universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Data analysis in intracranial EEG recordings of epileptic patients to examine neuronal activity around interictal epileptiform discharges“

verfasst von / submitted by

Tamás Horváth

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2023 / Vienna, 2023

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 921
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Informatik
Betreut von / Supervisor:	Univ.-Prof. Dipl.-Ing. Dr.-Ing. Moritz Grosse-Wentrup

Acknowledgements

I would like to express my gratitude and appreciation to all people who helped me along the journey of my Master's course and writing this thesis. First and foremost, I would like to thank the professors and lecturers of the Faculty of Computer Science at the University of Vienna who provided me with comprehensive and valuable knowledge that has proven to be a great asset not only for this thesis, but also for my professional life. I owe gratitude to my supervisor, Prof. Moritz Grosse-Wentrup who guided me through the journey of writing this academic piece of work. His ideas and responsiveness contributed greatly to the fact that I can present this work. I would like to thank the people of the Epilepsy Centre at the Department of Neurology at the University Hospital Erlangen who, in any way, directly or indirectly, contributed to this thesis by providing all the recordings that served as the basis of the analysis. I would like to particularly highlight Dr. med. Johannes D. Lang who tirelessly kept answering all my questions as they came up throughout the evolution of this project. I truly appreciate all his efforts, eagerness, curiosity and professionalism. Without his inputs this thesis could have never been complete. Furthermore, I need to mention the speaker of the Epilepsy Centre, Prof. Dr. med. Hajo M. Hamer who provided constructive feedback in all our meetings discussing the progress of the project.

Tamás Horváth

Vienna, June 9, 2023

Abstract

Interictal epileptiform discharges (IED) are excessive surges of electrical cortical activity that can be detected in the interval between two successive seizures in patients with epilepsy. Utilizing intracranial electroencephalography (iEEG) the collective electrical activity of different cortical regions can be recorded. These recordings are obtained from two types of electrodes: high impedance 'microwire' contacts for the recording of single unit activity (SUA) and 'macro' contacts for capturing IEDs. Differentiating the types of neurons (pyramidal cells and interneurons) that fire around IEDs can shed light on the nature of IEDs and tell whether they are a product of inhibition or hyperexcitability. In this thesis we aim to explore how hippocampal neurons behave in the pre- and post-IED phase when IEDs occur in the different regions using data from ten patients with epilepsy. As existing IED detection methodologies are cumbersome, we introduce two novel methods that show great potential in classifying them: autoencoders and generative adversarial networks. Besides, the performance of three binary classifier models (logistic regression, SVM and artificial neural networks) that can differentiate the two types of neurons are compared. We argue that either of these binary classifier models can be superior to the k-means clustering algorithm when the intracluster distance is low, albeit the neural network outperforms the two classical models in terms of accuracy (95.27%) and precision (98.92%) averaged across all patients. The analyses revealed that neurons are inclined to fire in consistent patterns in the pre- and post-IED phase when IEDs occur within and near the hippocampus, however, no strong evidence of neuronal activity of the two cell types was found that is consistent in all patients.

Kurzfassung

Interiktale epileptiforme Entladungen (IED) sind übermäßige Stromstöße der kortikalen Aktivität, die im Intervall zwischen zwei aufeinanderfolgenden Anfällen bei Patient*innen mit Epilepsie festgestellt werden können. Mit Hilfe der intrakraniellen Elektroenzephalographie (iEEG) kann die kollektive elektrische Aktivität verschiedener kortikaler Regionen aufgezeichnet werden. Diese Aufzeichnungen erfolgen über zwei Arten von Elektroden: hochohmige "Mikrodraht"-Kontakte für die Aufzeichnung der Aktivität einzelner Einheiten (SUA) und "Makro"-Kontakte für die Erfassung von IED. Die Differenzierung der Neuronenarten (Pyramidenzellen und Interneuronen), die um IED herum feuern, kann Aufschluss über die Art der IED geben und zeigen, ob sie ein Produkt von Hemmung oder Übererregbarkeit sind. In dieser Arbeit wollen wir anhand von Daten von zehn Epilepsiepatient*innen untersuchen, wie sich die Neuronen des Hippocampus in der Prä- und Post-IED-Phase verhalten, wenn IED in den verschiedenen Regionen auftreten. Da die bestehenden Methoden zur Erkennung von IED umständlich sind, stellen wir zwei neue Methoden vor, die ein großes Potenzial zur Klassifizierung von IED aufweisen: Autocoder und generative adversarische Netzwerke. Außerdem wird die Leistung von drei binären Klassifizierungsmodellen (logistische Regression, SVM und künstliche neuronale Netze), die die beiden Arten von Neuronen unterscheiden können, verglichen. Wir argumentieren, dass jedes dieser Modelle dem k-means-Clusteralgorithmus überlegen sein kann, wenn die Intracluster-Distanz gering ist, obwohl das neuronale Netz die beiden klassischen Modelle in Bezug auf Genauigkeit ($95,27\%$) und Präzision ($98,92\%$) im Durchschnitt über alle Patient*innen übertrifft. Die Analysen ergaben, dass Neuronen dazu neigen, in der Prä- und Post-IED-Phase in übereinstimmenden Mustern zu feuern, wenn IED innerhalb und in der Nähe des Hippocampus auftreten, jedoch wurden keine eindeutigen Hinweise auf neuronale Aktivität der beiden Zelltypen gefunden, die bei allen Patient*innen übereinstimmen.

Contents

List of Tables

List of Figures

List of Algorithms

1. Introduction and Background

1.1. Introduction

Some say that the brain is the most complicated machine in the universe, and indeed, even today we know little about how this magnificent device works, and even though we do not possess a full understanding of its entire mechanism, current knowledge is already enormous that could easily fill libraries. Like any other organ, the brain is also susceptible to dysfunction. These may be caused by external physical factors such as trauma to the head, by inherited genetic factors, or by some other unknown reasons. These conditions may have different manifestations both within the brain and, consequently, in the behaviour of the affected person. For example, amongst the most common neurological diseases is Alzheimer's disease that shows visible structural degradation of the brain involving damage to the nervous system, and it is accompanied by negative changes to one's life as the disease progresses, ranging from mild memory loss to the person's inability to carry out daily activities. On the other hand, depression - a psychiatric condition - is not accompanied by similar structural changes, but has been attributed to chemical imbalance of neurotransmitters. However, recent research suggests that no consistent evidence exits to support this theory [1]. Epilepsy is yet another common neurological disorder affecting around 50 million people worldwide [2]. According to the International League Against Epilepsy [3], it is defined as a disorder of the brain characterized by an enduring predisposition to generate epileptic seizures, which has been applied as having two unprovoked seizures more than 24 hours apart. A more practical clinical approach extended the definition to include cases with one unprovoked seizure and an increased general recurrence risk as well as specific epilepsy syndromes [4]. The underlying cause remains unknown in many cases.

Electroencephalography (EEG) can be used to record brain activity via electrical currents, thus, it may provide insights into the mechanisms of epilepsy. It is well established that IEDs are a hallmark of epilepsy and therefore used for diagnosis [5]. Although IEDs have first been described decades ago, their true significance remains an enigma [6].

Epilepsy is an example of a condition that illustrates the need to better understand the brain. Even though the field has made significant progress in understanding many aspects of the brain's functionality, there is a lot of unknown ground and a necessity for multidisciplinary research involving medicine, neuroscience, and computer science to improve our understanding

In several cases there is a large amount of raw data collected from patients in various forms (in the current context it is EEG recordings), which needs translating into valuable

knowledge. This opens up possibilities for state-of-the-art data analytic methods which could potentially contribute to improved and faster diagnosis, and advancements in medical research. Furthermore, they motivate us to create and develop novel data science techniques.

1.2. Background

1.2.1. Epilepsy

Epilepsy is a brain disorder characterized by an enduring predisposition to generate epileptic seizures and by the neurobiologic, cognitive, psycho-logical, and social consequences of this condition and requires at least one epileptic seizures [3] [2]. These seizures are transient episodes of symptoms due to an abnormal excessive synchronous electrical activity of a group of neurons. While a single seizure may not necessarily indicate epilepsy and may occur as a symptom due to an acute brain condition, so-called acute symptomatic seizures, two (or more) unprovoked seizures fulfill the definition of the criteria of recurrence and constitute the diagnosis of epilepsy [3] [4] [7]. The current definition of epilepsy types includes mainly focal and generalized epilepsy, along with rare instances of combined generalized and focal epilepsy and an unknown epilepsy type [8]. The etiologies encompass genetic, structural (due to epileptogenic lesions or trauma), infectious, immune, metabolic, and unknown causes.

Epileptic seizures are classified according to their onset, that is the local source, and semiology, i.e. the corresponding signs and symptoms caused by the epileptic activity. The revised ILAE operational classification of seizure types differentiates focal, generalized, unknown onset and unclassified seizure types [9]. As the name implies, focal seizures (formerly also 'partial') originate in a circumscribed brain region and may remain regional or propagate, that is either migrate to another region or hemisphere, or spread to ultimately include both hemispheres (formerly also 'generalize'). During these seizures patients might be aware or have an impaired awareness, and show both motor and non-motor signs. The possible semiology includes automotor, hyperkinetic, cognitive, and many more symptoms, largely dependent on the localization of the epileptic activity. However, seizures with a so-called generalized onset are rather 'bilateral seizures' that feature a widespread epileptic activity early during the seizure and thus appear to involve the whole brain. They are usually unaware seizures featuring both, motor and non-motor symptoms in the form of tonic, clonic, tonic-clonic, myoclonic and absence seizures [9].

If a patient diagnosed with focal epilepsy does not respond to at least two drugs that have been appropriately chosen and administered in a sufficient dosage for an adequate time, the epilepsy is considered to 'drug resistant' [10]. In such cases, epilepsy surgery might be an option. The objective in epilepsy surgery is the resection or disconnection of the potential seizure onset. In cases of structural epilepsy, the surgical procedure usually includes a lesionectomy, i.e. the resection of the structural cause. To avoid postsurgical disabilities, the presurgical workup requires careful characterization and localization of the epileptogenic zone as well as good characterization of and clear delineation from functional

areas, such as memory, speech, motor and sensory functions and other important abilities.

1.2.2. Electroencephalography

The main diagnostic procedure in the diagnosis of epilepsy is the registration of the electrical activity of the brain called electroencephalography (EEG). Broadly speaking, there are two techniques of EEG: one is non-invasive, while the other is invasive.

In non-invasive EEG (often referred to as scalp EEG) [11], electrodes are placed on the scalp of the patient. Ease of use, cost efficiency and no need for any surgical procedures are its advantages. The downsides are, however, its poor temporal and spatial resolution, making it unsuitable for high frequency sampling (thus, some events can be potentially missed), and it can only record ensemble electrical activity of a very large population of neurons on or close to the surface of the brain surface (cortex). In summary, with non-invasive EEG, fine-grained events (e.g., single neuronal activity) cannot be examined deeper within the brain such as the hippocampus. Nevertheless, scalp EEG is still an efficient way to diagnose epilepsy.

The other type of EEG is invasive (intracranial EEG or iEEG) [11]. This method involves surgically placing electrodes within the skull. It can be further divided into two types: when an electrode array (or grid) is directly placed onto the cortex (electrocorticography - ECoG), and when electrodes are placed into deeper regions of the brain (also referred to as depth electrodes; stereotactic EEG - SEEG). The possibility of recording single (or an ensemble) neuronal activity with high temporal resolution deep within the brain is a great benefit. On the other hand, this method requires (risky) surgical procedures, and hence, the use of invasive methods must be well justified as in patients with drug resistant focal epilepsy.

Throughout this study, two different types of electrodes are distinguished: *macro* and *microwire* electrodes. The difference between the two is their capability to record at different resolutions. Macro contacts are used to capture ensembles of larger populations of neurons at lower sampling frequencies, and microwire electrodes are thin high impedance electrode wires that have the potential to record single neuronal activity at higher sampling frequencies. Each electrode can have multiple contacts, thus, they can record electrical activity concurrently at multiple locations of the brain. However, while macro SEEG contacts are circular electrodes that are usually placed in distinct distances along the shaft of the depth electrode, and their location can be radiologically examined, microwires are bundles of thin wires that protrude at the tip of a depth electrode and their exact location within the bundle cannot be determined.

1.2.3. Neurons and neuronal activity

The brain constitutes of highly interlinked nerve cells called neurons. On a high level, a typical neuron consists of a *soma* (cell body), *nucleus*, an *axon*, and *dendrites*. Multiple neurons are connected with one another through *synapses* (neuronal junctions), which connect the axon to the dendrites of other cells. Fundamentally, the dendrites are the inputs and the axon is the output of a neuron. In general two types of postsynaptic

potentials are differentiated: positively charged excitatory (EPSP) and negatively charged inhibitory (IPSP). While the EPSP facilitates the generation of an action potential (AP) on the postsynaptic membrane, the IPSP inhibits their generation. An AP is a brief increase of voltage, which when registered results in a sharp shape, a so-called spike. This is commonly termed as neuronal firing.

Cortical neurons can be largely divided into two cell types: projection neurons (pyramidal cells) and intrinsic neurons (interneurons) [12]. The difference is that projection neurons 'project' to another cortical or spinal area, whereas the intrinsic neurons are much more locally interconnected [12]. We can differentiate neurons based on their action: they can be either *excitatory* or *inhibitory*. When a neuron is excitatory, then the probability that its postsynaptic cells will fire upon receiving an electrical signal is high. An inhibitory neuron is the opposite of this. Interneurons are primarily inhibitory and pyramidal cells are excitatory.

Unit activity usually refers to timing of spikes elicited by individual neurons [13], and it is often referred to as single unit activity (SUA).

Recording SUA in vitro is possible with invasive methods. Figure 1.1 depicts possible extracellular recordings of a spike. The shape, polarity and amplitude changes depending on how close the probe is placed to the cell.

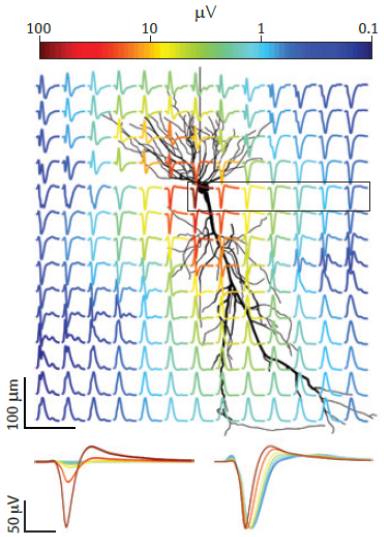


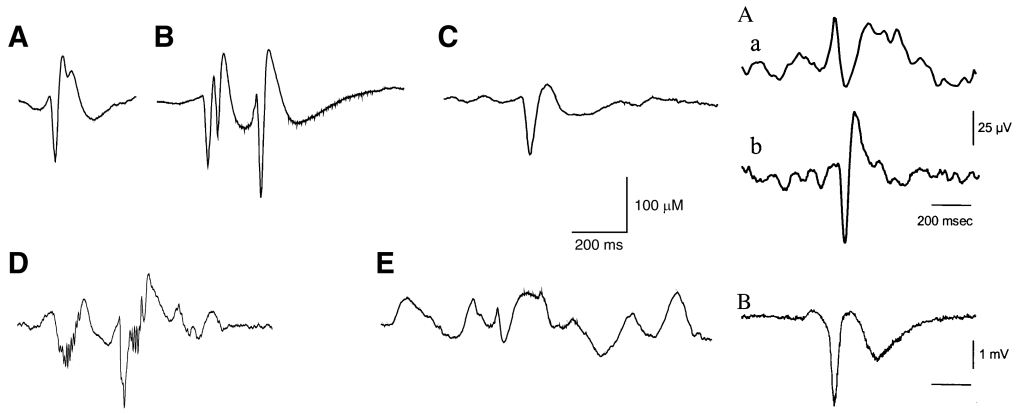
Figure 1.1.: Recorded spikes in the vicinity of a neuron. The shape, amplitude and polarity of the recording are highly dependent on the placement of the electrode. (Credit: [14])

1.2.4. Interictal epileptiform discharge

Epilepsy patients exhibit a common, characteristic electrical activity in their brain which is visible on EEG recordings and is not classified as a seizure. Such an electrical activity

is called an interictal (epileptiform) discharge (IED or ID) or an interictal spike (IIS or IS). An IED is a short, abnormal electrical discharge lasting for a couple of milliseconds (typically < 250 ms), and it occurs between seizures[1]. Such characteristic events are associated with epileptic disorders, have a distinct shape and are more frequent than seizures and therefore helpful in the diagnosis of epilepsy. IEDs are generated by the synchronous discharges of a group of neurons in a region referred to as the epileptic focus [5] and considered to resemble an epileptiform activity. While their true nature is still unknown, it is possible that IEDs actively inhibit seizures, or that a decreased IED rate is a secondary symptom of the brain approaching seizure [15], thus, interictal spikes could be imminent predictors for seizure events.

These discharges are very heterogeneous in nature as they can take on several shapes and forms with high level of diversity in their duration as well. Some resemble easily recognizable spikes with a clear voltage surge, while others may look like multiple components of waves added together. Figure 1.2 shows the diverse manifestations of IEDs.



(a) Credit: [16] (b) Credit: [17]

Figure 1.2.: Examples of IEDs. These electrical discharges occur between seizures and they are used as the basis for diagnosing epilepsy. They exhibit very diverse appearances on EEG which makes it hard for human experts and machines to detect them. IEDs can vary greatly even within patients.

1.2.5. Terminology

Terminology can vary in different literature and amongst experts in the field. Some do not differentiate between *action potential* and *spike*, whereas depending on the context, there can be a significant difference. An action potential is always the fast changing (rising then falling) electrical signal produced by exactly one neuron. A spike can be an action potential or a similar signal emitted by multiple neurons. Sometimes *IEDs* are

[1]Ictal: seizure, inter-: between.

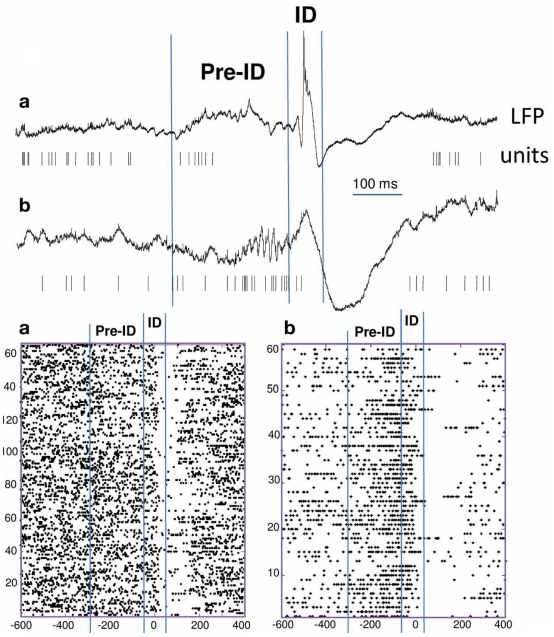
also referred to as *spikes* (interictal spike - IS or interictal discharge - ID), even though some may not exactly look like one.

1.3. Motivation

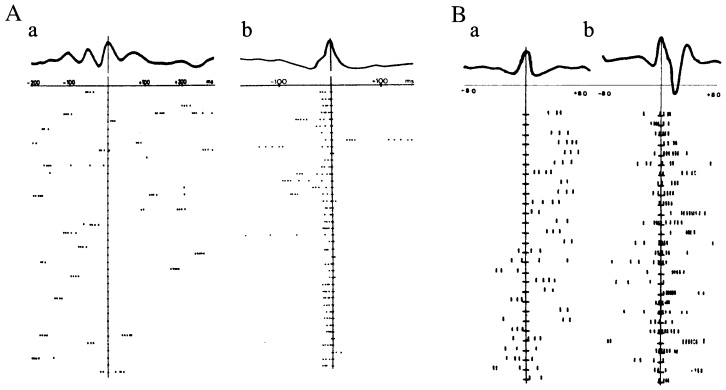
To better understand the nature and functions of IEDs, one needs to investigate the neurons that fire (synchronously) before, during and after IED events in different parts of the brain. Previous studies [18] [17] have shown that there is a correlation between neuronal firing and an IED occurring. Causal relationships, the role of brain regions and the relevance of different cell types were not covered. Figure 1.3 shows examples of the relationship between IEDs and SUA from these studies.

