of epileptic seizures with a modified heart rate variability algorithm based on Lorenz plot". In: *Seizure* 24.C (Jan. 2015), pp. 1–7. DOI: 10.1016/j.seizure.2014.11.004 (cit. on pp. 38, 97).

[238] Steven N. Baldassano, Benjamin H. Brinkmann, Hoameng Ung, Tyler Blevins, Erin C. Conrad, et al. "Crowdsourcing seizure detection: algorithm development and validation on human implanted device recordings". In: *Brain* 140.6 (June 2017), pp. 1680–1691. DOI: 10.1093/brain/awx098 (cit. on pp. 38, 56).

[239] Rachel E. Stirling, Mark J. Cook, David B. Grayden, and Philippa J. Karoly. "Seizure forecasting and cyclic control of seizures." In: *Epilepsia* 62 Suppl 1.S1 (Feb. 2021), S2–S14. DOI: 10.1111/epi.16541 (cit. on p. 39).

[240] William C. Stacey. "Seizure Prediction Is Possible-Now Let's Make It Practical." In: *EBioMedicine* 27 (Jan. 2018), pp. 3–4. DOI: 10.1016/j.ebiom.2018.01.006 (cit. on p. 39).

[241] Levin Kuhlmann, Klaus Lehnertz, Mark P. Richardson, Björn Schelter, and Hitten P. Zaveri. "Seizure prediction - ready for a new era." In: *Nature reviews. Neurology* 14.10 (Oct. 2018), pp. 618–630. DOI: 10.1038/s41582-018-0055-2 (cit. on p. 39).

[242] Maxime O. Baud, Jonathan K. Kleen, Emily A. Mirro, Jason C. Andrechak, David King-Stephens, et al. "Multi-day rhythms modulate seizure risk in epilepsy". In: *Nature Communications* 9.1 (Jan. 2018), pp. 1–10. DOI: 10.1038/s41467-017-02577-y (cit. on p. 39).

[243] Philippa J. Karoly, Vikram R. Rao, Nicholas M. Gregg, Gregory A. Worrell, Christophe Bernard, et al. "Cycles in epilepsy." In: *Nature reviews. Neurology* 17.5 (May 2021), pp. 267–284. DOI: 10.1038/s41582-021-00464-1 (cit. on p. 39).

[244] Helen J. Burgess, John Trinder, Young Kim, and David Luke. "Sleep and circadian influences on cardiac autonomic nervous system activity." In: *The American journal of physiology* 273.4 (Oct. 1997), H1761–8. DOI: 10.1152/ajpheart.1997.273.4.H1761 (cit. on p. 39).

[245] Braxton B. Wannamaker. "Autonomic nervous system and epilepsy." In: *Epilepsia* 26 Suppl 1.s1 (June 1985), S31–9. DOI: 10.1111/j.1528-1157.1985.tb05722.x (cit. on p. 39).

[246] Solveig Vieluf, Marta Amengual-Gual, Bo Zhang, Rima El Atrache, Claire Ufongene, et al. "Twenty-four-hour patterns in electrodermal activity recordings of patients with and without epileptic seizures". In: *Epilepsia* 62.4 (Apr. 2021), pp. 960–972. DOI: 10.1111/epi.16843 (cit. on pp. 39, 107).

[247] Solveig Vieluf, Rima El Atrache, Sarah Cantley, Michele Jackson, Justice Clark, et al. "Seizure-related differences in biosignal 24-h modulation patterns." In: *Scientific reports* 12.1 (Sept. 2022), p. 15070. DOI: 10.1038/s41598-022-18271-z (cit. on p. 39).

[248] Philippa J. Karoly, Rachel E. Stirling, Dean R. Freestone, Ewan S. Nurse, Matias I. Maturana, et al. "Multiday cycles of heart rate are associated with seizure likelihood: An observational cohort study." In: *EBioMedicine* 72 (Oct. 2021), p. 103619. DOI: 10.1016/j.ebiom.2021.103619 (cit. on p. 39).

[249] Christian Meisel, Rima El Atrache, Michele Jackson, Sarah Schubach, Claire Ufongene, et al. "Machine learning from wristband sensor data for wearable, noninvasive seizure forecasting". In: *Epilepsia* 61.12 (2020), pp. 2653–2666. DOI: 10.1111/epi.16719. arXiv: 1906.00511 (cit. on p. 39).

[250] Mona Nasseri, Tal Pal Attia, Boney Joseph, Nicholas M. Gregg, Ewan S. Nurse, et al. "Ambulatory seizure forecasting with a wrist-worn device using long-short term memory deep learning". In: *Scientific Reports* 11.1 (Dec. 2021), p. 21935. DOI: 10.1038/s41598-021-01449-2 (cit. on p. 39).

[251] Cristian Donos, Mihai Dragos Maliia, Matthias Dümpelmann, and Andreas Schulze-Bonhage. "Seizure onset predicts its type". In: *Epilepsia* 59.3 (Jan. 2018), pp. 650–660. DOI: 10.1111/epi.13997 (cit. on p. 39).

[252] Daniel M. Goldenholz, Shira R. Goldenholz, Juan Romero, Rob Moss, Haoqi Sun, et al. "Development and Validation of Forecasting Next Reported Seizure Using e-Diaries". In: *Annals of Neurology* (2020), pp. 1–8. DOI: 10.1002/ana.25812 (cit. on p. 39).

[253] Levin Kuhlmann, Philippa Karoly, Dean R Freestone, Benjamin H Brinkmann, Andriy Temko, et al. "Epilepsyecosystem.org: crowd-sourcing reproducible seizure prediction with long-term human intracranial EEG." In: *Brain : a journal of neurology* 141.9 (Aug. 2018), pp. 2619–2630. DOI: 10.1093/brain/awy210 (cit. on pp. 39, 56).