In this thesis some of these missing aspects are investigated. Concretely, we were interested in whether firing neurons are excitatory or inhibitory. Inhibition could suggest that firing cells try to counteract IEDs, and they are a rescuing effort of the brain. Furthermore, we look into the relationship of IEDs and SUA in different regions of the brain (hippocampus, temporal and extratemporal lobe). Additionally, problems with some existing methods are addressed with better approaches. A more robust clustering algorithm for spike detection [19] is used, and for the differentiation of the two cell types a benchmark of different algorithms is provided. Moreover, two novel methods are proposed for IED detection to overcome the difficulties of and flaws in existing methods [20].

This thesis aims to enrich knowledge within the medical field and to introduce state-of-the-art approaches to analyse large amounts of EEG data within the context of epilepsy.



(a) Credit: [18]



(b) Credit: [17]

Figure 1.3.: Different firing patterns of neurons can be observed around IEDs. a) There is significant correlation between IEDs and SUA. In the pre-IED phase, SUA is abundant, while in the post-IED phase it is sparse. b) In many cases such a correlation cannot be observed, and it can be dependent on several factors. Each line in the event plot represents an IED sample, and the dots represent the firing of a neuron.

2. Data and Methods

2.1. Patient data

The Epilepsy Centre at the Department of Neurology at the University Hospital Erlangen[1] provided anonymized raw iEEG data for ten distinct patients. Electrodes were implanted into the brain in order to record and localize epileptiform activity during a presurgical workup with the goal to define the area for resective epilepsy surgery. In stereotactic depth electrodes there's actually two types of contacts: macro and micro electrodes. The former is designed to capture IEDs covering a broader spectrum of the brain at a lower resolution (sampling frequency). While microwire electrodes aim to capture activity on a more granular level at a higher resolution, and hence, they can record SUA. These tiny wires are placed at the tip of an electrode. Figure 2.1 demonstrates an electrode and its recording capabilities. All recordings were achieved using Ad-Tech Medical[2] electrodes and the recording system manufactured by Neuralynx[3].

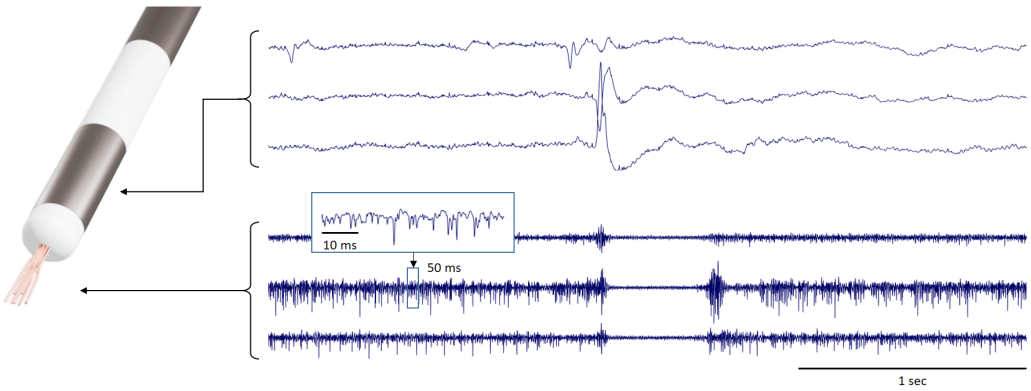


Figure 2.1.: The schematic view of an implanted Macro-Micro Depth Electrode. The gray-colored areas represent macro contacts, and inner microwire bundle extends past the tip of the macro depth electrode after the electrode has been implanted stereotactically to the region of interest. These two types of contacts are designed to record different activities in the brain. Credit: Dr. med. Johannes D. Lang, University Hospital Erlangen

The macro electrodes can record activity in the form of local field potentials along the

[1] https://www.epilepsiezentrum.uk-erlangen.de/
[2] https://adtechmedical.com/epilepsydepth-electrodes
[3] https://neuralynx.com/

trajectory of the depth electrode, while microwires can be protruded past the tip of the depth electrode into the adjacent tissue. When the macro electrode tip is placed in the hippocampus, the microwires can detect the local activity at the very tip of the electrode. Due to the 3D neuronavigation, the flexibility of the depth electrodes and the different compactness of the brain tissue that the electrode has to pass on a temporal trajectory to the hippocampus, the placement of the electrodes may not always be accurate, but usually varies within a few millimeters. For example, the first macro contact is supposed to be placed in the hippocampus or in direct proximity, but may sometimes be off and fail to reach the hippocampal formation which may result in a misplacement of the mircowire bundle outside the hippocampus, too.

All electrodes and contacts are labelled according to what parts of the brain they were implanted in. For example, RTA2 is interpreted as follows: the second contact of the first (A) electrode that is placed from the frontal to occipital lobe within the temporal lobe (T) on the right side (R). Or LPC6 is left (L), parietal (P), third (C) electrode from frontal to occipital, sixth contact. Even though the markings of electrodes are verbose, in this analysis we consider three major parts of the brain: hippocampal (temporomesial within the hippocampal formation), temporal (within the temporal lobe), and extratemoporal (outside the temporal lobe). Figure 2.2 shows the placements of electrodes on an MRI scan of one of the patients.

Table 2.1 summarizes the recordings of patients present in this thesis, and Table 2.2 lists where in the three major parts of the brain the macro contacts (for IEDs) are placed. Table 2.3 shows the microwire contacts considered for all patients. The channels fed from these electrodes show relevant electrical activity and were chosen after inspection by an expert physician.

Patient number	Sex	Age	Time of recording	Duration of recording
1	male	51	Dec 2019	05h 43m 24s
2	male	28	Jan 2020	04h 01m 28s
3	male	20	Oct 2020	04h 56m 03s
4	male	39	Jun 2021	17h 04m 14s
5	male	31	Nov 2020	03h 20m 07s
6	female	26	Nov 2019	05h 17m 40s
7	male	30	Oct 2019	05h 34m 30s
8	female	49	Jul 2019	01h 49m 54s
9	male	38	Apr 2019	01h 47m 06s
10	female	36	Nov 2021	04h 42m 05s

Table 2.1.: Patient data.

2.1.1. Data for training

The vast majority of the data provided was in raw format, and most of it did not come with labelling. However, some of the machine learning algorithms used in this thesis

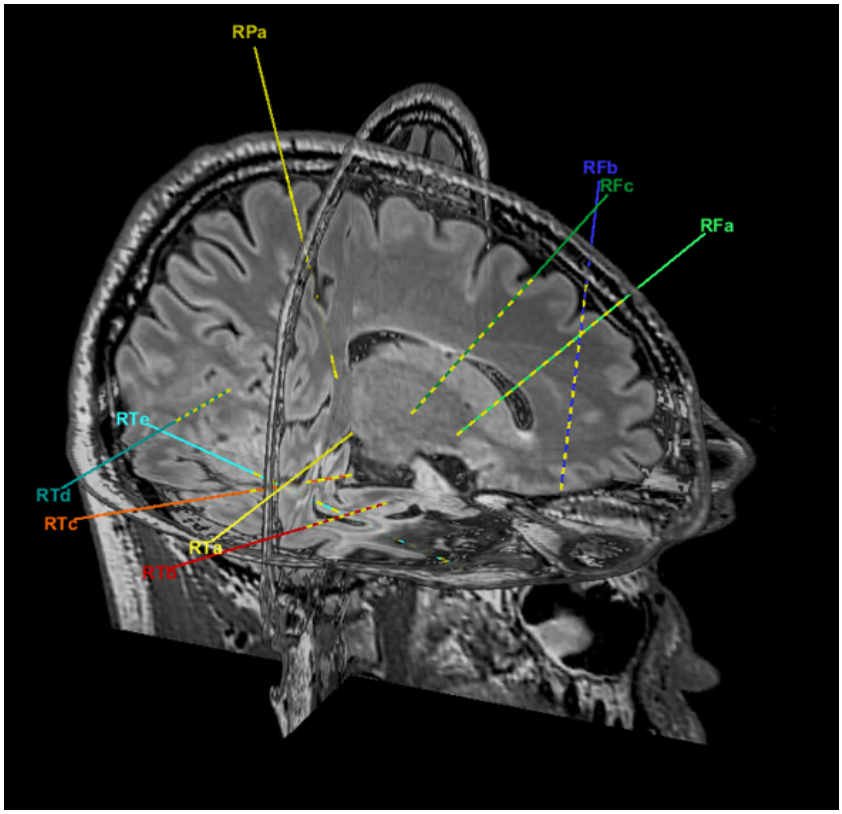


Figure 2.2.: MRI scan and electrode placement of patient 5. The colored lines represent the electrodes, and the short yellow lines show where the contacts are. All electrodes are implanted in the right side of the brain. Credit: figure courtesy of the University Hospital Erlangen

required labelled data. In order to gain some labelled dataset for training and testing, existing methods from previous studies (described later) were used. Details of these methods are described in later sections.

Single unit spikes

One crucial dataset is the one containing spikes and their classification into what neurons they originate from: pyramidal cells or interneurons. Using a clustering algorithm (described later), SUA could be categorized into these two classes, hence, a *pseudo ground truth*[4] could be determined that serve as the basis for the machine learning models.

The MRTC1 channel of patient 2 showed a reasonable difference in the characteristics

[4]Due to the limitations of extracellular recordings, an electrode may not capture the *real* shape of an action potential. This may result in spikes that are distorted. Therefore, determining the ground truth for them was not possible. Instead, we relied on existing techniques to come up with a pseudo ground truth. While 100% correctness cannot be guaranteed, the results give a feasible base.

Patient number	Temporal	Extratemporal	Hippocampal
1	LTA{1,6}, LTB{5,6}, LTE7	n/a	n/a
2	RTA{1,2,4,5,6}, RTB{1,6,7}, RTE{2,3}	n/a	RTC1
3	RTD1	n/a	RTA{1,2}, RTC1
4	RTA{1,2,3}, RTB7, RTC7, RTD5, RTF{2,3}	n/a	n/a
5	RTB7, RTC{3,5}, RTE{2,3,8}	n/a	RTB1, RTC1
6	n/a	RPC7, RPD4	RTA2
7	RTA6	RFA1, RFB2, RFC1, RFF4	n/a
8	n/a	n/a	LTA{1,2}, LTB{1,2}
9	n/a	LOA3, RPA{7,8}	LTB8
10	LTA{6,7}, LTH{6,7,8}	n/a	LTC1, LTE1

Table 2.2.: Electrode placement of contacts. Since these macro channels exhibited IED activity, they were selected for analysis.

Patient number	Microwire channels
1	MLTB{3,8}
2	MRTC{1,2,3,4,5,6,8}
3	MRTC{2,3,4,5,6,7}
4	MRTC{1,2,3,4,5,6,7,8}
5	MRTB{7,8}
6	MRTA7
7	MRTA{1,2,3,6,7,8}, MRTB{1,3,4,5}
8	MLTA{1,5,6}
9	MA{1,2,4}, MB2
10	MLTC{4,5,6,8}, MLTE{2,3,5,6,8}

Table 2.3.: Microwire channels used in this study. Similarly to macro channels, since these channels showed SUA, they were selected.

(shapes) of putative interneuronal and pyramidal spikes, thus, after careful visual examination, this dataset was selected to train the classifiers described later. The resulting dataset turned out to be imbalanced with a ratio of 1:3 for the two classes[5]. To mitigate this, downsampling of the majority class was performed with random selection. This resulted in a dataset of a total of 161566 samples of 2ms long (65 features) spikes with a 50-50% ratio between the two classes.

[5]Detected pyramidal spikes: 80783, interneuronal spikes: 256511

IEDs

The IED training dataset was chosen in a similar fashion, empirically. The LTA1 channel of patient 8 appeared to provide suitable samples for training. The samples were identified using a third-party software, BESA® (see later) fed with IED samples/templates hand-picked by a certified epileptologist. This way 640 samples of 300ms (2458 features) were identified on the channel. The dataset contains only positive IED samples, and samples from other hypothetical classes (e.g., baseline, artifacts) are not included.

2.2. Tools, libraries and frameworks used

To carry out the analyses, Python was used to create the software part of this thesis. The major libraries used are: NumPy (for vector and matrix storage and manipulation), Pandas (for in-memory storage), scikit-learn (for classical machine learning algorithms), SciPy (for signal processing), Matplotlib and Seaborn (for visualization). Initially, TensorFlow[6] with Keras[7] was used for the implementation of and as a proof of concept for artificial neural networks, which was later re-implemented in PyTorch[8].

2.3. Overview of the analytic processes

The process of the analysis can be thought of as a pipeline of jobs. Jobs represent building blocks performing an atomic operation such as filtering, spike detection, classification or clustering, and statistical analysis. The pipelines to detect and analyse SUA and IEDs are fundamentally similar, however, they differ to some extent.

Pipeline for SUA

The first stage in this pipeline is detection. The raw files are loaded, read, and then filtered to eliminate oscillations of undesired frequencies. In other words, the baseline is *flattened*. Using a threshold, spikes that cross this line are the candidates. As a result of the filter, a spike may cross both positive and negative thresholds. This way some spikes are detected redundantly and incorrectly. As a remedy, data cleaning is performed, and falsely detected spikes are eliminated. This set of procedures produces an output (a CSV file) which serves as the input to the next job (stage).

The next step is the analysis, and it is composed of two parts: sorting and clustering. Sorting aims to eliminate candidates that are in fact artifacts, but resemble spikes, while clustering assigns a label to them depending on whether they originate from a pyramidal cell or interneuron. These methods are described in detail in later sections.

The fact that there are multiple separate channels (recordings) for each patient results in multiple combinations as to how they can be analysed and aggregated. Merging the

[6] https://www.tensorflow.org/
[7] https://keras.io/
[8] https://pytorch.org/

channels at one point is important in order to gain a comprehensive overview of all electrical activities within the sampled volume of the hippocampal formation and to not lose valuable information that could lead to erroneous deductions. While there is no proven way to solve this problem, and the relevant studies reviewed [18] [17] do not explicitly mention how they overcame this, a flexible and rational solution is provided here.

The pipeline structure of the software allows to try multiple combinations of *detecting* SUA on each separate channel - *sorting* spikes independently from other spikes on other channels - *clustering* these spikes into one of the two types - *aggregating* data from all channels to get an overall view on the electrical activity at given time intervals. At first, the order was the following: detect, sort and cluster, then aggregate (merge) the spikes extracted from the channels. This resulted in a incomprehensible outcome, as spikes on different channels but with the same label were not clustered correspondingly as visually justified by an electrophysiologist. One set of spikes on one channel that were put into one cluster did not correspond to spikes on another channel in terms of their shape, albeit having the same label assigned[9]. While a different learning algorithm and access to the ground truth would be the desired solution, changing the order of jobs also led to feasible results. Therefore, the chosen order is the following: detect, aggregate, sort, and cluster.

It may happen that a channel exhibits spikes with both negative and positive polarity. For simplicity, in this thesis spikes with one polarity are considered, and it is chosen based on whose population is greater.

Pipeline for IED

IEDs are more cumbersome to detect due to their heterogeneous shapes and appearances. However, the principle of the pipeline is identical to the one used for SUA. That is, the first phase is detection, and the second one is clustering.

From the perspective of this thesis, one goal is to provide a novel method to identify IEDs. The results are very promising, albeit they do not provide fully reliable outcomes that can be used to deduct medical conclusions. The improvement and fine-tuning of these methods require further investigation which is beyond the scope of this thesis. Therefore, in the final statistical analysis IED sample annotations provided by physicians are used. The aim of exploring these state-of-the-art methods is to showcase what possibilities of detection there are and to set the foundations for further research. Detection methods, their details and some of the issues arising are discussed in detail in later sections.

Aggregated analysis

In the final step, results from the above two pipelines are merged and compared. SUA is explored in the vicinity of (around) a relatively large number of IEDs to examine firing

[9] This is an evident example that the clustering algorithm picked is relative to the underlying data, and it serves as a motivation to find other methods to classify spikes. This issue is described in more detail in later sections.

patterns and make deductions. The results are presented both visually, in the form of raster plots (event plots), and quantitatively, in the form of statistical hypothesis testing.

The IED samples that take part in the analysis are cleaned in the following way. If an IED has no SUA around it, or SUA is below the average firing rate in a given set (sparse firing), then it is discarded. More formally, given IED samples $S_{IED} \subset \mathbb{R}^{N \times D}$ and spike samples $S_{spikes} \subset \mathbb{R}^{N \times D}$, the sample-(row-)wise spike counts are summed: $c = \sum_{i=1}^{N} s_{spikes}^{(i)} \in \mathbb{R}^N$. Afterwards, the per-sample sums are averaged: $\mu = mean(c) \in \mathbb{R}$. If $c^{(i)} < \mu$ for the corresponding $s_{IED}^{(i)}$, then it is discarded.

2.4. Sorting and clustering of spikes

2.4.1. Detection and sorting

There are a number of existing methods capable of extracting and sorting spikes fired by neurons.

[19] propose a widely used method, and off-the-shelf tools such as Combinato[10] and Wave_clus[11], which are implementations of these techniques. The steps are defined as follows.

1. Given a raw recording, apply a band-pass filter (300-6000 Hz) in order to annihilate undesired frequencies (e.g., eliminate local field potentials). Single unit spikes fire in higher frequencies. (Sometimes only a high-pass filter is applied, e.g., 600 Hz).

2. Calculate the approximation of the standard deviation of the filtered signal: $\sigma = median\left\{\frac{|x|}{0.6745}\right\}$ and apply a voltage threshold of 4σ or 5σ, both above and below the mean to detect spikes of different polarity. The reason the standard deviation is replaced with the median is because it could lead to very high threshold values which may fail to detect spikes of lower amplitudes.

3. Once candidate spikes are detected, the most descriptive features of the data points are extracted. As one spike may contain 60-80 data points depending on the sampling frequency and the length of the snippet (usually 2ms), dimensionality reduction is necessary. The authors [19] propose two methods: principal component analysis (PCA) and wavelet transformation. Generally, the first three principal components are enough to sufficiently *describe* a spike. A problem with PCA is that it selects the directions of maximum variance from the data, which may not be the directions of the best separation. Wavelets overcome this issue.

4. The final step is clustering the spikes. The proposed method is superparamagnetic clustering (SPC) [21]. It is based on nearest neighbours interactions which group together a contiguous set of points given that the local density is larger than a certain value [22].

[10] https://github.com/jniediek/combinato
[11] https://github.com/csn-le/wave_clus

Rossant et al. [23] introduce a novel method for spike sorting for large and dense electrode arrays where a high number of channels are present. One advantage is that it can be applied to multiple concurrent channels. A disadvantage is the need to know the probe geometry beforehand, which can be cumbersome. To detect and sort spikes, the authors

1. filter the signal, then use a flood-fill algorithm to detect spikes across channels.
2. Afterwards, two vectors are created: a feature vector filled with values after running PCA (with the first three principal components), and a mask vector which is computed from the peak spike amplitude on each detected channel.
3. To cluster the spikes, a modification of the traditional expectation-maximization (EM) algorithm, the masked EM algorithm is used. This fits the data as a mixture of Gaussians.
4. As their last step, they provide the possibility for interactive manual curation.