[254] Benjamin H. Brinkmann, Joost Wagenaar, Drew Abbot, Phillip Adkins, Simone C. Bosshard, et al. "Crowdsourcing reproducible seizure forecasting in human and canine epilepsy." In: *Brain : a journal of neurology* 139.Pt 6 (June 2016), pp. 1713–22. DOI: 10.1093/brain/aww045 (cit. on pp. 39, 56).

[255] Dona E. Cragar, David T R Berry, Toufic A. Fakhoury, Jean E. Cibula, and Frederick A. Schmitt. "A review of diagnostic techniques in the differential diagnosis of epileptic and nonepileptic seizures." In: *Neuropsychology review* 12.1 (Mar. 2002), pp. 31–64. DOI: 10.1023/a:1015491123070 (cit. on pp. 40, 117).

[256] Sándor Beniczky, Isa Conradsen, Mihai Moldovan, Poul Jennum, Martin Fabricius, et al. "Quantitative analysis of surface electromyography during epileptic and nonepileptic convulsive seizures". In: *Epilepsia* 55.7 (June 2014), pp. 1128–1134. DOI: 10.1111/epi.12669 (cit. on p. 40).

[257] Shitanshu Kusmakar, Ramanathan Muthuganapathy, Bernard Yan, Terence J. O'Brien, and Marimuthu Palaniswami. "Gaussian mixture model for the identification of psychogenic non-epileptic seizures using a wearable accelerometer sensor". In: *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Vol. 2016-Octob. IEEE, Aug. 2016, pp. 1006–1009. ISBN: 978-1-4577-0220-4. DOI: 10.1109/EMBC.2016.7590872 (cit. on pp. 40, 45, 54).

[258] Massimiliano Pau, Silvia Caggiari, Alessandro Mura, Federica Corona, Bruno Leban, et al. "Clinical assessment of gait in individuals with multiple sclerosis using wearable inertial sensors: Comparison with patient-based measure." In: *Multiple sclerosis and related disorders* 10.October (Nov. 2016), pp. 187–191. DOI: 10.1016/j.msard.2016.10.007 (cit. on p. 40).

[259] Maddalena Sparaco, Luigi Lavorgna, Renata Conforti, Gioacchino Tedeschi, and Simona Bonavita. "The Role of Wearable Devices in Multiple Sclerosis." In: *Multiple sclerosis international* 2018.1 (Oct. 2018), p. 7627643. DOI: 10.1155/2018/7627643 (cit. on p. 40).

[260] Shaoxiong Sun, Amos A Folarin, Yuezhou Zhang, Nicholas Cummins, Shuo Liu, et al. "The utility of wearable devices in assessing ambulatory impairments of people with multiple sclerosis in free-living conditions." In: *Computer methods and programs in biomedicine* 227 (Oct. 2022), p. 107204. DOI: 10.1016/j.cmpb.2022.107204. arXiv: 2112.11903 (cit. on p. 40).

[261] Giampaolo Brichetto. "We should monitor our patients with wearable technology instead of neurological examination - Commentary." In: *Multiple sclerosis (Houndmills, Basingstoke, England)* 26.9 (Aug. 2020), pp. 1028–1030. DOI: 10.1177/1352458520930985 (cit. on p. 40).

[262] Sara K. Simblett, Joanne Evans, Ben Greer, Hannah Curtis, Faith Matcham, et al. "Engaging across dimensions of diversity: A cross-national perspective on mHealth tools for managing relapsing remitting and progressive multiple sclerosis." In: *Multiple sclerosis and related disorders* 32.January (July 2019), pp. 123–132. DOI: 10.1016/j.msard.2019.04.020 (cit. on p. 40).

[263] Marios Politis, Kit Wu, Sophie Molloy, Peter G Bain, K. Ray Chaudhuri, et al. "Parkinson's disease symptoms: the patient's perspective." In: *Movement disorders : official journal of the Movement Disorder Society* 25.11 (Aug. 2010), pp. 1646–51. DOI: 10.1002/mds.23135 (cit. on p. 40).

[264] Elina Kuosmanen, Florian Wolling, Julio Vega, Valerii Kan, Yuuki Nishiyama, et al. "Smartphone-Based Monitoring of Parkinson Disease: Quasi-Experimental Study to Quantify Hand Tremor Severity and Medication Effectiveness." In: *JMIR mHealth and uHealth* 8.11 (Nov. 2020), e21543. DOI: 10.2196/21543 (cit. on p. 40).

[265] Bjoern M Eskofier, Sunghoon I Lee, Jean-francois Daneault, Fatemeh N Golabchi, Gabriela Ferreira-Carvalho, et al. "Recent machine learning advancements in sensor-based mobility analysis: Deep learning for Parkinson's disease assessment". In: *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, Aug. 2016, pp. 655–658. ISBN: 978-1-4577-0220-4. DOI: 10.1109/EMBC.2016.7590787 (cit. on p. 40).

[266] Paola Pierleoni, Alberto Belli, Omid Bazgir, Lorenzo Maurizi, Michele Paniccia, et al. "A Smart Inertial System for 24h Monitoring and Classification of Tremor and Freezing of Gait in Parkinson's Disease". In: *IEEE Sensors Journal* 19.23 (Dec. 2019), pp. 11612–11623. DOI: 10.1109/JSEN.2019.2932584 (cit. on p. 40).

[267] Xueyi Wang, Joshua Ellul, and George Azzopardi. "Elderly Fall Detection Systems: A Literature Survey." In: *Frontiers in robotics and AI* 7.June (June 2020), p. 71. DOI: 10.3389/frobt.2020.00071 (cit. on p. 40).

[268] Tomoyasu Ikeno, Hideo Shigematsu, Masako Miyakoshi, Akira Ohba, Kazuichi Yagi, et al. "An analytic study of epileptic falls." In: *Epilepsia* 26.6 (Dec. 1985), pp. 612–21. DOI: 10.1111/j.1528-1157.1985.tb05701.x (cit. on p. 40).