The high level steps described above are visualized in Figure 2.3.

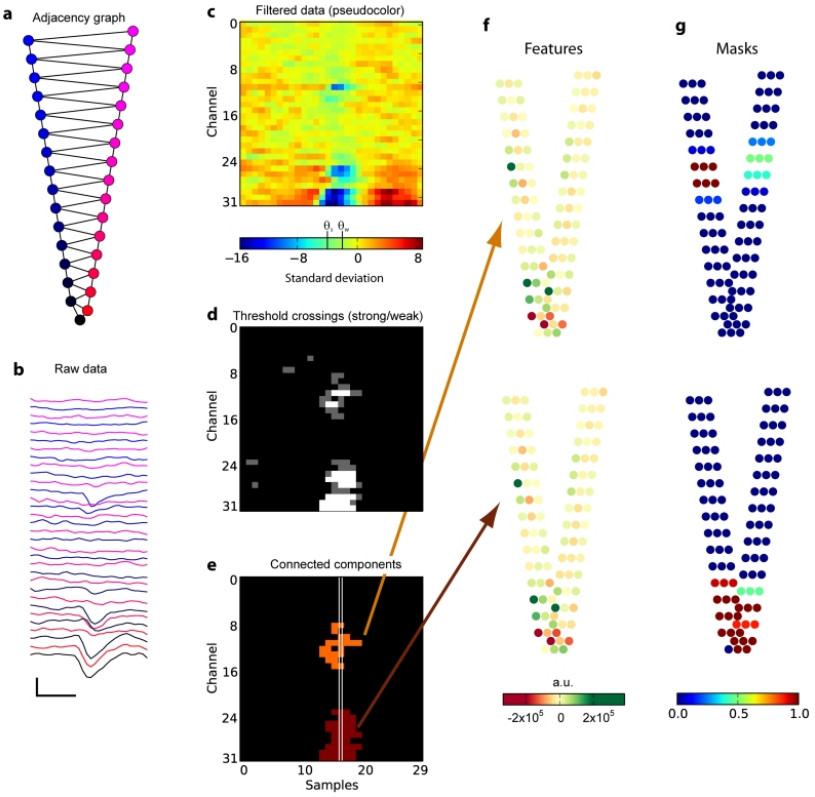


Figure 2.3.: Local spike-detection algorithm. (Credit: [23])

In this thesis, ideas from both methods were used, however, the detection algorithm resembles the work of [19] more with some modifications applied.

An important difference is that we can automatically detect spikes of both positive and negative polarity, i.e., the threshold crossing (standard deviation) is carried out below and above the mean of the signal. One byproduct of the bandpass filter is that spikes are slightly distorted and hence, they may cross the threshold on both sides as the threshold detector detects spikes with both positive and negative polarity. Therefore, an additional cleaning step is needed. That is, candidates with positive and negative polarity that are in close vicinity of each other (exceeding a temporal threshold τ) are checked. We empirically determined $\tau = 750$ microseconds to be an optimal value to distinguish bipolar threshold crossings of one potential from two potentials of opposite polarity in rapid succession. Given the event markings (timestamps), $t_{postitive}$ and $t_{negative}$ of the candidates with the two polarities, if $abs(t^{(i)}_{postitive} - t^{(j)}_{negative}) < \tau$, then the absolute peak values of the corresponding candidates i and j are compared, and the one with the higher value is kept and the other is discarded.

For simplicity, feature extraction is achieved using PCA with the first three principal components. SPC turned out to be too complex, outdated and cumbersome, therefore, it was not utilized, and instead, an expectation maximization (EM) algorithm (mixture of Gaussians) was used. To determine the number of clusters, the Bayesian Information Criterion (BIC) was used such that the maximum allowable number of clusters was set to 10. Initially, DBSCAN [24] was used as it has the capability of clustering without determining the number of clusters beforehand, and it can implicitly distinguish outliers. In spite of these advantages, it is prone to errors when the density of the points in one cluster differs from another one (when inter-cluster distances are significantly different). Moreover, setting the algorithm's hyperparameters *eps* and *minPts* is challenging as it varies greatly with the underlying dataset. The authors of DBSCAN propose a heuristic to determine these parameters using the k-nearest neighbors (KNN) algorithm. However, it is not an automatic remedy, and it requires human interaction as well. Furthermore, this heuristic did not turn out to be an effective way of defining the number of clusters with our dataset. One disadvantage of the EM algorithm is that it cannot detect outliers by itself.

Once the automatic sorting is finished, the human operator of the software is asked to select the clusters that should be used in the next stages. The unselected clusters are discarded. The operator is presented with a plot of sample spikes summing up to 1000 samples drawn uniformly at random from the detected population. The *fineness* of the clusters, that is, whether a cluster consists of actual spikes from neurons cannot be automatically decided, hence, the requirement of visual inspection. Making such a decision is subjective and needs domain knowledge.

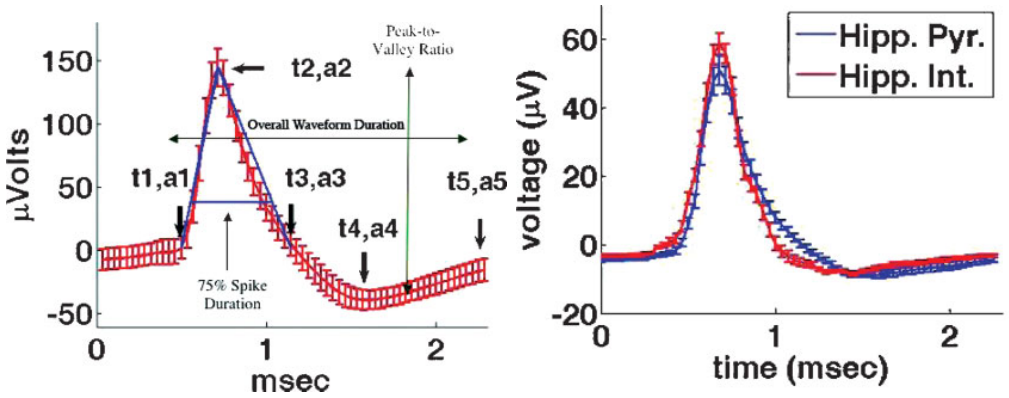
2.4.2. Clustering putative pyramidal cells and interneurons

Previous studies [25], [26], [27] have presented methods that morphologically characterize individual spikes, and using the gained feature, they differentiate between the two types

of neurons. Defining the characteristics of spikes in general, which could provide a reliable identification of distinct neurons (their type), is a hard task as the properties of recorded spikes is highly dependent on several factors such as the physical distance between the relative localizations of the neuron and the recording probes. Nevertheless, bearing in mind that we cannot have 100% confidence about correct identification, we examined and utilized the methods described below.

Viskontas et al. [26] characterize the electrophysiological features of spikes. They use three physiological criteria to distinguish between the two types of neurons: *firing rate*, *burst propensity*, and *waveform*. Figure 2.4a shows possible measurements of a spike, and Figure 2.4b depicts median differences between pyramidal cells and interneurons. Note how similar the two types are in the figure.

A comparison of pyramidal cells and interneurons was conducted by [25]. Figure 2.5 shows the different shapes of superimposed spikes elicited by these two types of neurons.



(a) (b)

Figure 2.4.: a) Measures of action potential duration and amplitude. b) Median waveforms. It shows the difference in the median waveform shape for interneurons and pyramidal cells in the hippocampus. (Credit: [26])

[27] use somewhat different features to characterise spikes. These are *peak amplitude asymmetry*, *half-width duration* and *trough to peak*. Figure 2.6 visually explains the definitions of these three features. To cluster the data points, the k-means algorithm is used, where $k = 2$ as two types of neurons are assumed.

The authors manage to group the two types of neurons into clusters that characterise the spikes well. This is shown in Figure 2.7.

In this thesis, the latter method [27] was followed, and in accordance with Figure 2.8, the features were defined as the following. The *peak amplitude asymmetry* is $(b-a)/(b+a)$, where $a = |0 - a_1|$ and $b = |0 - a_5|$. a_i is the corresponding amplitude at t_i. The *halfwidth duration* is the value of Δt at the spike's half width, i.e., $t_4 - t_2$. Lastly, the *trough to peak* is defined as $t_3 - t_1$. Note that this only applies when the spike has a positive polarity. When it is negative, this feature is intuitively calculated as $t_5 - t_3$. To differentiate putative neurons given these features, the k-means algorithm is applied, where $k = 2$.

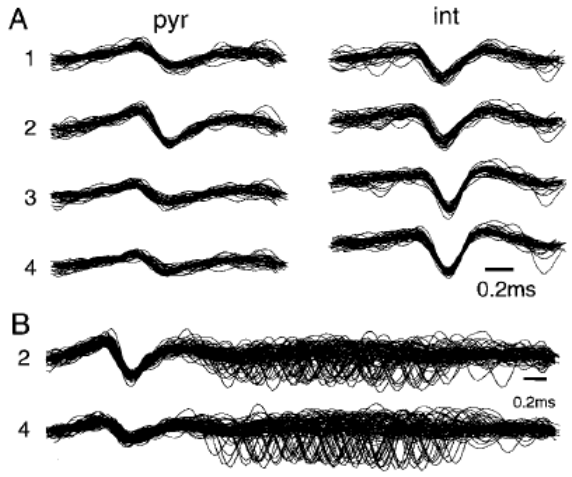


Figure 2.5.: Superimposed waveforms of a pyramidal cell and an interneuron. (Credit: [25])

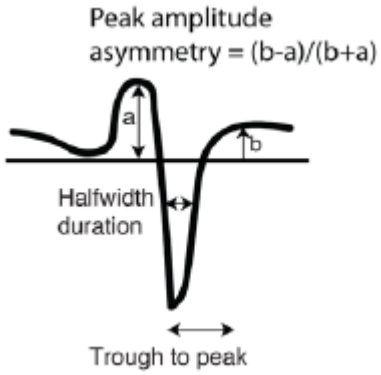


Figure 2.6.: The three features extracted from spikes. (Credit: [27])

This method is capable of separating the spikes based on the above features, however, it cannot determine with any certainty which clusters spikes fall into. Therefore, the final say is the human operator's responsibility who has to assign the labels.

The above method can be considered as a *naïve* solution. On the one hand, since the publication of these papers state-of-the-art machine learning algorithms have gained more popularity which these studies do not consider. On the other hand, the application of a clustering algorithm is *relative* to the underlying dataset. That is, if spikes in one dataset (e.g., one channel of one patient) are more alike than in another, then it is possible that some spikes are assigned an incorrect label depending on the variability of the dataset.

Let $X^{(1)}, X^{(2)} \subset \mathbb{R}^{N \times D}$ be two datasets from two sources (i.e., patients), and p_i be the distribution of the ith population where $i \in \{1, 2\}$ represents the two different types

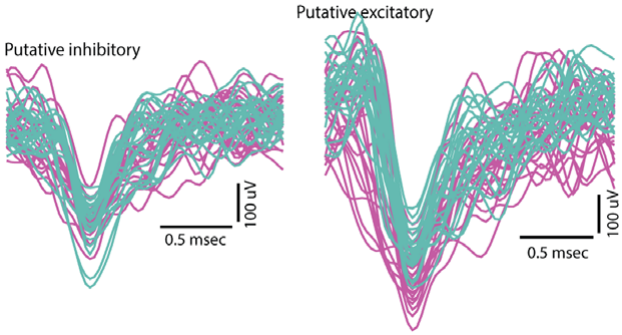


Figure 2.7.: Putative inhibitory and excitatory spikes. (Credit: [27])

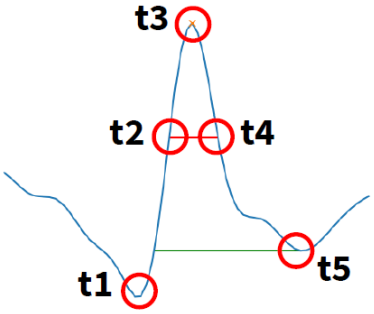


Figure 2.8.: The three features extracted from spikes.

of spikes from pyramidal cells and interneurons. Furthermore, assume that all samples are real, and no noise is present. Applying a clustering algorithm c on both datasets independently, we have $X_1^{(1)}, X_2^{(1)} \subset X^{(1)}$ and $X_1^{(2)}, X_2^{(2)} \subset X^{(2)}$ with $X_y^{(j)}$ being a clustered set with label y for the jth dataset. It may happen that $X_1^{(j)}, X_2^{(j)} \sim p_i(X)$, that is, the spikes in the two clustered datasets are equivalent or have a very similar shape. Two problems arise from this:

1. using the k-means algorithm, k clusters are created irrespective of the underlying distribution. For example, if spikes in one channel or patient only represent one certain type of neuronal activity, or their shapes are distorted such that they look alike, it would be more feasible to have only one cluster instead of $k = 2$;

2. in a different setup when $X'^{(1)}$ and $X'^{(2)}$ each have samples from both $X^{(1)}$ and $X^{(2)}$, it may happen that some samples $\hat{X} \subset X^{(j)}$ have a different label assigned by c only because the setup is different. That is, the output of c is dependent on the features of spikes in the dataset that is at hand. This means that a set of spikes in one setting could be identified as originating from pyramidal cells, which would otherwise be spikes of interneurons. Consequently, c itself, cannot

objectively provide a ground truth with respect to the true characteristics of the spikes representing the different types of neurons.

Despite these problems, the strength of clustering lies in the fact that it can learn in an unsupervised manner, and hence, no labelled data (ground truth) is needed. Nor is it required to define the number of clusters with several clustering algorithms. However, in this context having a clear distinction between the different neurons is important from a medical perspective. Furthermore, the above mentioned methods do not provide a concrete threshold for distinctions if the inter-cluster difference is low. (This was the case with the patient data used in this thesis.)

These problems motivate us to search for alternatives to clustering. Fortunately, there are several supervised methods that could prove to be useful. Thereby the clustering problem is translated into a binary classification problem. Supervised methods, however, require labelled data for training, and it was not provided due to how tedious the task of labelling would be. Therefore, after applying the *naïve* methods [27] to the data, the clustering provides a distinction between putative pyramidal cells and interneurons. The *best* dataset was handpicked, and used as training dataset for the supervised methods described below. This was done with visual inspection confirmed by a physician. For the sake of brevity, we assume that the clustering provides the ground truth of the spikes for the datasets across all patients. This assumption was used for the benchmark to measure the performance of the models.

Logistic regression

Logistic regression is one of the simplest classification algorithms. Despite its simplicity, it can be very powerful. The hypothesis $h(x, w) = h(Z) = \hat{y}$ (prediction of the input) is achieved as follows. Let $Z = W^T X + b$ be an intermediate step with $W \subset \mathbb{R}^N$ being the weights (parameters) and $b \in \mathbb{R}^N$ the bias of the model. Feeding Z into the sigmoid function $h(Z) = \frac{1}{1+e^{-Z}}$, the probability of the samples belonging to one of the classes is returned. The cost (or loss) for one sample is defined as

$$\ell(y, \hat{y}) = \begin{cases} -log(\hat{y}) & \text{if } y = 1 \\ -log(1 - \hat{y}) & \text{if } y = 0, \end{cases}$$

and for the entire dataset it is $L(y, \hat{y}) = \frac{1}{m} \sum_{i=1}^{m} \ell(y^{(i)}, \hat{y}^{(i)})$.

For the optimization (or training, i.e., finding the values of W), gradient descent (or some variant of it) is used to find the minimum of the loss function.

Support vector machine

Support vector machine (SVM) [28] is another simple yet popular supervised learning algorithm for classification problems. The idea behind SVM is to maximize the margin (distance) between the training samples and the decision boundary. The support vectors are the data points that are closest to the decision boundary. Let $\ell(y, \hat{y}) = max(0, 1 - y\hat{y})$ be the hinge loss. The objective is to minimize the parameters w of the model:

$\min_w ||w||_2 + \frac{1}{m} \sum_{1=1}^{m} \ell(y, \hat{y})$. With the hypothesis $h(x, w) = \hat{y}$, the label of a new sample is predicted. Figure 2.9 illustrates how SVM functions.

There is a soft-margin variant [28] of SVM, and the algorithm can be improved with the kernel trick if data is not linearly separable. These methods are beyond the scope of this thesis and are not discussed here.

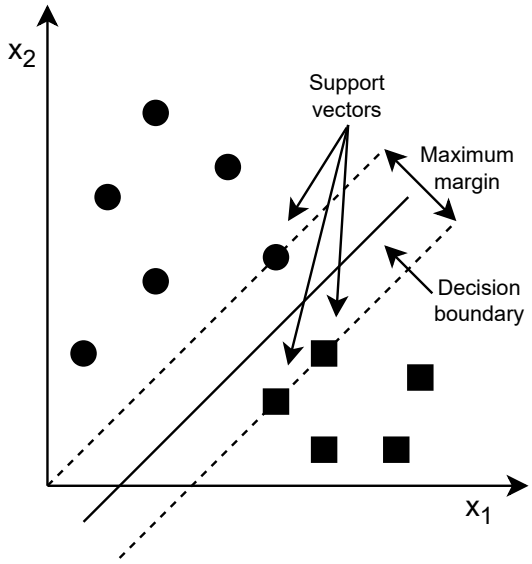


Figure 2.9.: Overview of vanilla SVM for binary classification in the two dimensional feature space. The objective is to maximize the margin between samples lying on the border of the two classes. The decision boundary is then the line (hyperplane) along this margin.

Artificial neural network

An artificial neural network (ANN) is a learning algorithm mimicking biological neurons and connectivity of the brain. A single unit is called the *perceptron*, and it is a function f that maps its input x to some output y. f performs a linear transformation on x using the associated parameters (or weights) w and bias b. The intermediate transformation $z = Wx$ is then fed into an activation function φ that outputs a. a either serves as the input of other neurons in the network, or it is the output y_{pred} of the network that provides the answer to a problem such as classification (e.g., it is a class label).

A neural network consists of multiple layers L: the input layer, the hidden layer, and the output layer. Each layer l may contain multiple units. Depending on the problem setting, the hidden layer may contain multiple layers of perceptrons resulting in a *deep* network. We refer to learning in deep neural networks as deep learning. In other cases, when the hidden layer contains only a few layers, the architecture is referred to as a *shallow* network.

The simplest form of ANNs are feedforward neural networks (FNNs) or multi-layer perceptrons (MLPs) as illustrated by Figure 2.10. In this architecture there are no cycles or loops. The lth layer simply passes the data to the $(l+1)$th layer. This is called forward propagation, and when not in training mode, it is inference, that is, the model provides an answer to a well defined problem such as classification. During training (or learning), the weights of the neurons are updated based on the resulting loss. This is called backpropagation. Backpropagation calculates the gradients of the loss with respect to the weights of the network.

One significant benefit of ANNs is that there is no need for feature engineering. Instead, the model learns the the underlying representation of the data layer by layer. The advantage is that these do not necessarily have to be human interpretable, and hidden features may also better describe the data.

There exist numerous other types of networks for different problem settings. Some of the most common ones include convolutional neural networks (CNNs) that are widely used in computer vision; recurrent neural networks (RNNs) that are utilized in text processing, natural language processing (NLP); long short-term memory (LSTM) networks that are a variant of RNNs and are popular in speech recognition and video analysis.

The model used in this thesis has four layers as illustrated by Figure 2.11. The input has a dimension of 65 which is the equivalent of a 2ms long sample in the current setting. All layers except the output layer have ReLU (Rectified Linear Unit) as their activation function. It is defined as $\max(0, z)$. It is a state-of-the-art choice, and it has the benefit of avoiding vanishing gradients which the sigmoid function is more prone to [29]. Binary cross entropy (BCE) [30] and RMSprop (Root Mean Squared Propagation) [31] [32] [33] were used as the loss function and optimizer, respectively. The parameters for the optimizer are as follows: $\eta = 0.001$ and $\epsilon = 10^{-7}$. 30 epochs were chosen with a batch size of 32. The number of layers and nodes in them along with the other hyperparameters were chosen empirically.