[269] Evelien E. Geertsema, Gerhard H. Visser, Max A. Viergever, and Stiliyan N. Kalitzin. "Automated remote fall detection using impact features from video and audio". In: *Journal of Biomechanics* 88 (May 2019), pp. 25–32. DOI: 10.1016/j.jbiomech.2019.03.007 (cit. on pp. 40, 88).

[270] Fabio Bagalà, Clemens Becker, Angelo Cappello, Lorenzo Chiari, Kamiar Aminian, et al. "Evaluation of accelerometer-based fall detection algorithms on real-world falls." In: *PloS one* 7.5 (May 2012). Ed. by Antony Bayer, e37062. DOI: 10.1371/journal.pone.0037062 (cit. on p. 40).

[271] Akin Avci, Stephan Bosch, Mihai Marin-Perianu, Raluca Marin-Perianu, and Paul Havinga. "Activity recognition using inertial sensing for healthcare, wellbeing and sports applications: A survey". In: *23th International Conference on Architecture of Computing Systems 2010, ARCS 2010 - Workshop Proceedings*. 2010, pp. 167–176. ISBN: 9783800732227 (cit. on p. 40).

[272] M. Mähs, J. S. Pithan, I. Bergmann, L. Gabrys, J. Graf, et al.
"Activity tracker-based intervention to increase physical activity in patients with type 2 diabetes and healthy individuals: study protocol for a randomized controlled trial." In: *Trials* 23.1 (July 2022), p. 617. DOI: 10.1186/s13063-022-06550-z (cit. on p. 40).

[273] Giulia Regalia, Giulia Gerboni, Matteo Migliorini, Matteo Lai, Jonathan Pham, et al.
"Sleep assessment by means of a wrist actigraphy-based algorithm: agreement with polysomnography in an ambulatory study on older adults." In: *Chronobiology international* 38.3 (Mar. 2021), pp. 400–414. DOI: 10.1080/07420528.2020.1835942 (cit. on p. 40).

[274] Florian Wolling, Kristof Van Laerhoven, Jonas Bilal, Philipp M. Scholl, and Benjamin Volker.
"WetTouch: Touching Ground in the Wearable Detection of Hand-Washing Using Capacitive Sensing". In: *2022 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*. IEEE, Mar. 2022, pp. 769–774. ISBN: 978-1-6654-1647-4. DOI: 10.1109/PerComWorkshops53856.2022.9767345 (cit. on p. 40).

[275] Bjoern Eskofier, Sunghoon Lee, Manuela Baron, André Simon, Christine Martindale, et al.
"An Overview of Smart Shoes in the Internet of Health Things: Gait and Mobility Assessment in Health Promotion and Disease Monitoring". In: *Applied Sciences* 7.10 (Sept. 2017), p. 986. DOI: 10.3390/app7100986 (cit. on p. 40).

[276] David Duncker, Wern Yew Ding, Susan Etheridge, Peter A. Noseworthy, Christian Veltmann, et al.
"Smart Wearables for Cardiac Monitoring-Real-World Use beyond Atrial Fibrillation."
In: *Sensors (Basel, Switzerland)* 21.7 (Apr. 2021), p. 2539. DOI: 10.3390/s21072539 (cit. on p. 40).

[277] Giacomo Cappon, Giada Acciaroli, Martina Vettoretti, Andrea Facchinetti, and Giovanni Sparacino.
"Wearable Continuous Glucose Monitoring Sensors: A Revolution in Diabetes Treatment".
In: *Electronics* 6.3 (Sept. 2017), p. 65. DOI: 10.3390/electronics6030065 (cit. on p. 40).

[278] Philipp M. Scholl and Kristof van Laerhoven.
"A Feasibility Study of Wrist-Worn Accelerometer Based Detection of Smoking Habits". In: *2012 Sixth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*.
IEEE, July 2012, pp. 886–891. ISBN: 978-1-4673-1328-5. DOI: 10.1109/IMIS.2012.96
(cit. on p. 40).

[279] Beata Jarosiewicz and Martha Morrell.
"The RNS System: brain-responsive neurostimulation for the treatment of epilepsy".
In: *Expert Review of Medical Devices* 18.2 (Feb. 2021), pp. 129–138.
DOI: 10.1080/17434440.2019.1683445 (cit. on p. 45).

[280] Paul Boon, Kristl Vonck, Kenou van Rijckevorsel, Riem El Tahry, Christian E. Elger, et al. "A prospective, multicenter study of cardiac-based seizure detection to activate vagus nerve stimulation".
In: *Seizure* 32 (Nov. 2015), pp. 52–61. DOI: 10.1016/j.seizure.2015.08.011
(cit. on pp. 45, 104, 121).

[281] Abhinav V Kurada, Tarun Srinivasan, Sarah Hammond, Adriana Ulate-Campos, and Jonathan Bidwell.
Seizure detection devices for use in antiseizure medication clinical trials: A systematic review. Mar. 2019.
DOI: 10.1016/j.seizure.2019.02.007 (cit. on p. 45).

[282] Robert S. Fisher. "Bad information in epilepsy care". In: *Epilepsy & Behavior* 67 (Feb. 2017), pp. 133–134.
DOI: 10.1016/j.yebeh.2016.10.022 (cit. on p. 45).

[283] Sándor Beniczky, Isa Conradsen, Mihai Moldovan, Poul Jennum, Martin Fabricius, et al.
"Automated differentiation between epileptic and nonepileptic convulsive seizures".
In: *Annals of Neurology* 77.2 (Feb. 2015), pp. 348–351. DOI: 10.1002/ana.24338 (cit. on p. 45).

[284] Anca A. Arbune, Isa Conradsen, Damon P. Cardenas, Luke E. Whitmire, Shannon R. Voyles, et al.
"Ictal quantitative surface electromyography correlates with postictal EEG suppression".
In: *Neurology* 94.24 (June 2020), e2567–e2576. DOI: 10.1212/WNL.0000000000009492
(cit. on p. 45).

[285] Laura Vilella, Nuria Lacuey, Johnson P. Hampson, M.R. R.Sandhya Rani, Rup K. Sainju, et al.
"Postconvulsive central apnea as a biomarker for sudden unexpected death in epilepsy (SUDEP)".
In: *Neurology* 92.3 (Jan. 2019), E171–E182. DOI: 10.1212/WNL.0000000000006785
(cit. on p. 45).

[286] Charles Ákos Szabó, Lola C. Morgan, Kameel M. Karkar, Linda D. Leary, Octavian V. Lie, et al.
"Electromyography-based seizure detector: Preliminary results comparing a generalized tonic–clonic seizure detection algorithm to video-EEG recordings". In: *Epilepsia* 56.9 (Sept. 2015), pp. 1432–1437.
DOI: 10.1111/epi.13083 (cit. on p. 46).

[287] Chad Carlson, Vanessa Arnedo, Maria Cahill, and Orrin Devinsky. "Detecting nocturnal convulsions: Efficacy of the MP5 monitor". In: *Seizure* 18.3 (Apr. 2009), pp. 225–227. DOI: 10.1016/j.seizure.2008.08.007 (cit. on p. 46).

[288] PP Rickham. "Responsibility in Investigations on Human Subjects: M.R.C. Statement". In: *BMJ* 2.5402 (July 1964), pp. 178–180. DOI: 10.1136/bmj.2.5402.178 (cit. on p. 46).

[289] Lu Ann Aday. *At Risk in America*. John Wiley & Sons, 2001, pp. 1–372 (cit. on p. 46).

[290] William Hagman, David Andersson, Daniel Västfjäll, and Gustav Tinghög. "Public Views on Policies Involving Nudges". In: *Review of Philosophy and Psychology* 6.3 (Sept. 2015), pp. 439–453. DOI: 10.1007/s13164-015-0263-2 (cit. on p. 46).

[291] Yatharth Ranjan, Zulqarnain Rashid, Callum Stewart, Pauline Conde, Mark Begale, et al. "RADAR-base: Open source mobile health platform for collecting, monitoring, and analyzing data using sensors, wearables, and mobile devices". In: *Journal of Medical Internet Research* 21.8 (Aug. 2019), e11734. DOI: 10.2196/11734 (cit. on pp. 49, 59, 81, 82, 106).

[292] Elisa Bruno, Andrea Biondi, Sebastian Böttcher, Simon Lees, Andreas Schulze-Bonhage, et al. "Day and night comfort and stability on the body of four wearable devices for seizure detection: A direct user-experience". In: *Epilepsy & Behavior* 112 (Nov. 2020), p. 107478. DOI: 10.1016/j.yebeh.2020.107478 (cit. on pp. 50–52).

[293] Giulia Regalia, Francesco Onorati, Matteo Lai, Chiara Caborni, and Rosalind W. Picard. *Multimodal wrist-worn devices for seizure detection and advancing research: Focus on the Empatica wristbands*. July 2019. DOI: 10.1016/j.eplepsyres.2019.02.007 (cit. on pp. 51, 96, 122).

[294] A. Schulze-Bonhage, S. Böttcher, M. Glasstetter, N. Epitashvili, E. Bruno, et al. "Mobiles Anfallsmonitoring bei Epilepsiepatienten". In: *Der Nervenarzt* 90.12 (Dec. 2019), pp. 1221–1231. DOI: 10.1007/s00115-019-00822-x (cit. on p. 51).

[295] Frans S. S. Leijten, J. Andel, C. Ungureanu, J. Arends, F. Tan, et al. "Multimodal seizure detection: A review". In: *Epilepsia* 59.S1 (June 2018), pp. 42–47. DOI: 10.1111/epi.14047 (cit. on pp. 51, 66).

[296] Elisa Bruno, Sara Simblett, Alexandra Lang, Andrea Biondi, Clarissa Odoi, et al. "Wearable technology in epilepsy: The views of patients, caregivers, and healthcare professionals". In: *Epilepsy and Behavior* 85 (2018), pp. 141–149. DOI: 10.1016/j.yebeh.2018.05.044 (cit. on p. 52).

[297] Mary E. Thompson, Jennifer Langer, and Meheret Kinfe. "Seizure detection watch improves quality of life for adolescents and their families". In: *Epilepsy & Behavior* 98 (Sept. 2019), pp. 188–194. DOI: 10.1016/j.yebeh.2019.07.028 (cit. on p. 52).

[298] Pirgit Meritam, Philippe Ryvlin, and Sándor Beniczky. "User-based evaluation of applicability and usability of a wearable accelerometer device for detecting bilateral tonic-clonic seizures: A field study". In: *Epilepsia* 59.November 2017 (June 2018), pp. 48–52. DOI: 10.1111/epi.14051 (cit. on pp. 52, 80).

[299] Wynne W. Chin, Norman Johnson, and Andrew Schwarz. "A fast form approach to measuring technology acceptance and other constructs". In: *MIS Quarterly: Management Information Systems* 32.4 (2008), pp. 687–703. DOI: 10.2307/25148867 (cit. on p. 52).

[300] Mathias Benedek and Christian Kaernbach. "A continuous measure of phasic electrodermal activity". In: *Journal of Neuroscience Methods* 190.1 (2010), pp. 80–91. DOI: 10.1016/j.jneumeth.2010.04.028 (cit. on p. 54).

[301] I. A. Rezek and S. J. Roberts. "Stochastic complexity measures for physiological signal analysis". In: *IEEE Transactions on Biomedical Engineering* 45.9 (1998), pp. 1186–1191. DOI: 10.1109/10.709563 (cit. on p. 54).