2.5. Detection of intericatal epileptiform discharges

The detection of IEDs can be a cumbersome task. As they can differ both in shape and duration even within one patient, it is difficult to provide a reliable and robust detection method. In certain cases, even certified epileptologists may find it hard to visually differentiate IEDs from other artifacts. This makes automation more challenging. Since manual extraction is very laborious, the need for automatization is obvious, and therefore, several methods have been proposed. [20] provide a comprehensive and recent review on existing approaches. They conclude that methods that utilize classical machine learning and deep learning algorithms show the most promising results, however, their use in clinical settings is not prevalent.

In the sections below two simple methods are described which are used more frequently. Furthermore, we propose two techniques utilizing state-of-the-art neural network architectures with promising results.

Classical supervised machine learning algorithms (e.g., logistic regression, SVM) are

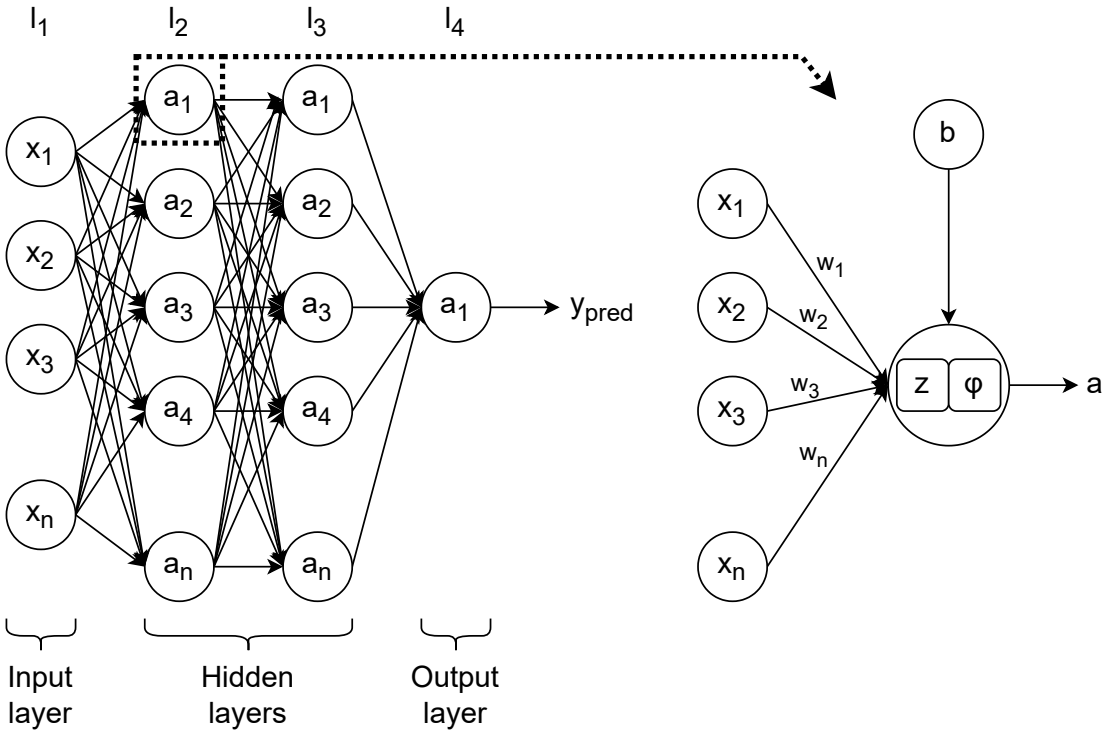


Figure 2.10.: Schematic overview of a fully connected feedforward neural network with two hidden layers. A neuron is fed with the data (either from the input or from other nodes), the corresponding weights and a bias term. The neuron calculates the dot product of the input and weights, and adds the bias. The activation function is applied at the end to introduce non-linearity.

not covered here, but they could also potentially achieve satisfying detection capabilities. One drawback is, however, the need for a pre-defined number of outcomes or classes (i.e., the different clusters/classes of IEDs). Additionally, experimenting with clustering algorithms could be intriguing as well.

2.5.1. Template matching with sliding window

A common way to detect IEDs is template matching methods [34] [35] [36], where a template IED is created from an annotated set of samples. This template (or templates) is slided across the signal, and if the snippet of the signal is similar to the template, it is a hit. The matching part can be done using different techniques such as mean squared error (MSE) [34], cross-correlation [35], or dynamic time warping (DTW) [36].

Generally, template matching works as the following. Given n manually marked spikes, we select via visual inspection m of these such that they have similar characteristics. Afterwards, an average of these spikes is calculated resulting in a template. A recording (from the same subject) is then partitioned into snippets, and the template is slided

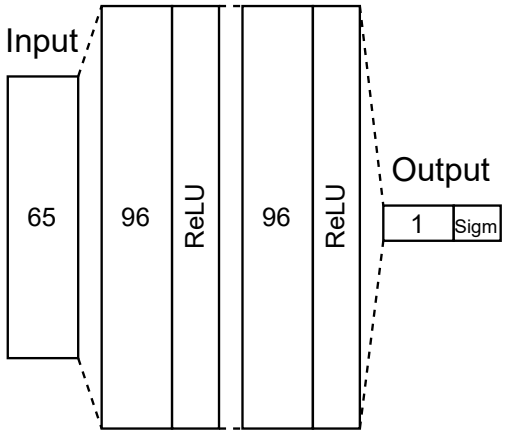


Figure 2.11.: Architecture of the feedforward neural network for the classification of spikes. The boxes represent the layers of the network, and the numbers in them indicate the number of neurons in that layer.

through the recording. In each snippet the sample is centered according to its maximum or minimum amplitude. As the last step, the correlation (or some similarity measure) between the template and the snippet is calculated. If this value reaches a certain threshold, the snippet is marked as an IED. [15] also utilize this method for detection. This approach is simple, but it has several drawbacks. Firstly, manual selection of the IEDs that add up to the template is inevitable. Furthermore, this has to be repeated multiple times to detect a wide range of the discharges. Secondly, multiple patients exhibit different types of IEDs, meaning that one set of templates from one patient cannot be used for another. Again, the whole procedure needs to be repeated for every patient.

BESA®[12] is a proprietary software that allows the semi-automatic detection of IEDs in multiple channels. In order to do this, a human expert with enough domain knowledge operating the software has to manually select a few samples that are used to create a template from. These templates are then compared with snippets of the EEG, and if a certain threshold is met, the snippet is marked as a positive sample. IEDs used in this thesis were extracted with this software.

2.5.2. Threshold method

[37] revisit the threshold detector for IEDs and propose a fast thresholding method to achieve detection. It consists of the following steps:

1. Preprocessing of the signal. Initially, apply a 60 Hz notch-filter to eliminate noise (frequency of power line noise in the US, in Europe 50 Hz should be used), then normalize the filtered signal to a range [-1, 1], where -1 represents the xth percentile, and 1 represents the $(100 - x\text{th})$ percentile. In order to eliminate the changing baseline, a moving average of one second is subtracted from the data.

[12] https://www.besa.de/

2. Split the normalized data into two halves. One half is the training dataset, and the other is the testing dataset. For the training half, an array is created which marks whether a spike occurs at a given time. Run the data through a threshold detector (with various thresholds, from -0.3 to 0.7 in increments of 0.02). Mark the spikes that cross the thresholds, then compare the detected spikes with the manually marked IEDs in order to determine the true positive and false positive rates for each threshold. Choose the optimal threshold (which minimizes the error of false positives and negatives), then run the test data through the detector with the chosen threshold.

3. Manually inspect the results of the detector.

With this method the authors claim to have achieved a true positive rate of 98.8% and a false positive rate of 2%. Although the presented method seems very promising, there are still some questions that remain open. A significant disadvantage is that for training, manual markings are still necessary for each patient as the algorithm splits the data into training and verification (test) sets. Consider a case when a patient is monitored for 24 hours. In such a scenario, it is very likely that the IEDs are unevenly distributed throughout the recording. How can one tell where to mark IEDs to have a sufficient amount of training data? What if this one-day-long recording captures events in both awake and sleep states, which may differ in terms of distribution and characteristics of the IEDs? How to efficiently handle cases when the sampling for training is randomly distributed throughout the entire recording (to get a good variability of IED types), and as a result, the data cannot be evenly split? What happens when an IED is an ensemble of multiple spikes (peaks), and this single IED is detected multiple times? Moreover, there are the cases when IEDs overlap, and what can be seen on the recording is a sequence of IEDs.

Unfortunately, the implemented method turned out to be insufficient for large scale analysis, and therefore, it was not utilized in the end.

2.5.3. Autoencoder

Given the homogeneous nature of IEDs across and also within patients, implementing an n-class or binary classifier is not a feasible option. As we are not aware of the number of possible classes beforehand, training such a model is not possible. A binary classifier could, however, work and even provide intriguing results, although in practice such an implementation can be tedious work. Namely, one would train a model with two classes: IEDs vs. everything else. While positive samples (IEDs) are available, collecting samples from the other class and covering all signal fluctuations and possibilities would be impractical.

To overcome these difficulties, where samples from only one class are available, we explore a promising artificial neural network architecture: autoencoders [38] [39]. These networks have the advantage that, for training, having samples of one class is sufficient.

An autoencoder is a type of neural network that has the same number of input nodes as the number of output perceptrons. The first half hidden layers compress the data (encode

it) as the number of nodes in these consecutive layers decrease. Once a pre-defined level of compression is achieved, the number of nodes towards the output layer start to increase again. This part is responsible for the re-construction of the data (decode it). In this case the model does not assign a label (y) to the input values (x), i.e., the model is not a classifier by itself ($f(x) \neq y$), but instead, it learns the representation of the data and is capable of creating an artificial object (x') that resembles the data the model was trained on ($f(x) = x'$). The power of autoencoders lies in this characteristic. If x is from the distribution of IEDs, then the reconstruction error (e) between x and the corresponding x' is low. If, however, x does not come from this distribution, then e is high. Exploiting this feature, an autoencoder can be used as a classifier in our context. That is, feeding signal segments into the model, if e is below a certain threshold, then it is positively identified as an IED.

Let $X_{train} \in \mathbb{R}^{N \times D}$ be the dataset used for training, $X_{test} \in \mathbb{R}^{N \times D}$ the test (or validation) dataset, f the autoencoder as a function, $X'_{train} = f(X_{train})$, $X'_{test} = f(X_{test})$ the reconstructed sample sets respectively, and ℓ the L1 (or Mean Absolute Error - MAE) loss, where N is the number of samples and D is the number of features. Then $t = g(X_{train}, X'_{train})$ calculates the threshold, where $g(X, X') = mean(\ell(X, X')) + std(\ell(X, X'))$. If $x_{test} \in_R X_{test}$, whose $g(x_{test}, x'_{test}) < t$, then x_{test} is a positively classified sample.

Figure 2.12 shows the structure of the autoencoder used as the IED detector. The size of the input and output was selected to be the number of features in a 300ms snippet, and the number of hidden layers and nodes in the layers were chosen arbitrarily. ReLU and sigmoid were used as activation functions for the same reasons as in the case of the pyramidal cell-interneuron classifier: in larger networks vanishing gradients could potentially have more adverse effect on performance. 20 epochs with a batch size of 64 were also empirically chosen along with the Adam optimizer [40] [41] and L1 (MAE) as the loss function. The parameters for Adam are as follows: $\eta = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 10^{-7}$. The LTA1 channel of patient 8 with 640 samples served as the training and validation dataset with a 65-35% split.

2.5.4. Generative adversarial network

A generative adversarial network (GAN) [42] is another type of neural network architecture. Their advantage is similar to the one of autoencoders. That is, having positive samples is sufficient. GANs are also generative models, however, their complexity is higher. One remarkable, and perhaps, mostly known example with this architecture is the possibility to create *deepfakes*. This type of model can be very much useful within the context of this thesis.

A GAN is composed of two neural networks: a generator G, and a discriminator D. Random noise ($X_{noise} \sim U(0, 1)$) is fed into G, whose aim is to generate samples $X_{generated}$ that resemble X_{real}. The goal of D is to differentiate between $X_{generated}$ and X_{real}. In essence, D and G compete with one another, and this type of problem is known as the *minimax game* in game theory [43]. The goal of the training of both networks is to find the Nash equilibrium [44]. This means that there is an optimal solution such that

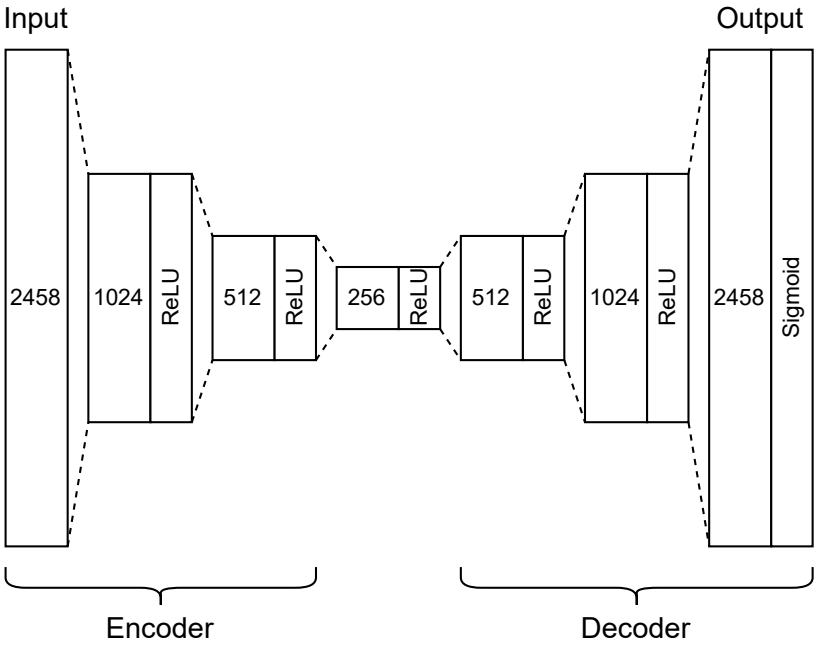


Figure 2.12.: Architecture of the autoencoder on which IEDs were trained on. In training mode, IED samples serve as input. The encoder compresses them and the decoder reconstructs them. When samples that do not resemble IEDs (e.g., random noise) are used as input, the trained model outputs an artificially generated IED. This way the reconstruction error is high, and the input snippet is not regarded as IED.

G can generate fake samples which very much resemble real IEDs, posing a challenge even to humans to differentiate between the two set of samples. If the goal is merely the generation of fake samples, then the 'confusion' of D after n training iterations is achieved. In other words, the accuracy of G gets better at the expense of D. In our scenario the goal is to differentiate IEDs from noise, baseline and other artifacts that are present on the EEG channels. That is, a sufficiently trained D achieves this task, and we are not interested in artificially generating IEDs[13]. The schematic overview of a GAN is shown in Figure 2.13.

Despite the power and promising results of GANs, their training remains a cumbersome and difficult tasks. As other studies [45] have pointed out, one of the difficulties lies in having to train two separate networks simultaneously as they compete with each other. Finding the Nash equilibrium is a difficult problem, and feasible algorithms to apply to GANs where the cost functions are non-convex and the parameter space is

[13]It is worth noting that generated IEDs exhibit visually appealing results, and even though we do not utilize this 'byproduct' of GANs, generated samples could be a base to make training sets larger, which can then later be used to train other models.

high-dimensional is currently not known [45]. Therefore, it may happen that the GAN fails to converge resulting in an unstable state.

Another issue that arises is called *mode collapse* [43]. If a GAN was trained on data that has multiple class labels (e.g., the MNIST dataset), and G fails to generate samples corresponding to all output labels, then we are faced with this problem. With regard to IED detection, G is not used, this issue was disregarded. However, it is worth looking into what effect this phenomenon has on D.

The idea in our context is to use the discriminator to tell the difference between IEDs and noise, baseline and artifacts. The generator's ability to create artificial IEDs is a byproduct in this case, and although it is not utilized, it has the potential to enrich scarce datasets. Figures 2.14 and 2.15 show the two neural networks of the GAN used here, the generator and the disriminator, respectively. The number of neurons and hidden layers were empirically chosen. All hidden layers have ReLU as their activation function, and sigmoid at both outputs guarantees that the results are normalized [0,1]. Both neural networks have the same type of loss function, binary cross entropy (BCE), and the Adam optimizer with the same parameters: $\eta = 0.0001$, $\beta_1 = 0.5$, $\beta_2 = 0.999$ and $\epsilon = 10^{-7}$. The model was trained in 100 epochs with a batch size of 16. To train and validate this model the same dataset as in the case of the autoencoder was used.

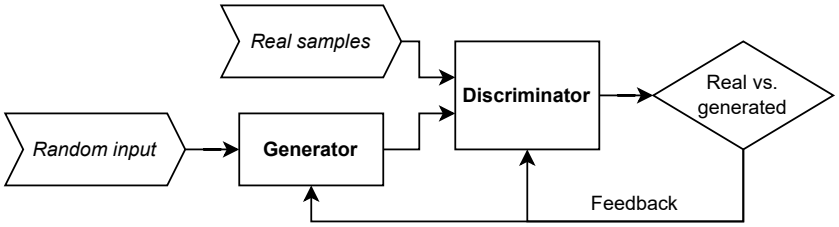


Figure 2.13.: Schematic architecture of a GAN. Random noise is fed into the generator whose output serves as the input of the discriminator. The latter also receives real training samples, and its task is to differentiate between real and generated samples. Parameters of both neural networks are updated based on how well the discriminator performs. After training, if the generator becomes good at creating artificial samples, the discriminator will be unable to tell the difference between real and fake inputs. At the same time, the discriminator should be able to differentiate between real samples and fake ones up to a certain degree. In our context, to tell the difference between IED and non-IED snippets, a good discriminator is sufficient and favorable.

2.6. Visualization and quantitative analysis of IED-SUA correlation

To compare the samples and to examine how neurons fire in the vicinity of IEDs, visual and quantitative measures were used. Visual representation of the connection of IEDs

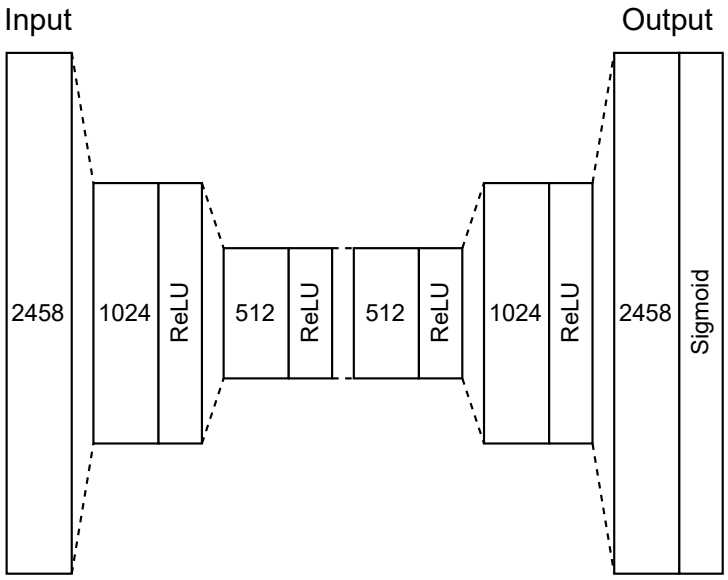


Figure 2.14.: The generator of the GAN aims at creating artificial IEDs.