[302] Amos Folarin, The Hyve, and The RADAR-CNS Consortium. *RADAR-base*. 2022. URL: https://github.com/RADAR-base (visited on June 11, 2023) (cit. on p. 59).

[303] Philippe Ryvlin, Lina Nashef, Samden D. Lhatoo, Lisa M. Bateman, Jonathan Bird, et al. "Incidence and mechanisms of cardiorespiratory arrests in epilepsy monitoring units (MORTEMUS): A retrospective study". In: *The Lancet Neurology* 12.10 (2013), pp. 966–977. DOI: 10.1016/S1474-4422(13)70214-X (cit. on pp. 65, 80, 87, 88).

[304] Christian Hoppe, Annkathrin Poepel, and Christian E. Elger. "Epilepsy: Accuracy of patient seizure counts". In: *Archives of Neurology* 64.11 (2007), pp. 1595–1599. DOI: 10.1001/archneur.64.11.1595 (cit. on pp. 65, 104).

[305] Andreas Schulze-Bonhage, Francisco Sales, Kathrin Wagner, Rute Teotonio, Astrid Carius, et al. "Views of patients with epilepsy on seizure prediction devices". In: *Epilepsy and Behavior* 18.4 (2010), pp. 388–396. DOI: 10.1016/j.yebeh.2010.05.008 (cit. on p. 65).

[306] Ying Gu, Evy Cleeren, Jonathan Dan, Kasper Claes, Wim Van Paesschen, et al. "Comparison between Scalp EEG and Behind-the-Ear EEG for Development of a Wearable Seizure Detection System for Patients with Focal Epilepsy". In: *Sensors* 18.2 (Dec. 2017), p. 29. DOI: 10.3390/s18010029 (cit. on p. 65).

[307] Isabelle Guyon and André Elisseeff. "An introduction to variable and feature selection". In: *Journal of Machine Learning Research* 3 (2003), pp. 1157–1182 (cit. on p. 67).

[308] Ulf Großekathöfer, Nikolay V Manyakov, Vojkan Mihajlović, Gahan Pandina, Andrew Skalkin, et al. "Automated detection of stereotypical motor movements in autism spectrum disorder using recurrence quantification analysis". In: *Frontiers in Neuroinformatics* 11 (Feb. 2017). DOI: 10.3389/fninf.2017.00009 (cit. on pp. 67, 106).

[309] Jerome H. Friedman. "Greedy function approximation: A gradient boosting machine". In: *Annals of Statistics* 29.5 (2001), pp. 1189–1232. DOI: 10.1214/aos/1013203451 (cit. on pp. 69, 109).

[310] Yoav Freund and Robert E. Schapire. "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting". In: *Journal of Computer and System Sciences* 55.1 (1997), pp. 119–139. DOI: 10.1006/jcss.1997.1504 (cit. on pp. 70, 110).

[311] MathWorks. *Estimates of predictor importance for classification ensemble of decision trees - MATLAB.* 2022. URL: https://mathworks.com/help/stats/compactclassificationensemble.predictorimportance.html (visited on June 11, 2023) (cit. on pp. 74, 110).

[312] Sándor Beniczky, Tilman Polster, Troels W. Kjaer, and Helle Hjalgrim. "Detection of generalized tonic-clonic seizures by a wireless wrist accelerometer: A prospective, multicenter study". In: *Epilepsia* 54.4 (Apr. 2013), e58–e61. DOI: 10.1111/epi.12120 (cit. on p. 80).

[313] Kris Cuppens, Peter Karsmakers, Anouk Van de Vel, Bert Bonroy, Milica Milosevic, et al. "Accelerometry-Based Home Monitoring for Detection of Nocturnal Hypermotor Seizures Based on Novelty Detection". In: *IEEE Journal of Biomedical and Health Informatics* 18.3 (May 2014), pp. 1026–1033. DOI: 10.1109/JBHI.2013.2285015 (cit. on p. 80).

[314] Juliana Lockman, Robert S. Fisher, and Donald M. Olson. "Detection of seizure-like movements using a wrist accelerometer". In: *Epilepsy & Behavior* 20.4 (Apr. 2011), pp. 638–641. DOI: 10.1016/j.yebeh.2011.01.019 (cit. on p. 80).

[315] Amy L. Patterson, Basanagoud Mudigoudar, Stephen Fulton, Amy McGregor, Kate Van Poppel, et al. "SmartWatch by SmartMonitor: Assessment of Seizure Detection Efficacy for Various Seizure Types in Children, a Large Prospective Single-Center Study". In: *Pediatric Neurology* 53.4 (Oct. 2015), pp. 309–311. DOI: 10.1016/j.pediatrneurol.2015.07.002 (cit. on pp. 80, 104, 121).

[316] Robert Bentley Todd. *Clinical lectures on paralysis, disease of the brain, and other affections of the nervous system.* Lindsay & Blakiston, 1855 (cit. on p. 80).

[317] Elisa Bruno and Mark P. Richardson. "Postictal generalized EEG suppression and postictal immobility: what do we know?" In: *Epileptic Disorders* 22.3 (June 2020), pp. 245–251. DOI: 10.1684/epd.2020.1158 (cit. on p. 80).

[318] Mira Semmelroch, Robert D. C. Elwes, Dora A. Lozsadi, and Lina Nashef.
"Retrospective audit of postictal generalized EEG suppression in telemetry".
In: *Epilepsia* 53.2 (Feb. 2012), e21–e24. DOI: 10.1111/j.1528-1167.2011.03296.x
(cit. on pp. 80, 87).