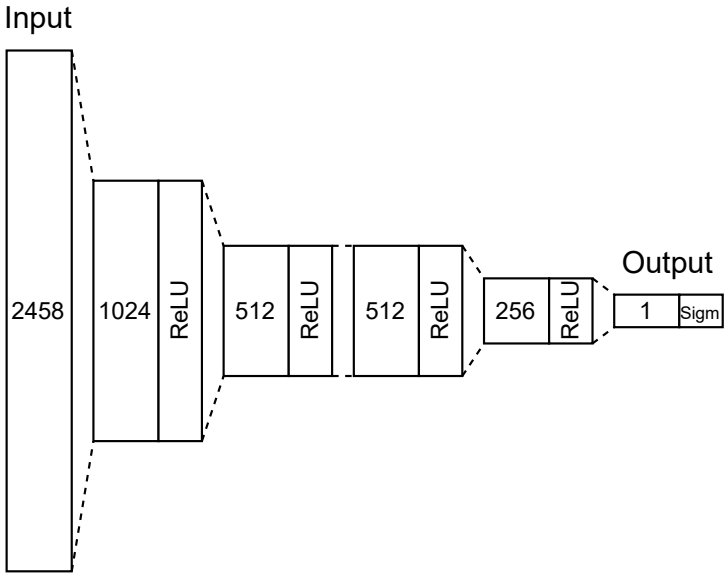


Figure 2.15.: The discriminator of the GAN differentiates between input samples whether they are real or not. Hence, it is essentially a binary classifier.

and SUA is achieved using event plots (raster plots) in which SUA around IED samples are stacked over each other. This results in a 300ms window divided between a pre- and post-IED window. These two windows are separated by the event marking of the IED.

This is conventionally defined by the maximum or minimum amplitude of a discharge. The stacking of samples allows to discover patterns that occur around an IED or a randomly selected non-IED sample.

Some forms of analyses may require samples where there is no abnormal activity in the recording. That is, 'silent' event markings are needed that do not contain IEDs. Extraction of such samples uses a greedy method to find relevant intervals using IED markings (timestamps). Algorithm 1 describes how non-IED samples are searched and extracted.

For quantitative measurements, permutation tests as statistical hypothesis analysis were carried out. There are several settings that could serve as the basis for the test depending on what is most relevant and what knowledge needs to be extracted. Given the pre- and post-windows of IEDs, multiple clusters of IEDs, the two types of neurons, the different brain regions, and the way how patients can be organised, the combination of possible test cases increases rapidly. Due to the many possibilities which may or may not be useful from a medical perspective, only one test case is presented here to serve as a proof of concept.

Let H_0 be the null hypothesis stating that pyramidal cells and interneurons fire around IEDs the same way for the different clusters of IEDs. More formally, let $z_j^{(i)} = count(spikes)$ be the number of neurons of type $j \in \{pyr, int\}$ firing in a given pre- and post-window of length t for the ith IED sample in cluster $c \in \mathcal{C}$ with $\mathcal{C}$ containing all IED clusters for patient m. Then, if $p(z_{pyr}^{(i)} = z_{int}^{(i)}) \forall i < 0.05$, H_0 is rejected.

Algorithm 1 Find non-IED samples

function InitializeIEDIntervals(T_{IED})
  $I_{IED} \leftarrow [\,]$
  for t in T_{IED} **do**
    $t_{start} \leftarrow t - \tau_{pre}$ ▷ τ represents the pre and post lengths in terms of time
    $t_{end} \leftarrow t + \tau_{post}$
    Append (t_{start}, t_{end}) to I_{IED}
  end for
  return I_{IED}
end function

function FindNonIEDIntervals(I_{IED})
  $I_{non-IED} \leftarrow [\,]$
  for $i \leftarrow 1$ to length of $I_{IED} - 1$ **do**
    $\Delta t \leftarrow t_{start}^{(i+1)} - t_{end}^{(i)}$
    if $\Delta t < 0$ **then** ▷ IED intervals overlap
      continue
    else
      $t_{start} \leftarrow t_i - \tau_{pre}$
      $t_{end} \leftarrow t_i + \tau_{post}$
      Append (t_{start}, t_{end}) to $I_{non-IED}$
    end if
  end for
  return $I_{non-IED}$
end function

function FindNonIEDTimestamps($I_{non-IED}$)
  $T_{non-IED} \leftarrow [\,]$
  for (t_{start}, t_{end}) in $I_{non-IED}$ **do**
    $N \leftarrow \lfloor (t_{end} - t_{start})/(\tau_1 + \tau_2) \rfloor$ ▷ No. of max events in an interval
    $k \leftarrow t_{start}$
    for $i \leftarrow 1$ to N **do**
      $k \leftarrow k + \tau_{pre}$
      Append k to $T_{non-IED}$
      $k \leftarrow k + \tau_{post}$
    end for
  end for
  return $T_{non-IED}$
end function

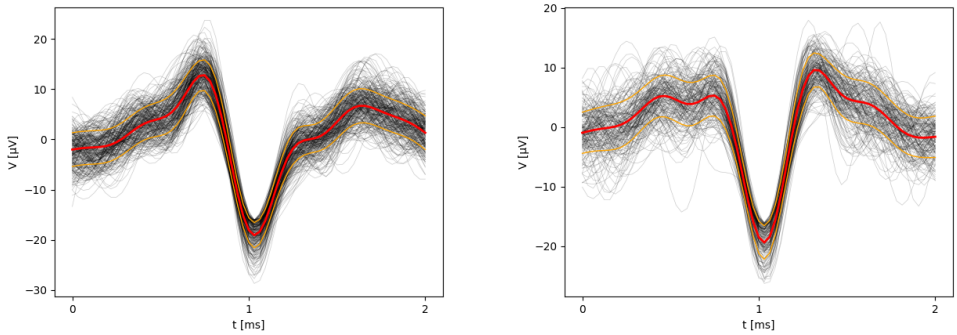
3. Results

3.1. Spike detection and clustering

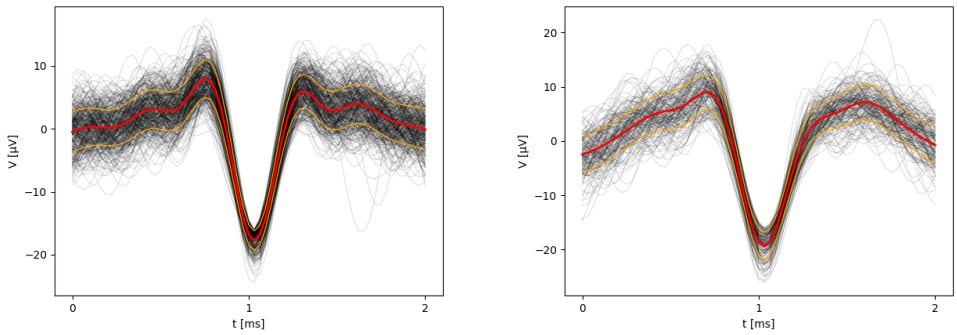
The threshold detector applied to the channels detected potential spikes on both polarities of the signal. These candidate spikes were aggregated and then cleaned to exclude duplicates. For the sorting the sets of spikes whose cardinality was higher were selected as they were more likely to contain a larger amount of proper SUA. The semi-automatic sorting method placed the spikes into clusters based on their shapes. Clusters with potential artifacts were discarded. Three features of the remaining spikes were used to determine the cell type. Table 3.1 summarizes the results for all patients from this pipeline. The number of detected spikes and clusters varied significantly on the given patient. There could potentially be confounders as to why these numbers differ including recording length, recording circumstances and medical history. Due to the lack of access to several of these data, conclusions cannot be made using merely the numbers in the table. Figures 3.1, 3.2, 3.3 show the visualization of the results for patient 5. These plots provided aid for manual inspection and classification.

Patient number	No. of candidate spikes detected	Spike polarity	No. of clusters detected/selected	No. of putative pyr. cells/interneurons
1	166059	Negative	6/4	48935/101072
2	3184873	Negative	7/5	1821634/1233754
3	748699	Negative	7/5	469370/221086
4	56256	Negative	7/5	7559/39368
5	63756	Negative	6/4	16773/44152
6	386761	Negative	7/5	226964/119924
7	2403397	Positive	7/5	1132199/1059989
8	19740	Positive	7/3	5622/9339
9	90619	Positive	7/3	23516/59564
10	526259	Negative	7/1	119718/247137

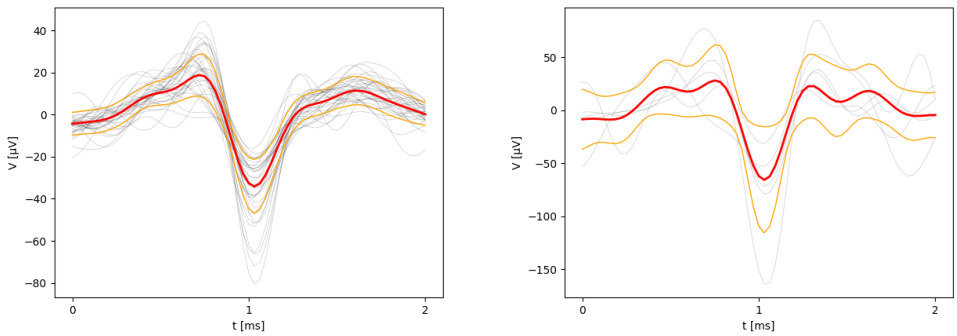
Table 3.1.: Detection, sorting and clustering statistics for all patients. Candidate spikes are not cleaned and contain outliers.



(a) Cluster with label 1. Total spikes: 17290. (b) Cluster with label 2. Total spikes: 10190.



(c) Cluster with label 4. Total spikes: 22482. (d) Cluster with label 5. Total spikes: 10963.



(e) Cluster with label 0. Total spikes: 2504. (f) Cluster with label 3. Total spikes: 327.

Figure 3.1.: Spikes of patient 5 clustered together after sorting. a) b) c) and d) with labels 1, 2, 4 and 5 were selected for further analysis after manual inspection as they exhibit the characteristics of SUA. The remaining set of spikes are likely to be noise and artifacts. In the figures only a subset of random samples are displayed for better visualization. The red line represents the average spike in that cluster and the orange lines are one standard deviation above and below the mean.

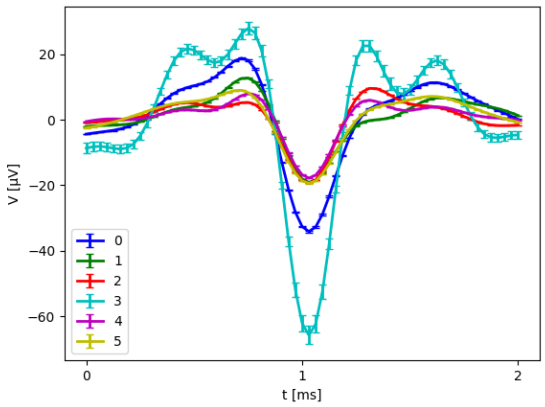


Figure 3.2.: Average spikes in the respective clusters as seen plotted against each other. The bars on the curves represent the standard error in that cluster. The numbers correspond to the cluster labels.

3.2. Classification benchmark for SUA

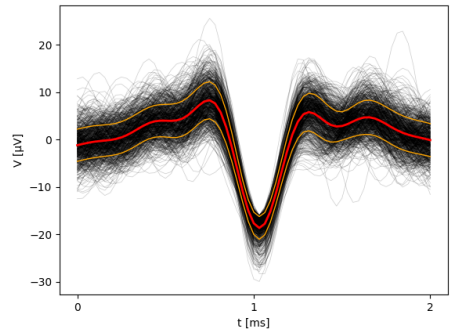
The aggregated SUA datasets clustered into one of the two categories, putative pyramidal cells and interneurons, for each patient were evaluated on the models trained on a subset of data from patient 2. Table 3.2 shows the accuracy scores for the different models. While the two classical methods, linear regression and SVM result in an accuracy range of 84%-91% across patients, the neural network model clearly outperforms these, resulting in an accuracy consistently above 90% for most patients even when the sample sizes are relatively large. Figure 3.5 shows the confusion matrix and receiver operating characteristic (ROC) curve with area under the ROC curve (AUC) to demonstrate the performance of the different models for patient 1.

3.3. IED detector with autoencoder and GAN

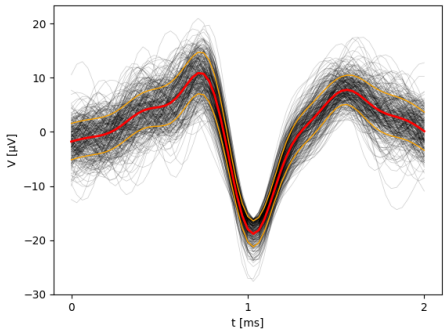
3.3.1. Autoencoder

The autoencoder showed promising results, and the model learnt the characterisitcs of IEDs well. The training and validation losses have decreased over the course of training, and at the 20th epoch they are 0.081449 and 0.083496, respectively. Figure 3.6a depicts how the model progressed as it was learning. At the end of learning, the training dataset was evaluated by the model once again to calculate the threshold based on the reconstruction error. This threshold was automatically set to 0.120026. Figure 3.6b shows how these losses are distributed. The majority of the losses are found below the threshold.

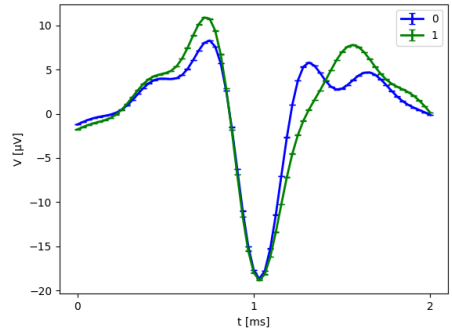
As a generative model, it is primarily capable of creating artificial IEDs, or reconstruct them. An example for reconstruction and generation can be seen in Figure 3.7. If an



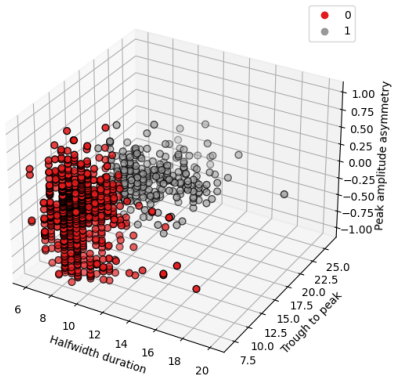
(a) Putative interneurons. Total spikes: 44152



(b) Putative pyramidal cells. Total spikes: 16773



(c) Average of the two types of spikes with the standard error indicated.



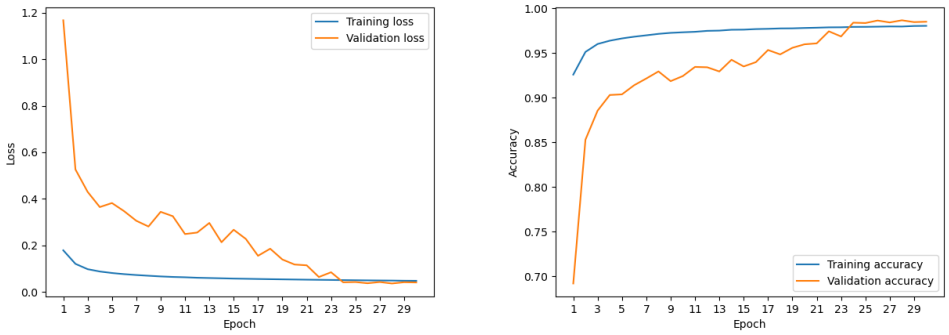
(d) Scatter plot of a subset of spikes represented by three features.

Figure 3.3.: The two clusters of spikes of putative interneurons (label 0) (a) and pyramidal cells (label 1) (b). In this case, the difference between the shape of the two types of spikes is visually observable. While spikes from interneurons are narrower, pyramidal cells tend to have a more skewed shape.

IED serves as input, the model can encode and then decode the sample such that the reconstructed output resembles the original input with relatively low error. If, however, the input differs from an IED, e.g., it is some random noise or baseline signal, the reconstruction error is high, and therefore, the sample snippet is discarded.

3.3.2. GAN

During the 100 epochs, the fluctuation of the training losses can be observed as shown in Figure 3.8. In two segments, the two losses fork, and while one increases, the other decreases suggesting that the two networks are competing with each other. Simply by looking at the losses, one could conclude that the discriminator is better at distinguishing between real and artificial samples than the generator is at generating fake ones. However,



(a) (b)

Figure 3.4.: a) Training and validation loss decrease over multiple epochs. b) Training and validation accuracy increase over multiple epochs.

such a statement would be false without looking at other measures. Since a GAN is a generative model, a generator that excels at creating artificial samples, thus, confusing the discriminator is an ideal result. In other words, a model where the loss of the discriminator is higher than that of the generator is desirable. Although the model was able to learn the high-level characteristics of IEDs, it produces a noisy output as seen in Figure 3.9. This allows us to state that the generator underfits the data and does not create perfect artificial samples. Due to this, we cannot make a conclusion that a good discriminator has been trained as it is likely that the discriminator was fed with poor quality samples by the generator. Nonetheless, the results are promising, and they are worth further investigating.

3.4. Findings for individual patients

The event plots from individual patients and channels allow a visual insight into how neurons fire in the vicinity of IEDs in the different parts of the brain. Table 3.3 shows the results of the statistical analysis on the examined channels grouped by IED clusters. The results indicate that the presence of pyramidal cells and interneurons around IED samples differs. Figure 3.10 illustrates an example of neurons firing around clusters of IEDs for patient 10, and Figure 3.11 depicts how the two types of neurons differ.

For individual patients, it can be visually observed that neurons exhibit different firing patterns around IEDs depending on the brain region. When an IED occurs in the hippocampus, spikes either accumulate or get suppressed. An example of such a scenario can be seen in Figure 3.10a where there is increased neuronal activity. On the other hand, when an IED is observed in the other parts of the brain (temporal and extratemporal lobe), SUA shows no or less consistent patterns (Figure 3.10b). Furthermore, the different types of neurons showed no observable difference in their firing patterns across multiple patients even though their number and distribution often deviate.

Patient number	Number of samples	Logistic regression	SVM	Feedforward ANN
1	150007	0.890498	0.890118	0.961822
		0.981545	0.982808	0.991069
		0.853530	0.851819	0.951915
2	3055388	0.866217	0.866387	0.925990
		0.964989	0.966185	0.991209
		0.693861	0.693374	0.824023
3	690456	0.898728	0.899200	0.961026
		0.972157	0.973175	0.990497
		0.703884	0.704622	0.886791
4	46927	0.909072	0.912886	0.977241
		0.986473	0.987266	0.982635
		0.904008	0.907869	0.990373
5	60925	0.888683	0.889848	0.966270
		0.988752	0.988977	0.993552
		0.856133	0.857560	0.959685
6	346888	0.843137	0.842586	0.899241
		0.910309	0.911811	0.978867
		0.605967	0.602990	0.724184
7	2192188	0.859602	0.860127	0.939018
		0.964902	0.966851	0.987003
		0.736427	0.735958	0.885541
8	14961	0.877481	0.877816	0.958425
		0.964595	0.965654	0.989224
		0.834351	0.833922	0.943677
9	83080	0.891695	0.892682	0.969018
		0.987900	0.988786	0.994447
		0.859462	0.860066	0.962158
10	366855	0.895046	0.895861	0.968554
		0.987335	0.987594	0.993829
		0.855173	0.856169	0.959278

Table 3.2.: Accuracy (1st in row), precision (2nd in row) and recall (3rd in row) scores for the putative pyramidal cell vs. interneuron classification models.