[319] James X. Tao, Iris Yung, Anthony Lee, Sandra Rose, John Jacobsen, et al. "Tonic phase of a generalized convulsive seizure is an independent predictor of postictal generalized EEG suppression".
In: *Epilepsia* 54.5 (May 2013), pp. 858–865. DOI: 10.1111/epi.12094 (cit. on pp. 80, 87).

[320] Masud Seyal, Lisa M. Bateman, and Chin-Shang Li. "Impact of periictal interventions on respiratory dysfunction, postictal EEG suppression, and postictal immobility".
In: *Epilepsia* 54.2 (Feb. 2013), pp. 377–382. DOI: 10.1111/j.1528-1167.2012.03691.x
(cit. on pp. 80, 87).

[321] Jonathan Kuo, Wenjie Zhao, Chin-Shang Li, Jeffrey D. Kennedy, and Masud Seyal. "Postictal immobility and generalized EEG suppression are associated with the severity of respiratory dysfunction".
In: *Epilepsia* 57.3 (Mar. 2016), pp. 412–417. DOI: 10.1111/epi.13312 (cit. on pp. 80, 82, 87).

[322] Marjan Asadollahi, Mahyar Noorbakhsh, Leila Simani, Mahtab Ramezani, and Kurosh Gharagozli. "Two predictors of postictal generalized EEG suppression: Tonic phase duration and postictal immobility period". In: *Seizure* 61.June (Oct. 2018), pp. 135–138.
DOI: 10.1016/j.seizure.2018.08.009 (cit. on pp. 80, 87).

[323] S. J. Purves, M. Wilson-Young, and V. P. Sweeney.
"Sudden death in epilepsy: single case report with video-EEG documentation".
In: *Epilepsia* 33.s3 (1992), p. 123 (cit. on pp. 80, 87, 88).

[324] J. M. Bird, K. A. T. Dembny, D. Sandeman, and S. Butler.
"Sudden Unexplained Death in Epilepsy: An Intracranially Monitored Case".
In: *Epilepsia* 38.SUPPL.11 (Nov. 1997), S52–S56.
DOI: 10.1111/j.1528-1167.1997.tb06295.x (cit. on pp. 80, 87, 88).

[325] B. N. McLean and S. Wimalaratna. "Sudden death in epilepsy recorded in ambulatory EEG".
In: *Journal of Neurology, Neurosurgery & Psychiatry* 78.12 (Dec. 2007), pp. 1395–1397.
DOI: 10.1136/jnnp.2006.088492 (cit. on pp. 80, 87, 88).

[326] Lisa M. Bateman, Mark Spitz, and Masud Seyal. "Ictal hypoventilation contributes to cardiac arrhythmia and SUDEP: Report on two deaths in video-EEG-monitored patients".
In: *Epilepsia* 51.5 (May 2010), pp. 916–920. DOI: 10.1111/j.1528-1167.2009.02513.x
(cit. on pp. 80, 87, 88).

[327] Samden D. Lhatoo, Howard J. Faulkner, Krystina Dembny, Kathy Trippick, Claire Johnson, et al.
"An electroclinical case-control study of sudden unexpected death in epilepsy".
In: *Annals of Neurology* 68.6 (Dec. 2010), pp. 787–796. DOI: 10.1002/ana.22101
(cit. on pp. 80, 82, 87, 88).

[328] James X. Tao, Shuo Qian, Maria Baldwin, Xiang-Jun Chen, Sandra Rose, et al.
"SUDEP, suspected positional airway obstruction, and hypoventilation in postictal coma".
In: *Epilepsia* 51.11 (Nov. 2010), pp. 2344–2347. DOI: 10.1111/j.1528-1167.2010.02719.x
(cit. on pp. 80, 87, 88).

[329] J. A. Liebenthal, S. Wu, S. Rose, J. S. Ebersole, and J. X. Tao.
"Association of prone position with sudden unexpected death in epilepsy".
In: *Neurology* 84.7 (Feb. 2015), pp. 703–709. DOI: 10.1212/WNL.0000000000001260
(cit. on pp. 80, 87, 88).

[330] Veriano Alexandre, Blanca Mercedes, Luc Valton, Louis Maillard, Fabrice Bartolomei, et al.
"Risk factors of postictal generalized EEG suppression in generalized convulsive seizures".
In: *Neurology* 85.18 (Nov. 2015), pp. 1598–1603. DOI: 10.1212/WNL.0000000000001949
(cit. on p. 81).

[331] Philippe Ryvlin, Carolina Ciumas, Ilona Wisniewski, and Sandor Beniczky.
"Wearable devices for sudden unexpected death in epilepsy prevention".
In: *Epilepsia* 59.December 2017 (June 2018), pp. 61–66. DOI: 10.1111/epi.14054 (cit. on p. 87).

[332] Michael A. Pfeifer, Clare R. Weinberg, Daniel Cook, James D. Best, Andrew Reenan, et al.
"Differential changes of autonomic nervous system function with age in man".
In: *The American Journal of Medicine* 75.2 (Aug. 1983), pp. 249–258.
DOI: 10.1016/0002-9343(83)91201-9 (cit. on p. 87).

[333] Milena Pavlova, Kanwaljit Singh, Myriam Abdennadher, Eliot S. Katz, Barbara A. Dworetzky, et al. "Comparison of cardiorespiratory and EEG abnormalities with seizures in adults and children". In: *Epilepsy & Behavior* 29.3 (Dec. 2013), pp. 537–541. DOI: 10.1016/j.yebeh.2013.09.026 (cit. on p. 87).

[334] Tomor Harnod, Cheryl C.H. Yang, Yue-Loong Hsin, Kun-Ruey Shieh, Pen-Jung Wang, et al. "Heart rate variability in children with refractory generalized epilepsy". In: *Seizure* 17.4 (June 2008), pp. 297–301. DOI: 10.1016/j.seizure.2007.09.002 (cit. on p. 87).

[335] William Richard Gowers. "Epilepsy and other Chronic Convulsive Diseases, Their Causes, Symptoms and Treatment". In: *The Journal of Nervous and Mental Disease* 29.3 (Mar. 1902), pp. 185–186. DOI: 10.1097/00005053-190203000-00023 (cit. on p. 87).

[336] Rosalind W. Picard, Matteo Migliorini, Chiara Caborni, Francesco Onorati, Giulia Regalia, et al. "Wrist sensor reveals sympathetic hyperactivity and hypoventilation before probable SUDEP". In: *Neurology* 89.6 (Aug. 2017), pp. 633–635. DOI: 10.1212/WNL.0000000000004208 (cit. on p. 87).

[337] Milica Milosevic, Anouk Van de Vel, Bert Bonroy, Berten Ceulemans, Lieven Lagae, et al. "Automated Detection of Tonic-Clonic Seizures Using 3-D Accelerometry and Surface Electromyography in Pediatric Patients." In: *IEEE journal of biomedical and health informatics* 20.5 (Sept. 2016), pp. 1333–1341. DOI: 10.1109/JBHI.2015.2462079 (cit. on p. 91).

[338] Rainer Surges, Roland D Thijs, Hanno L Tan, and Josemir W Sander. *Sudden unexpected death in epilepsy: Risk factors and potential pathomechanisms*. Aug. 2009. DOI: 10.1038/nrneurol.2009.118 (cit. on p. 96).

[339] Jonathan J Halford, Michael R Sperling, Dileep R Nair, Dennis J Dlugos, William O Tatum, et al. "Detection of generalized tonic–clonic seizures using surface electromyographic monitoring". In: *Epilepsia* 58.11 (Oct. 2017), pp. 1861–1869. DOI: 10.1111/epi.13897 (cit. on p. 96).

[340] Robert S. Fisher, Pegah Afra, Micheal Macken, Daniela N. Minecan, Anto Bagić, et al. "Automatic Vagus Nerve Stimulation Triggered by Ictal Tachycardia: Clinical Outcomes and Device Performance—The U.S. E-37 Trial". In: *Neuromodulation: Technology at the Neural Interface* 19.2 (Feb. 2016), pp. 188–195. DOI: 10.1111/ner.12376 (cit. on pp. 104, 121).

[341] Shitanshu Kusmakar, Chandan K. Karmakar, Bernard Yan, Terence J. O'Brien, Ramanathan Muthuganapathy, et al. "Automated Detection of Convulsive Seizures Using a Wearable Accelerometer Device". In: *IEEE Transactions on Biomedical Engineering* 66.2 (Feb. 2019), pp. 421–432. DOI: 10.1109/TBME.2018.2845865 (cit. on pp. 104, 120, 121).

[342] A. Ulate-Campos, F. Coughlin, M. Gaínza-Lein, I. Sánchez Fernández, P. L. Pearl, et al. *Automated seizure detection systems and their effectiveness for each type of seizure*. 2016. DOI: 10.1016/j.seizure.2016.06.008 (cit. on pp. 104, 130).

[343] Judith van Andel, Constantin Ungureanu, Johan Arends, Francis Tan, Johannes Van Dijk, et al. "Multimodal, automated detection of nocturnal motor seizures at home: Is a reliable seizure detector feasible?" In: *Epilepsia Open* 2.4 (Dec. 2017), pp. 424–431. DOI: 10.1002/epi4.12076 (cit. on pp. 104, 121).

[344] Uri Kramer, Svetlana Kipervasser, Arie Shlitner, and Ruben Kuzniecky. "A Novel Portable Seizure Detection Alarm System: Preliminary Results". In: *Journal of Clinical Neurophysiology* 28.1 (Feb. 2011), pp. 36–38. DOI: 10.1097/WNP.0b013e3182051320 (cit. on pp. 104, 121).

[345] J.-P Eckmann, S. Oliffson Kamphorst, and D. Ruelle. "Recurrence Plots of Dynamical Systems". In: *Europhysics Letters (EPL)* 4.9 (Nov. 1987), pp. 973–977. DOI: 10.1209/0295-5075/4/9/004 (cit. on p. 106).

[346] C. L. Webber and J. P. Zbilut. "Dynamical assessment of physiological systems and states using recurrence plot strategies". In: *Journal of Applied Physiology* 76.2 (Feb. 1994), pp. 965–973. DOI: 10.1152/jappl.1994.76.2.965 (cit. on p. 106).

[347] Jianjie Lu and Kai-Yu Tong.
"Robust Single Accelerometer-Based Activity Recognition Using Modified Recurrence Plot".
In: *IEEE Sensors Journal* 19.15 (Aug. 2019), pp. 6317–6324. DOI: 10.1109/JSEN.2019.2911204
(cit. on p. 106).

[348] Martin Glasstetter, Sebastian Böttcher, Nicolas Zabler, Nino Epitashvili, Matthias Dümpelmann, et al.
"Identification of Ictal Tachycardia in Focal Motor- and Non-Motor Seizures by Means of a Wearable PPG Sensor". In: *Sensors* 21.18 (Sept. 2021), p. 6017. DOI: 10.3390/s21186017
(cit. on pp. 107, 122, 129).

[349] Felix Scholkmann, Jens Boss, and Martin Wolf.
"An efficient algorithm for automatic peak detection in noisy periodic and quasi-periodic signals".
In: *Algorithms* 5.4 (2012), pp. 588–603. DOI: 10.3390/a5040588 (cit. on p. 107).

[350] Stephenie C. Lemon, Jason Roy, Melissa A. Clark, Peter D. Friedmann, and William Rakowski.
"Classification and regression tree analysis in public health: Methodological review and comparison with logistic regression". In: *Annals of Behavioral Medicine* 26.3 (Dec. 2003), pp. 172–181.
DOI: 10.1207/S15324796ABM2603_02 (cit. on p. 110).