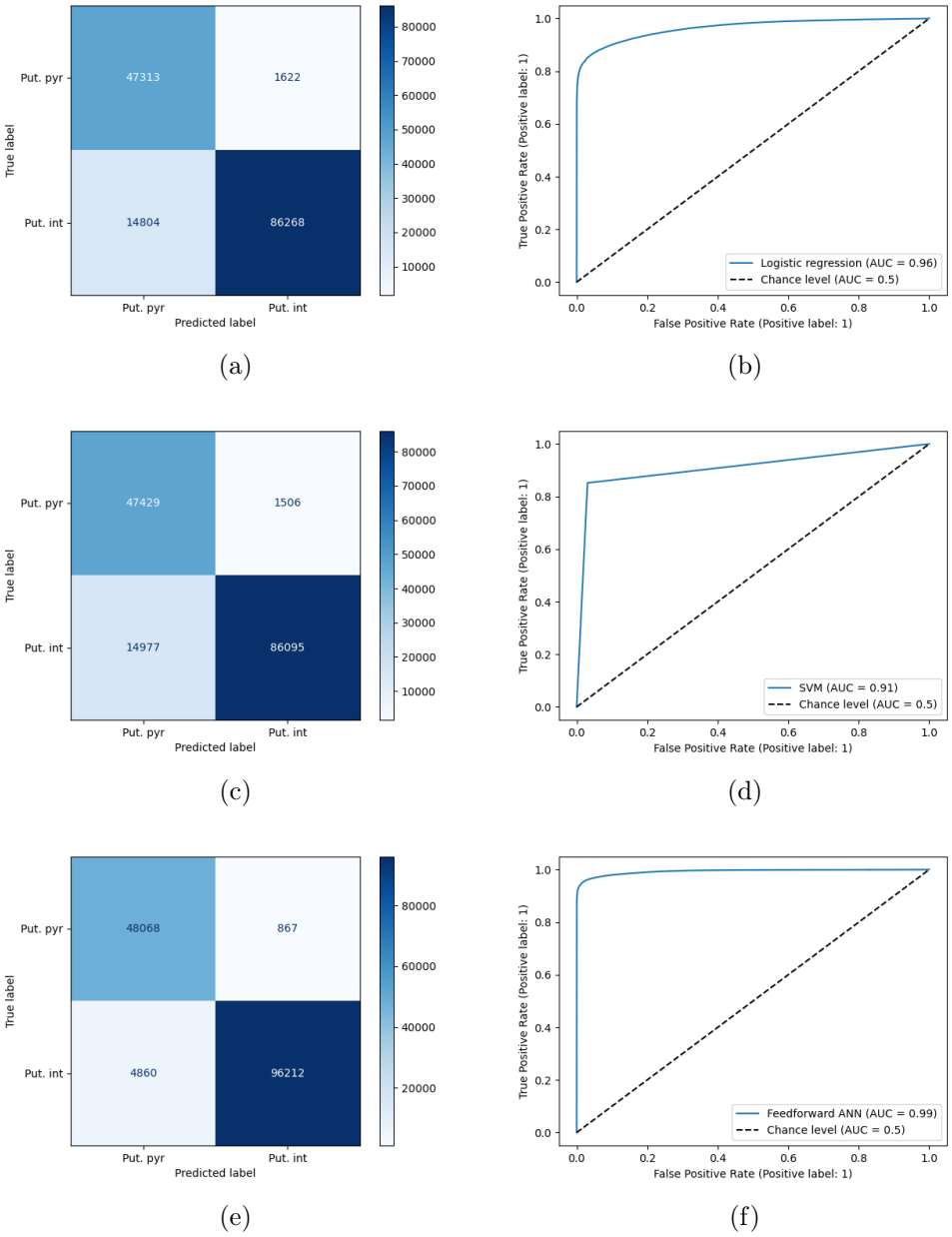
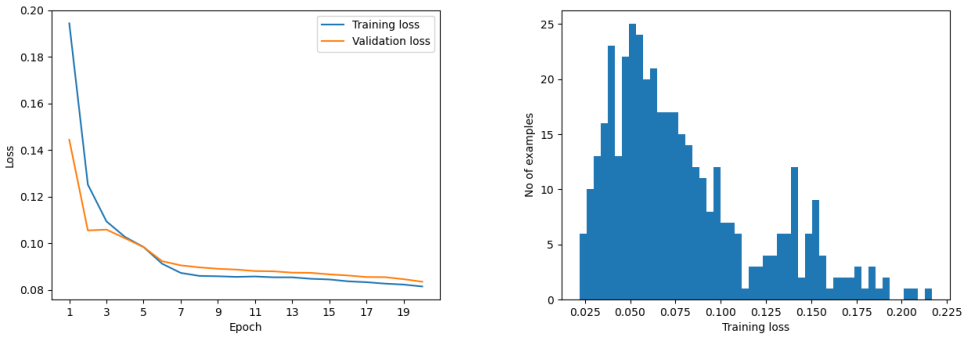
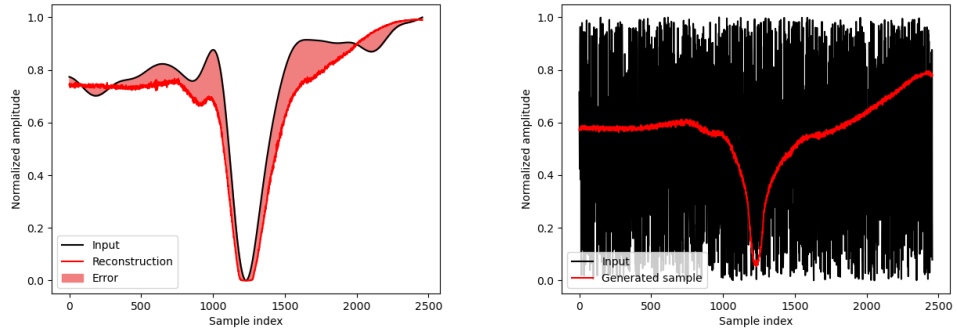


Figure 3.5.: Visualization of the benchmark for patient 1. The confusion matrix and ROC with AUC are as follows: a-b) Logistic regression. c-d) SVM. e-f) Feedforward neural network.



(a) (b)

Figure 3.6.: a) Training and validation loss over 20 epochs. b) Histogram of losses over 416 training samples.



(a) (b)

Figure 3.7.: a) A randomly selected IED, its reconstruction and error. The reconstruction error is 0.077468. As it is below the threshold, the input sample is considered an IED by the model. b) Randomly generated noise fed into the model. The output strongly resembles (artificially generates) an IED that the model was trained on. The reconstruction error in this case is 0.278971, and since it is above the threshold, the input is not considered as a valid IED.

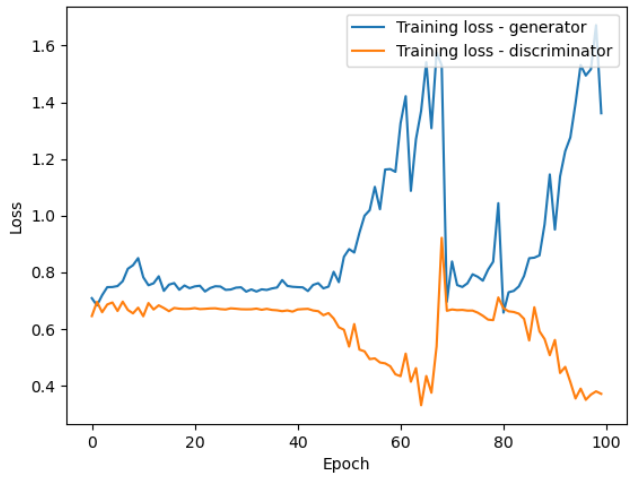


Figure 3.8.: Generator and discriminator training losses.

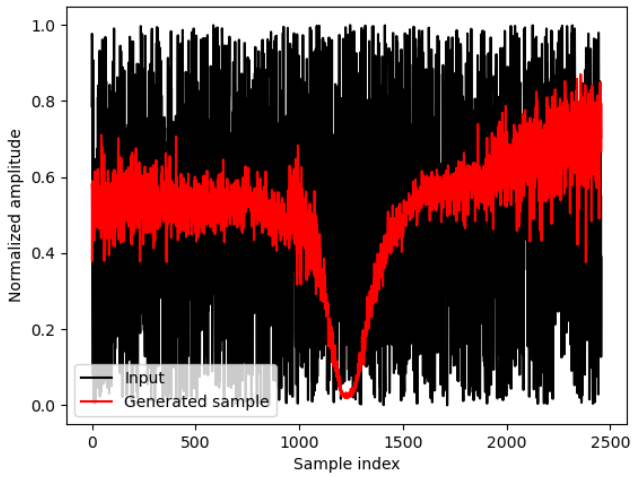
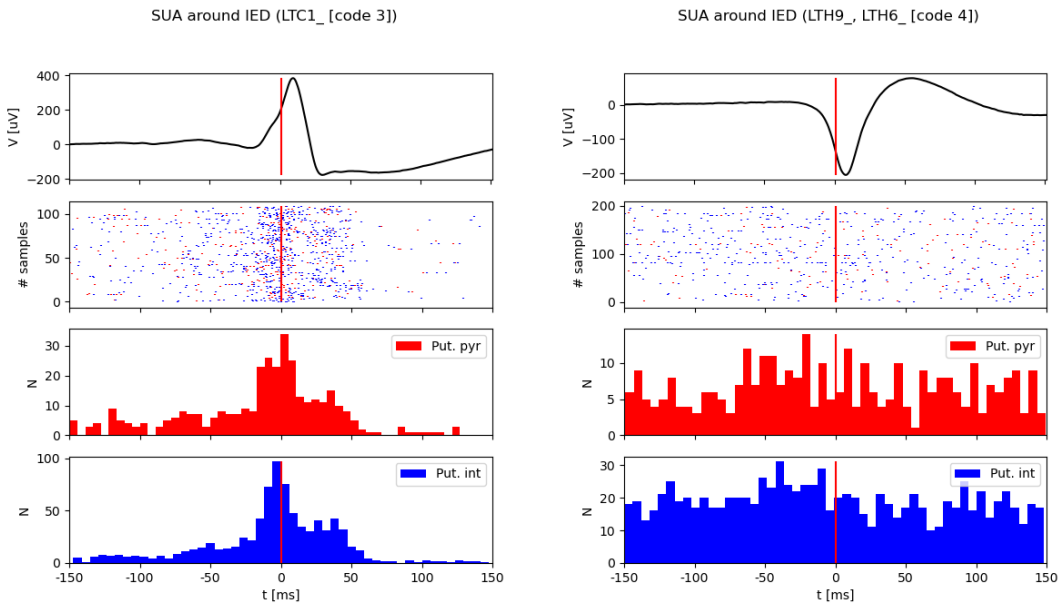


Figure 3.9.: Randomly generated noise fed into the generator. The output resembles the high-level characteristics of an IED that the model was trained on. The output is noisy as the generator appears to underfit the data.

#	IED cluster	N	$N'/\mu/\tilde{x}/\sigma$ pyramidal cells	$N'/\mu/\tilde{x}/\sigma$ interneurons	p-value	H_0
1	LTA1	200	449/2.245/2/2.036	989/4.945/4/3.058	$9.99e-5$	R
	LTB5	200	444/2.22/2/1.877	984/4.92/4/3.557	$9.99e-5$	R
	LTE7	200	503/2.515/2/2.419	948/4.74/4/3.076	$9.99e-5$	R
	LTA6	200	451/2.255/2/2.195	1022/5.11/4/3.534	$9.99e-5$	R
2	RTC1	55	1700/30.91/29/7.479	1453/26418/26/6.178	0.003	R
	RTA1	200	8364/41.82/42/5.541	5788/28.94/28/5.46	$9.99e-5$	R
	RTA1	181	7694/42.508/43/5.303	5279/29.166/29/5.662	$9.99e-5$	R
	RTB1	83	2870/34.578/34/6.682	2254/27.157/26/5.138	$9.99e-5$	R
	RTE2	200	8534/42.67/43/5.302	5762/28.81/28/4.671	$9.99e-5$	R
3	RTC1	200	2366/11.83/11/4.243	1777/8.885/7/6.044	$9.99e-5$	R
	RTA2	200	2330/11.65/11/3.364	1169/5.845/5.5/2.592	$9.99e-5$	R
	RTA1	200	2246/11.23/11/3.631	1206/6.03/5/3.132	$9.99e-5$	R
	RTD1	200	2341/11.705/11/3.178	1352/6.76/6/3.886	$9.99e-5$	R
4	RTA1	200	45/0.225/0/0.494	223/1.115/1/0.743	$9.99e-5$	R
	RTA2	119	27/0.227/0/0.457	146/1.227/1/0.772	$9.99e-5$	R
	RTF2	38	11/0.289/0/0.603	47/1.237/1/0.705	$9.99e-5$	R
	RTB7	87	21/0.241/0/0.567	120/1.379/1.464/1	$9.99e-5$	R
	RTC7	34	7/0.206/0/0.404	32/0.941/1/0.481	0.0002	R
5	RTC1	100	109/1.09/1/0.981	311/3.11/3/1.462	$9.99e-5$	R
	RTB1	200	213/1.065/1/0.975	430/2.15/2/1.615	$9.99e-5$	R
	RTB7	200	169/0.845/1/0.788	501/2.505/2/1.612	$9.99e-5$	R
	RTE2	200	195/0.975/1/0.982	720/3.6/3/1.849	$9.99e-5$	R
	RTC5	200	208/1.04/1/0.989	647/3.235/3/1.599	$9.99e-5$	R
	RTE8	200	195/0.975/1/0.919	659/3.295/3/1.737	$9.99e-5$	R
6	RTA2	200	1191/5.955/6/2.646	605/3.025/3/1.782	$9.99e-5$	R
	RPD4	200	1184/5.92/6/2.101	590/2.95/3/1.74	$9.99e-5$	R
7	RFA1	200	3825/19.125/19/3.848	3646/18.23/18/3.573	0.053	R
	RFA1	200	4054/20.27/20/3.717	3659/18.295/18/3.629	$9.99e-5$	R
	RFB2	200	3970/19.85/19.5/3.981	3769/18.845/19/3.647	0.032	R
	RFF4	200	3807/19.035/19/4.114	3531/17.655/18/3.728	0.004	R
	RFC1	200	4074/20.37/20/3.824	3829/19.145/19/3.599	0.007	R
	RTA6	200	3898/19.49/19/3.748	3565/17.825/18/3.253	0.0001	R
	RTA6	200	4017/20.085/20/4.126	3538/17.69/17/3.384	$9.99e-5$	R
8	LTA1,LTB1	200	224/1.12/1/1.023	474/2.37/2/1.286	$9.99e-5$	R
	LTA1,LTB1	200	289/1.445/1/1.117	610/3.05/3/1.642	$9.99e-5$	R
9	RPA7	200	247/1.235/1/1.144	1970/9.85/9/3.103	$9.99e-5$	R
	LTB1	27	44/1.629/1/1.127	319/11.815/12/2.695	$9.99e-5$	R
	LOA3	200	434/2.17/2/1.622	1186/5.93/5/2.718	$9.99e-5$	R
	RPA7	148	226/1.527/1/1.249	2142/14.473/14/2.757	$9.99e-5$	R
10	LTA6,LTH6	200	236/1.18/1/1.469	583/2.915/2/4.427	$9.99e-5$	R
	LTC1	200	488/2.44/2/1.66	1097/5.485/5/2.953	$9.99e-5$	R
	LTC1	109	352/3.229/3/1.942	833/7.642/7/2.359	$9.99e-5$	R
	LTH9,LTH6	200	334/1.67/1/1.858	953/4.765/3/5.012	$9.99e-5$	R

Table 3.3.: Results of the statistical analysis of pyramidal cells and interneurons around IED samples for each patient. N: sample size (number of IEDs in a window stacked above each other), N': number of spikes of the given neuronal type, μ: sample mean, $\tilde{x}$: sample median, σ: standard deviation. The IED cluster column may contain repetitive electrode labels for one patient. This means that on a given channel multiple clusters of IEDs were present. The R in H_0 means that the null hypothesis was rejected.



(a) (b)

Figure 3.10.: Single unit activity around two clusters of IED in the hippocampus (a) and the temporal lobe (b) of patient 10. The neurons show a consistent firing pattern in the hippocampus, while they do not exhibit a correlation with IEDs further away in the brain. The histograms represent the absolute count of spikes in a bin, and the top panels containing the IED are the mean of all samples in that cluster and window.

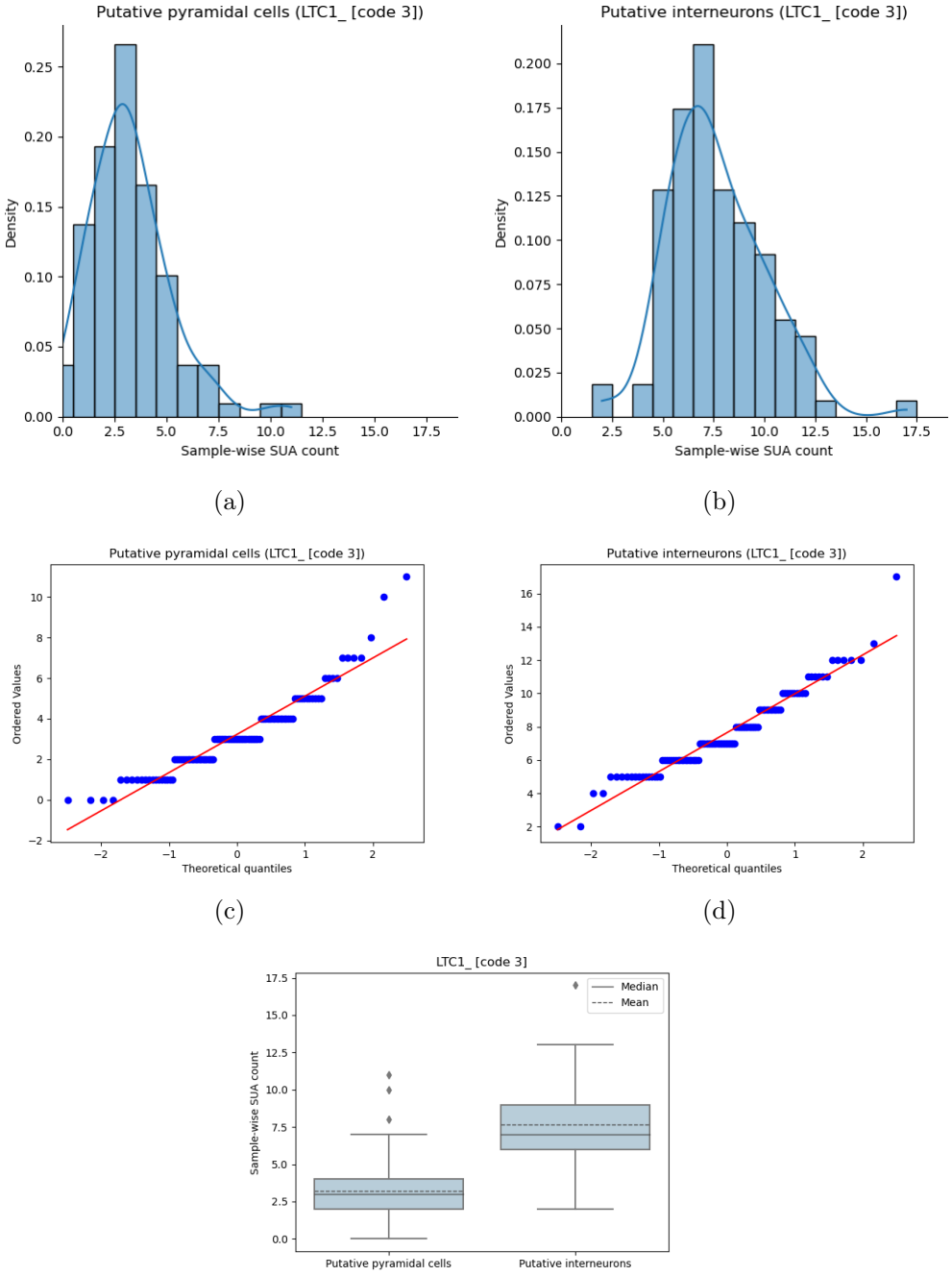


Figure 3.11.: Visualization of some of the statistics of the hippocampal IED window of patient 10 as seen in Figure 3.10a. a-b) Normalized histogram and density of the two types of neurons firing. c-d) Probability plots of the two distributions. Data points more aligned with the red line represent a more normally distributed dataset. e) Boxplots of the two distributions with the mean, median and outliers indicated.

4. Discussion

4.1. Detection and classification of SUA

Conventional methods for the detection of SUA such as threshold crossing provide a reliable outcome, but they still have the overhead of the need for manual verification. That is, clusters of outliers have to be eliminated by hand, and clusters that are deemed sufficient are passed forward. This method is, unfortunately, prone to errors and time-consuming, especially when dealing with large number of samples.

The differentiation between spikes from pyramidal cells and interneurons using the methods described earlier has its problems as pointed out previously. As the benchmark implies, supervised methods provide a good alternative to the k-means algorithm in the current scenario. Their major advantage is that they address the problem of having imbalanced data, when one type of neurons is scarce and the other type is largely present in the dataset. In other words, issues with clustering algorithms that are sensitive to inter- and intra-cluster distances of points disappear. Supervised methods, however, require labelled data and, hence, access to the ground truth, i.e., the true characteristics of spikes from different neurons, and use it as a basis for training. While in this thesis such a trained model is provided as a sufficient amount of data was available, it was acquired using *pseudo ground truth* as unfortunately, in-vivo intracellular recording of spikes is not possible in a clinical setting.

The benchmark compared three binary classifiers, and all of them showed promising results with the simple feedforward neural network outperforming the two classical models. Logistic regression and vanilla SVM are considered linear models, while ANNs with multiple nodes and (hidden) layers are non-linear. This non-linearity allows the model to learn more complex features of the input, however, they are also more prone to overfitting. This non-linear characteristic results in better accuracy. Spikes, even though their feature vector is relatively low dimensional, they can be too complex for linear models. However, SVMs, for example, can be used for classification when data is not linearly separable. This can be achieved with the kernel trick.

It may be worthwhile exploring different supervised methods to solve the problem of detection and classification. Different architectures of neural networks could potentially achieve better results in both detecting spikes and differentiating between the neuronal types. CNN could be a promising choice as it can also be applied on time series data. For example, the spectogram of the signal serves as the input to the model, and then it is treated and processed as an image by the network. *Dynamic time warping (DTW)* is a similar method to template matching. It can compare the similarity between two vectors of varying length, i.e., a spike template and a raw snippet that do not necessarily

have the same length. This way the problem of slight distortion and shift in the temporal domain can be remedied. DTW could be particularly handy in the detection phase.

Regardless of the chosen model used to differentiate the two putative cell types and the high accuracy achieved for each patient, results might be biased due to inaccessibility to the ground truth, i.e., the true characteristics of spikes at training. Therefore, it is important to stress that for more robust classification, other data acquiring (e.g., intracellular recording) technique is needed as a basis for training. Furthermore, another obstacle is that not all patients may exhibit clear and easily distinguishable spikes from pyramidal cells and interneurons.

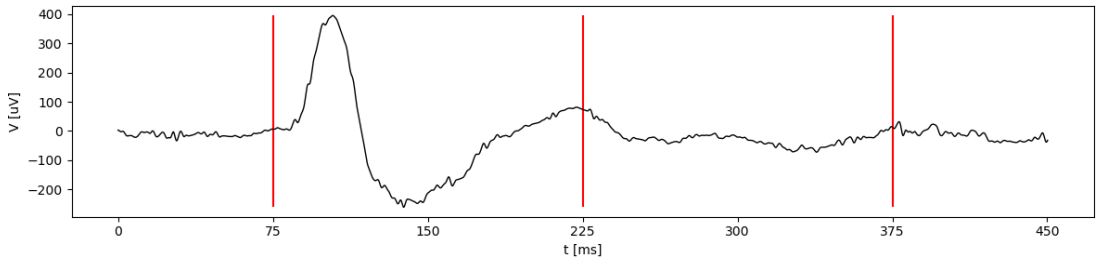
4.2. IED detection using autoencoder and GAN

As seen in the results, both models provide intriguing outcomes. They both learnt how to represent IEDs relatively well. The autoencoder, however, performs better as it can reproduce an IED with less noise than the GAN. Additionally, more objective and concrete measures such as the reconstruction error are provided with the former model. In contrast, to evaluate the performance of the GAN, we relied more on subjective measures such as the comparison of generated IEDs with the training data visually. As discussed in previous sections, autoencoders are also more straightforward and less cumbersome to train, while GANs prove to be more difficult to handle in that regard. This is contributed to the fact that in a GAN two separate neural networks have to be trained simultaneously while they are competing against each other to minimize their own loss. In addition, training in the latter case can be significantly more time-consuming.

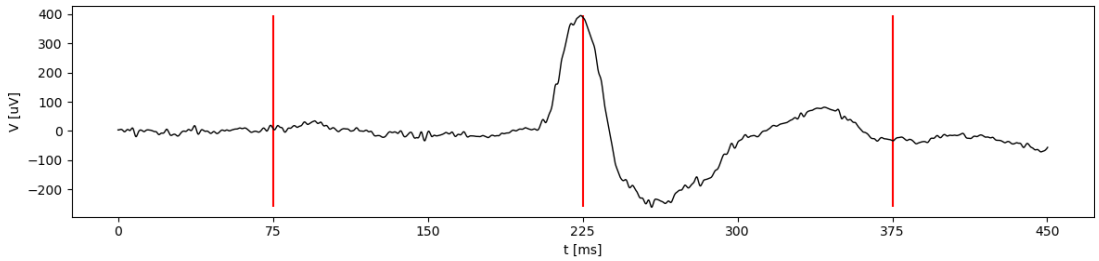
As an improvement, the training dataset should be expanded to include a higher amount of samples with a lot more homogeneous IEDs from multiple patients. The models were trained using only 416 samples with a comparably higher dimension (2458), and data from only one channel of patient 8. This is not the most ideal setup, and therefore, a more diverse dataset is desirable. Besides, the models themselves could potentially contribute to (artificially) creating more samples given their generative property. The models could be further tuned by adding or removing layers and nodes; using dropout; using different activation functions and optimizers; and experimenting with other hyperparameters such as the number of epochs and batch sizes. Efficient quantitative measures should be used to evaluate the performance of the GAN.

The windowing technique is a straightforward and easy way to segment a large signal into smaller snippets that serve as the input for classification models. The simplest idea is to cut the signal into snippets of a target duration (in this case, a window of 300ms or 2458 data points). Then the signal within the snippet is centered according to its maximum or minimum amplitude (whose absolute value is larger). The normalized sample is then fed into the model that decides whether it is an IED or not. (In more sophisticated cases, IEDs can also be classified.) Some problems, however, arise with this method. IEDs take on very homogeneous forms as discussed earlier. Some resemble a sharp spike, some are more curvy, and some are complex with multiple spikes and bumps. This means that the duration of IEDs may vary greatly. It can also happen that an EEG exhibits high

frequency IED occurrence resulting in them being too close to each other. Choosing a fixed length for snippets is not flexible enough to handle such edge cases. If a window is too large, it may include multiple IEDs and other artifacts, but if it is too small, the entirety of an IED is not captured. (The widely accepted convention is to mark IEDs at their highest (lowest) amplitude. Not enclosing all oscillations within one window may be acceptable in some cases.) Another issue is the centering within a window. If an IED's maxima (minima) lies on the border separating adjacent snippets, both these windows will center it locally. This means that the left-neighbor window will shift the IED to the left, and the right-neighbor will shift it to the right. This will result in detecting the IED twice. Figure 4.1 illustrates the problem. To mitigate the issue, overlapping windows could be introduced, and IED markings that are duplicates or are too close to each other could be dropped or merged.



(a) Ideal segmentation. The IED is well represented in the window of 75-225ms.



(b) Segmentation that is not ideal as the IED is cut in half at 225ms.

Figure 4.1.: The problem of the windowing technique with centering. The vertical lines represent the border of the snippets. a) The IED fits in one window (there is still a cut-off of the accompanying bump). Here, the IED is detected once. b) The IED lies on the border of two windows and it stretches across them. Here, both windows will locally center it according to its maximum amplitude, and two copies will result it duplicate detection.

Due to the homogeneous training dataset and issues with the windowing technique, the ratio of false positives and false negatives may be high, and therefore, neither models were used as detectors in this study. Nevertheless, both of them show great potential and benefit over other supervised and unsupervised methods due to having to have merely positive samples. Moreover, they show superiority over other existing traditional

techniques as well. With a well-trained model and a diverse and large training dataset, the semi-automatic template matching algorithm may become obsolete. Overall, this proof of concept sets the foundations for further research and exploration for a more robust and less error-prone IED detector.

4.3. Medically relevant findings

The event plots gained from the individual patients exhibit some consistent single unit activity around IEDs within the hippocampus, and in the other two parts of the brain such a pattern is less likely to occur. This confirms that there is a disruption of SUA by IED. Spatial distribution of epileptiform activity might be an important determinant of functional disruption in the hippocampus in epilepsy. The statistical tests suggest that putative pyramidal cells and interneurons fire differently around IEDs. There is, however, no clear evidence consistent across all patients that one type of neurons fire in a significantly different way than the other. Therefore, further analyses are necessary.

4.4. Conclusion

The analyses presented in this thesis exhibit intriguing and promising outputs from both a medical and data analysis perspective. Some of the methods provide rudimentary insights into their potential and give a direction into future research. Robust detection of IEDs on a large scale still poses challenges, however, as shown in this work, different neural network architectures such as autoencoders and GAN already provide convincing and plausible results. Similarly, the models distinguishing spikes from pyramidal cells and interneurons is more vigorous than the k-means clustering algorithm. Even though the usage of novel machine learning models could be revolutionary both in research and diagnosis, ethical, legal and technical obstacles may appear. Additionally, a relatively large amount of high quality data is necessary to train these models as described in previous sections.

While in this thesis only one possibility of quantitative analysis was presented, based on the medical insight needed, there are other statistical test cases that can be carried out. For example, comparing single unit activity around IEDs in the different brain regions could quantitatively support the visual results. Such analyses are left for further research.

This thesis also aimed to prove that there is clearly a need for medical experts and computer scientist to work together efficiently in a multidisciplinary way. Older and cumbersome methods ought to be reviewed on a regular basis, and experimenting with novel technologies should be encouraged as keeping up with state-of-the-art advancements in the field of data analysis and machine learning is key to research not only in epilepsy, but in the medical field as a whole.

Bibliography

[1] J. Moncrieff, R. E. Cooper, T. Stockmann, S. Amendola, M. P. Hengartner, and M. A. Horowitz, "The serotonin theory of depression: a systematic umbrella review of the evidence," *Molecular Psychiatry*, July 2022.

[2] WHO, "Epilepsy." `https://www.who.int/news-room/fact-sheets/detail/epilepsy`, 2023 (accessed 2023-05-07).

[3] R. S. Fisher, W. van Emde Boas, W. Blume, C. Elger, P. Genton, P. Lee, and J. Engel, "Epileptic seizures and epilepsy: Definitions proposed by the international league against epilepsy (ILAE) and the international bureau for epilepsy (IBE)," *Epilepsia*, vol. 46, pp. 470–472, Apr. 2005.

[4] R. S. Fisher, C. Acevedo, A. Arzimanoglou, A. Bogacz, J. H. Cross, C. E. Elger, J. Engel, L. Forsgren, J. A. French, M. Glynn, D. C. Hesdorffer, B. Lee, G. W. Mathern, S. L. Moshé, E. Perucca, I. E. Scheffer, T. Tomson, M. Watanabe, and S. Wiebe, "ILAE official report: A practical clinical definition of epilepsy," *Epilepsia*, vol. 55, pp. 475–482, Apr. 2014.

[5] K. J. Staley and F. E. Dudek, "Interictal spikes and epileptogenesis," *Epilepsy Currents*, vol. 6, pp. 199–202, Oct. 2006.

[6] R. E. D. Bautista, "On the nature of interictal epileptiform discharges," *Clinical Neurophysiology*, vol. 124, pp. 2073–2074, Nov. 2013.

[7] NIH, "Epilepsy and seizures." `https://www.ninds.nih.gov/health-information/disorders/epilepsy-and-seizures`, 2023 (accessed 2023-05-07).

[8] I. E. Scheffer, S. Berkovic, G. Capovilla, M. B. Connolly, J. French, L. Guilhoto, E. Hirsch, S. Jain, G. W. Mathern, S. L. Moshé, D. R. Nordli, E. Perucca, T. Tomson, S. Wiebe, Y.-H. Zhang, and S. M. Zuberi, "Ilae classification of the epilepsies: Position paper of the ilae commission for classification and terminology," *Epilepsia*, vol. 58, pp. 512–521, Mar. 2017.

[9] R. S. Fisher, J. H. Cross, C. D'Souza, J. A. French, S. R. Haut, N. Higurashi, E. Hirsch, F. E. Jansen, L. Lagae, S. L. Moshé, J. Peltola, E. R. Perez, I. E. Scheffer, A. Schulze-Bonhage, E. Somerville, M. Sperling, E. M. Yacubian, and S. M. Zuberi, "Instruction manual for the ilae 2017 operational classification of seizure types," *Epilepsia*, vol. 58, pp. 531–542, Mar. 2017.

[10] P. Kwan, A. Arzimanoglou, A. T. Berg, M. J. Brodie, W. A. Hauser, G. Mathern, S. L. Moshé, E. Perucca, S. Wiebe, and J. French, "Definition of drug resistant epilepsy: Consensus proposal by the ad hoc task force of the ILAE commission on therapeutic strategies," *Epilepsia*, vol. 51, pp. 1069–1077, Nov. 2009.

[11] R. P. N. Rao, *Brain-Computer Interfacing: An Introduction*. USA: Cambridge University Press, 2013.

[12] R. H. Masland, "Neuronal cell types," *Current Biology*, vol. 14, pp. R497–R500, July 2004.

[13] B. Teleńczuk and A. Destexhe, "Local field potential, relationship to unit activity," in *Encyclopedia of Computational Neuroscience*, pp. 1–6, Springer New York, 2014.

[14] G. Buzsáki, C. A. Anastassiou, and C. Koch, "The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes," *Nature Reviews Neuroscience*, vol. 13, pp. 407–420, May 2012.

[15] P. J. Karoly, D. R. Freestone, R. Boston, D. B. Grayden, D. Himes, K. Leyde, U. Seneviratne, S. Berkovic, T. O'Brien, and M. J. Cook, "Interictal spikes and epileptic seizures: their relationship and underlying rhythmicity," *Brain*, vol. 139, pp. 1066–1078, Feb. 2016.

[16] M. A. Marco de Curtis, John G R Jefferys, "Interictal epileptiform discharges in partial epilepsy: Complex neurobiological mechanisms based on experimental and clinical evidence," *Jasper's Basic Mechanisms of the Epilepsies*, 2012.

[17] M. de Curtis and G. Avanzini, "Interictal spikes in focal epileptogenesis," *Progress in Neurobiology*, vol. 63, pp. 541–567, Apr. 2001.

[18] C. Alvarado-Rojas, K. Lehongre, J. Bagdasaryan, A. Bragin, R. Staba, J. Engel, V. Navarro, and M. LE VAN QUYEN, "Single-unit activities during epileptic discharges in the human hippocampal formation," *Frontiers in Computational Neuroscience*, vol. 7, p. 140, 2013.

[19] R. Q. Quiroga, Z. Nadasdy, and Y. Ben-Shaul, "Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering," *Neural Computation*, vol. 16, pp. 1661–1687, Aug. 2004.

[20] C. da Silva Lourenço, M. C. Tjepkema-Cloostermans, and M. J. van Putten, "Machine learning for detection of interictal epileptiform discharges," *Clinical Neurophysiology*, vol. 132, no. 7, pp. 1433–1443, 2021.

[21] M. Blatt, S. Wiseman, and E. Domany, "Superparamagnetic clustering of data," *Physical Review Letters*, vol. 76, pp. 3251–3254, Apr. 1996.

[22] R. Q. Quiroga, "Spike sorting." `http://www.scholarpedia.org/article/Spike_sorting`, 2007 (accessed 2023-05-20).

[23] C. Rossant, S. N. Kadir, D. F. M. Goodman, J. Schulman, M. L. D. Hunter, A. B. Saleem, A. Grosmark, M. Belluscio, G. H. Denfield, A. S. Ecker, A. S. Tolias, S. Solomon, G. Buzsáki, M. Carandini, and K. D. Harris, "Spike sorting for large, dense electrode arrays," *Nature Neuroscience*, vol. 19, pp. 634–641, Mar. 2016.

[24] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," KDD'96, p. 226–231, AAAI Press, 1996.

[25] J. Csicsvari, H. Hirase, A. Czurko, and G. Buzsáki, "Reliability and state dependence of pyramidal cell–interneuron synapses in the hippocampus," *Neuron*, vol. 21, pp. 179–189, July 1998.

[26] I. V. Viskontas, A. D. Ekstrom, C. L. Wilson, and I. Fried, "Characterizing interneuron and pyramidal cells in the human medial temporal lobe in vivo using extracellular recordings," *Hippocampus*, vol. 17, no. 1, pp. 49–57, 2006.

[27] B. Elahian, N. E. Lado, E. Mankin, S. Vangala, A. Misra, K. Moxon, I. Fried, A. Sharan, M. Yeasin, R. Staba, A. Bragin, M. Avoli, M. R. Sperling, J. Engel, and S. A. Weiss, "Low-voltage fast seizures in humans begin with increased interneuron firing," *Annals of Neurology*, vol. 84, pp. 588–600, Oct. 2018.

[28] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, Sept. 1995.

[29] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics* (G. Gordon, D. Dunson, and M. Dudík, eds.), vol. 15 of *Proceedings of Machine Learning Research*, (Fort Lauderdale, FL, USA), pp. 315–323, PMLR, 11–13 Apr 2011.

[30] PyTorch, "Bce loss - pytorch." `https://pytorch.org/docs/stable/generated/torch.nn.BCELoss.html`, 2023 (accessed 2023-05-19).

[31] G. Hinton, "Overview of mini-batch gradient descent." `https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf`, 2014 (accessed 2023-05-19).

[32] A. Graves, "Generating sequences with recurrent neural networks," *CoRR*, vol. abs/1308.0850, 2013.

[33] PyTorch, "Rmsprop - pytorch." `https://pytorch.org/docs/stable/generated/torch.optim.RMSprop.html`, 2023 (accessed 2023-05-19).

[34] M. El-Gohary, J. McNames, and S. Elsas, "User-guided interictal spike detection," in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 821–824, 2008.

[35] P. C. Horak, S. Meisenhelter, M. E. Testorf, A. C. Connolly, K. A. Davis, and B. C. Jobst, "Implementation and evaluation of an interictal spike detector," in *SPIE Proceedings* (P. J. Bones, M. A. Fiddy, and R. P. Millane, eds.), SPIE, Sept. 2015.

[36] J. Jing, J. Dauwels, T. Rakthanmanon, E. Keogh, S. Cash, and M. Westover, "Rapid annotation of interictal epileptiform discharges via template matching under dynamic time warping," *Journal of Neuroscience Methods*, vol. 274, pp. 179–190, Dec. 2016.