[351] Beeke E. Heldberg, Thomas Kautz, Heike Leutheuser, Rudiger Hopfengartner, Burkhard S. Kasper, et al.
"Using wearable sensors for semiology-independent seizure detection - towards ambulatory monitoring of epilepsy". In: *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Vol. 2015-Novem. IEEE, Aug. 2015, pp. 5593–5596. ISBN: 978-1-4244-9271-8.
DOI: 10.1109/EMBC.2015.7319660 (cit. on p. 122).

[352] Ming Zher Poh, Tobias Loddenkemper, Claus Reinsberger, Nicholas C. Swenson, Shubhi Goyal, et al.
"Convulsive seizure detection using a wrist-worn electrodermal activity and accelerometry biosensor".
In: *Epilepsia* 53.5 (Mar. 2012), e93–e97. DOI: 10.1111/j.1528-1167.2012.03444.x
(cit. on p. 122).

[353] Lauren Swinnen, Christos Chatzichristos, Katrien Jansen, Lieven Lagae, Chantal Depondt, et al.
"Accurate detection of typical absence seizures in adults and children using a two-channel electroencephalographic wearable behind the ears." In: *Epilepsia* 62.11 (Nov. 2021), pp. 2741–2752.
DOI: 10.1111/epi.17061 (cit. on p. 129).

[354] Empatica Inc. *E4 Sunsetting – FAQs – Empatica Support*. 2022.
URL: https://support.empatica.com/hc/en-us/articles/6093022446 37-E4-Sunsetting-FAQs (visited on June 11, 2023) (cit. on p. 129).

[355] Kristof Van Laerhoven, Alexander Hoelzemann, Iris Pahmeier, Andrea Teti, and Lars Gabrys.
"Validation of an open-source ambulatory assessment system in support of replicable activity studies".
In: *German Journal of Exercise and Sport Research* 52.2 (June 2022), pp. 262–272.
DOI: 10.1007/s12662-022-00813-2 (cit. on p. 130).

[356] Stijn Luca, Peter Karsmakers, Kris Cuppens, Tom Croonenborghs, Anouk Van de Vel, et al.
"Detecting rare events using extreme value statistics applied to epileptic convulsions in children".
In: *Artificial Intelligence in Medicine* 60.2 (Feb. 2014), pp. 89–96.
DOI: 10.1016/j.artmed.2013.11.007 (cit. on p. 131).

[357] Nizar Islah, Jamie Koerner, Roman Genov, Taufik A. Valiante, and Gerard O'Leary.
"Machine Learning with Imbalanced EEG Datasets using Outlier-based Sampling". In: *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. Vol. 2020-July. IEEE, July 2020, pp. 112–115. ISBN: 978-1-7281-1990-8.
DOI: 10.1109/EMBC44109.2020.9175401 (cit. on p. 131).

[358] Oleg E. Karpov, Vadim V. Grubov, Vladimir A. Maksimenko, Semen A. Kurkin, Nikita M. Smirnov, et al.
"Extreme value theory inspires explainable machine learning approach for seizure detection."
In: *Scientific reports* 12.1 (Dec. 2022), p. 11474. DOI: 10.1038/s41598-022-15675-9
(cit. on p. 131).

Acknowledgments

First and foremost, I would like to express my sincerest gratitude to my thesis advisor Prof. Dr. Kristof Van Laerhoven, for the extensive, essential, and enduring support he has given me over the past few years. His guidance, inspiration, and reassurance were fundamental to my being able to complete this thesis, and I am profoundly grateful for his mentorship.

I would also like to thank Prof. Dr. Björn Eskofier for being the second examiner for my thesis, as well as the chair and associate chair of the board of examiners, Prof. Dr. Malte Lochau and Prof. Dr. Roman Obermaisser, for considering my work.

Furthermore, I would like to thank Prof. Dr. Andreas Schulze-Bonhage for enabling me to work in this field of research and on the resulting thesis, employing and supporting me for the past six years, and giving me the opportunities to travel to a multitude of different international conferences and meetings. Moreover, I want to thank PD Dr. Matthias Dümpelmann for his mentorship and constant advice and for always being my first place to go to in all things epilepsy and UKF. Speaking of which, I would also like to thank all my colleagues at UKF who I have met and worked with over the years, of which there are too many to name individually. Specifically, however, I would like to thank Nino Epitashvili, Lorenza Puma, Armin Brandt, the MTAs, and the clinicians, who have all supported and actively helped with the data collection procedures described in this thesis and without whom none of this would have been possible. In addition, thank you to Farrokh, Martin, Nico, and Fabio for always being helpful, insightful, and fun office mates.

Being a member of an international research project like RADAR-CNS naturally entails meeting numerous new colleagues and I would like to thank them all for being so absolutely welcoming and always open to collaboration. Special thanks go to Prof. Dr. Mark Richardson, Dr. Elisa Bruno, and Andrea Biondi at KCL for their uniquely clinical perspectives, Dr. Amos Folarin and the whole team at RADAR-base for their momentous work on, and support of, the technical data collection platform, Gergely Vértes for his illuminating skills in project management, and Dr. Nikolay "Kolya" Manyakov and Dr. Kasper Claes for their invaluable advice and perspectives on all things data science.

Making good, long-time friends does not happen often and I am deeply grateful to have met this group of wonderful people and their steadily growing families. Thank you, Benny, Jan, Johannes, Jonas, and Marc (a.k.a. the *ESE Guyz*)! Similarly, thank you as well to Dunja, Lino, and Thomas, hopefully there will be many a board game played in the future.

Lastly and most importantly, I would like to thank my family, my mom and dad and my sister, for always being there for me and helping me through the ups and downs that inherently accompany a Ph.D. student's life. Thank you!

Freiburg, June 2023
Sebastian Frank Calvin Böttcher