[37] A. Palepu, S. Premanathan, F. Azhar, M. Vendrame, T. Loddenkemper, C. Reinsberger, G. Kreiman, K. Parkerson, S. Sarma, and W. Anderson, "Automating interictal spike detection: Revisiting a simple threshold rule," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, July 2018.

[38] D. E. Rumelhart and J. L. McClelland, *Learning Internal Representations by Error Propagation*, pp. 318–362. 1987.

[39] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[40] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings* (Y. Bengio and Y. LeCun, eds.), 2015.

[41] PyTorch, "Adam - pytorch." `https://pytorch.org/docs/stable/generated/torch.optim.Adam.html`, 2023 (accessed 2023-05-19).

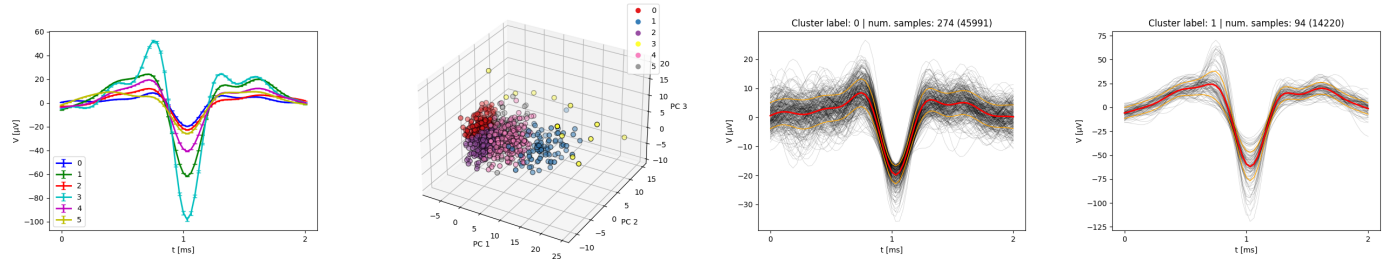
[42] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'14, (Cambridge, MA, USA), p. 2672–2680, MIT Press, 2014.

[43] I. J. Goodfellow, "NIPS 2016 tutorial: Generative adversarial networks," *CoRR*, vol. abs/1701.00160, 2017.

[44] M. Lucic, K. Kurach, M. Michalski, O. Bousquet, and S. Gelly, "Are gans created equal? a large-scale study," in *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NIPS'18, (Red Hook, NY, USA), p. 698–707, Curran Associates Inc., 2018.

[45] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training gans," in *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, (Red Hook, NY, USA), p. 2234–2242, Curran Associates Inc., 2016.

A. Spike sorting results



(a) Average of spikes with standard error.

(b) Scatter plot of a subset of spikes with first three PCs.

(c)

(d)

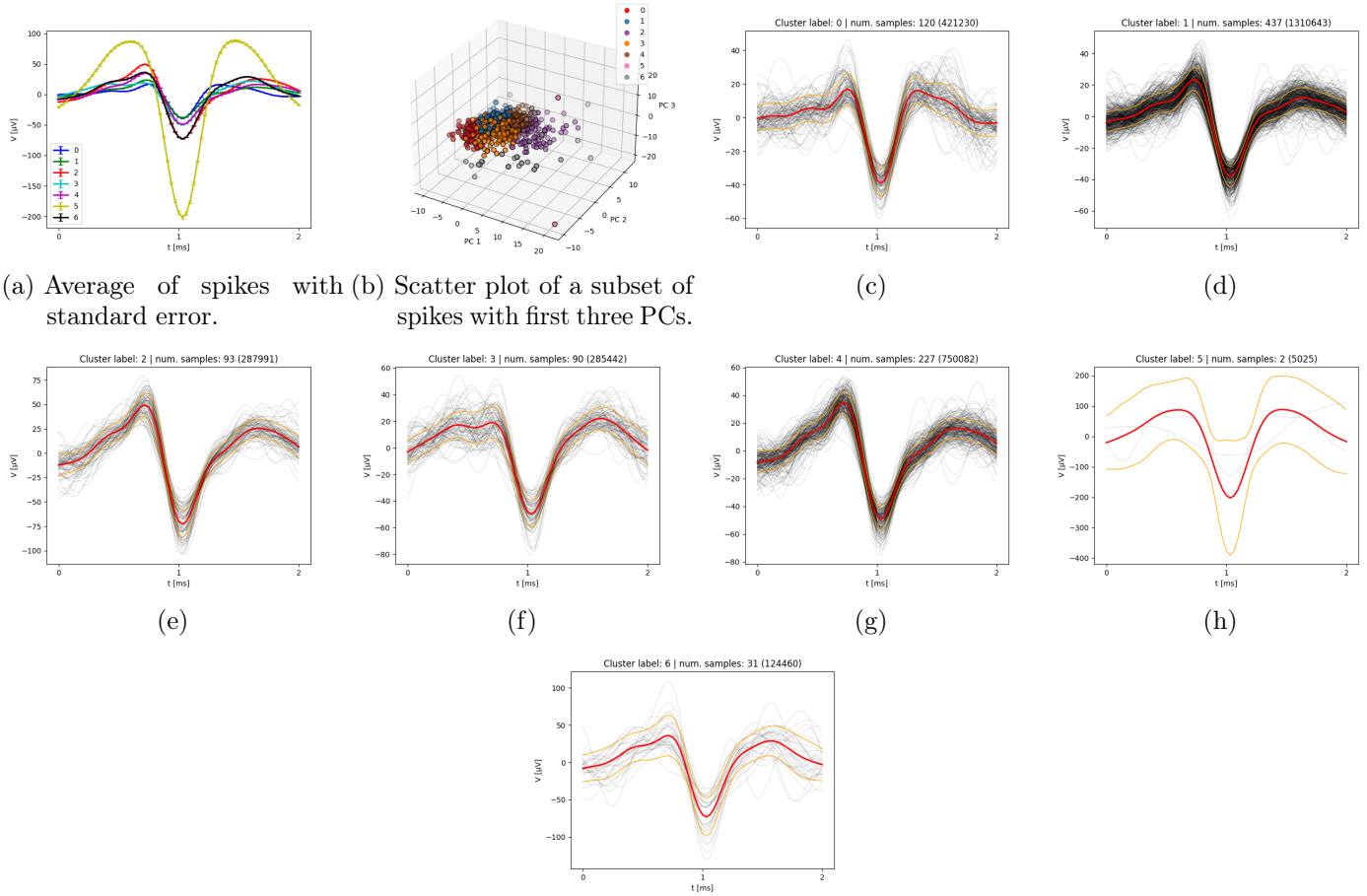
(e)

(f)

(g)

(h)

Figure A.1.: Clusters of spikes of patient 1. Assigned labels, randomly drawn and total number of samples are indicated in the title and legend. Clusters with the following labels were kept: 0, 2, 4, 5.



(a) Average of spikes with standard error.

(b) Scatter plot of a subset of spikes with first three PCs.

(c)

(d)

(e)

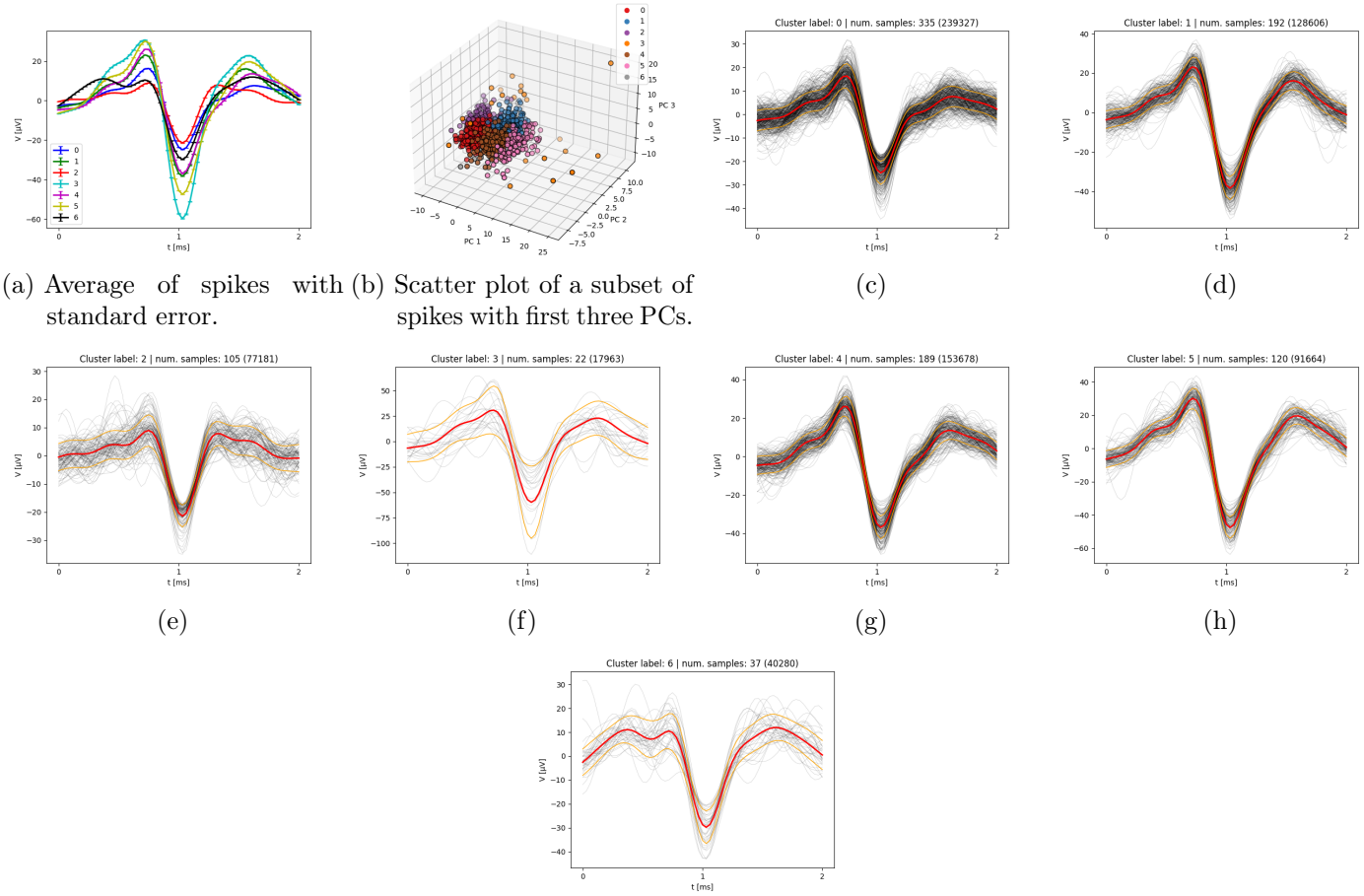
(f)

(g)

(h)

(i)

Figure A.2.: Clusters of spikes of patient 2. Assigned labels, randomly drawn and total number of samples are indicated in the title and legend. Clusters with the following labels were kept: 0, 1, 2, 3, 4.



(a) Average of spikes with standard error.

(b) Scatter plot of a subset of spikes with first three PCs.

(c)

(d)

(e)

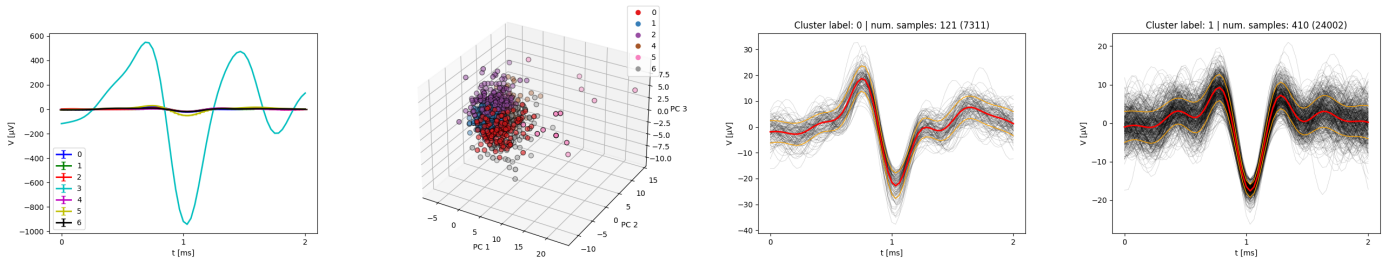
(f)

(g)

(h)

(i)

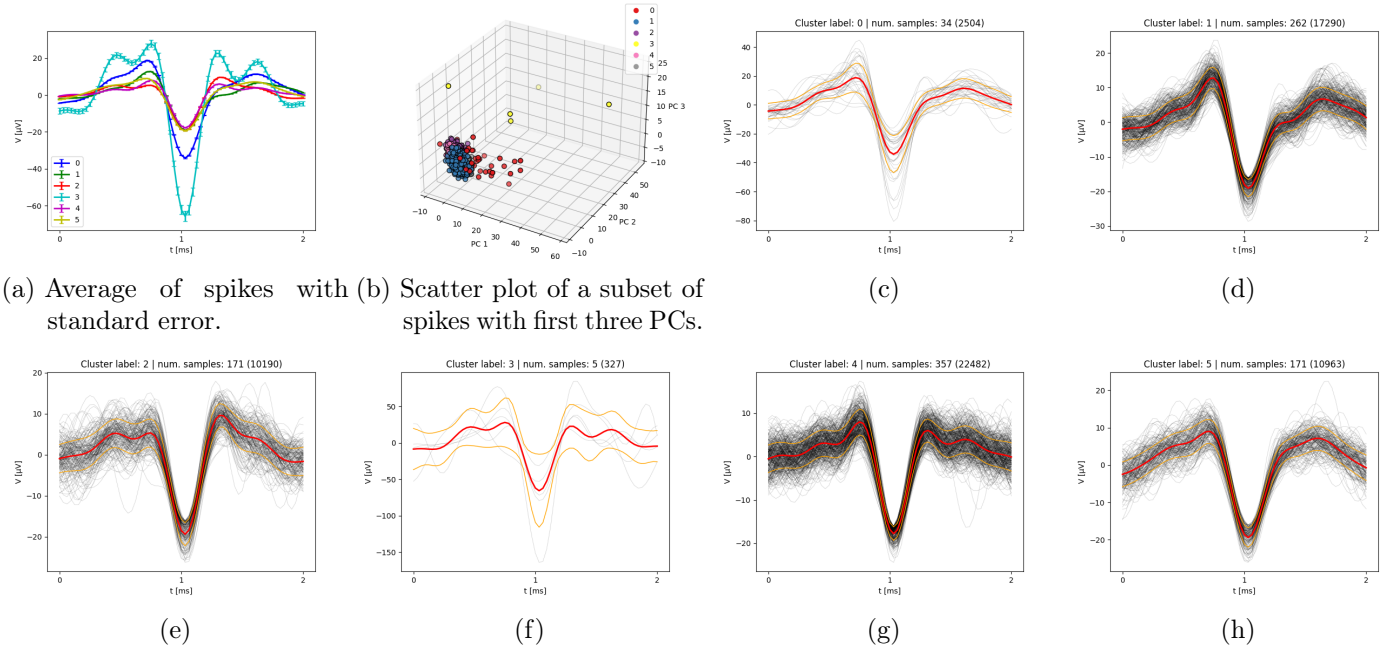
Figure A.3.: Clusters of spikes of patient 3. Assigned labels, randomly drawn and total number of samples are indicated in the title and legend. Clusters with the following labels were kept: 0, 1, 2, 4, 5.



(a) Average of spikes with standard error. (b) Scatter plot of a subset of spikes with first three PCs. (c) (d)

(e) (f) (g) (h)

Figure A.4.: Clusters of spikes of patient 4. Assigned labels, randomly drawn and total number of samples are indicated in the title and legend. Clusters with the following labels were kept: 0, 1, 2, 3, 6.



(a) Average of spikes with standard error.

(b) Scatter plot of a subset of spikes with first three PCs.

(c)

(d)

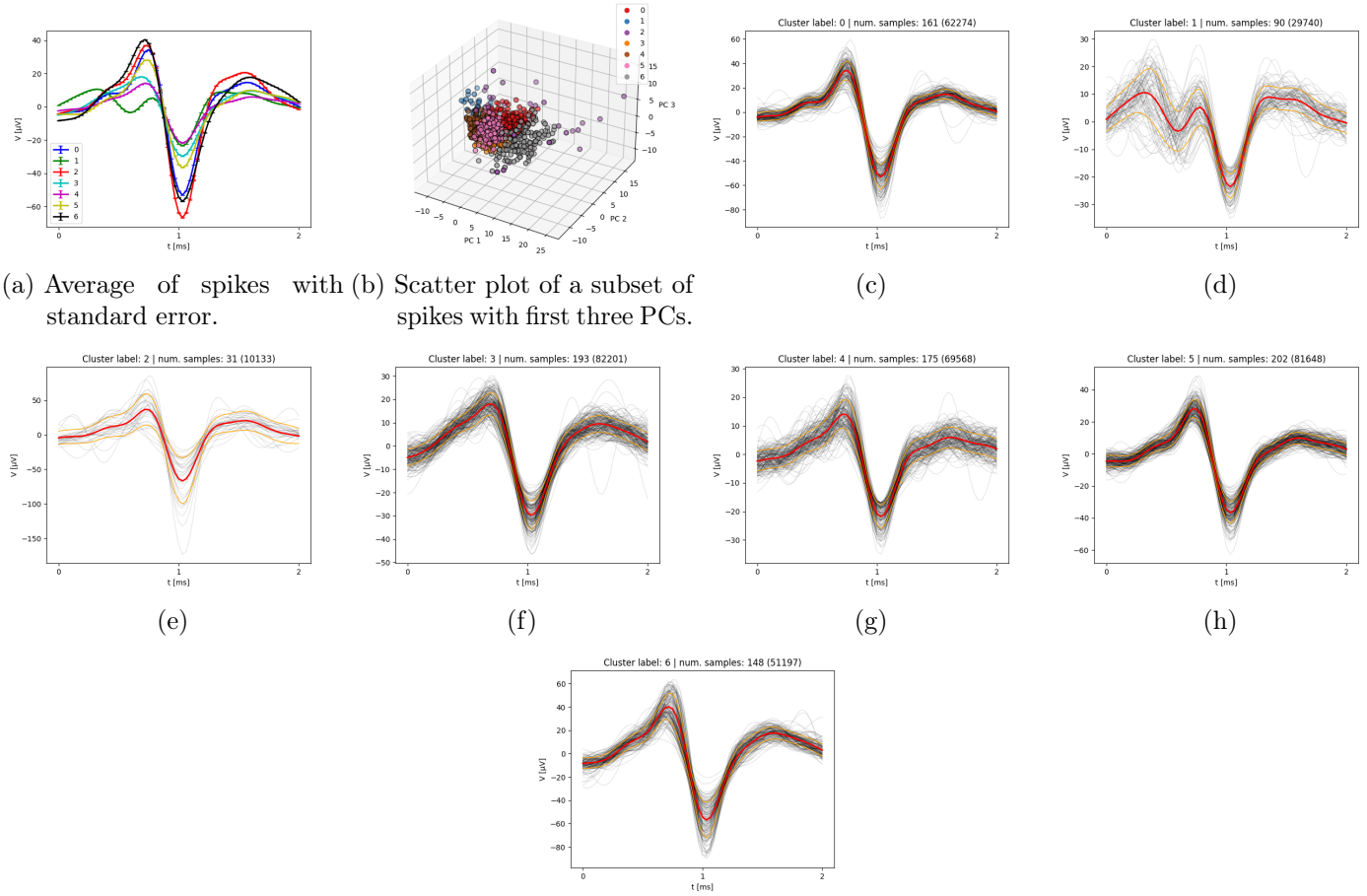
(e)

(f)

(g)

(h)

Figure A.5.: Clusters of spikes of patient 5. Assigned labels, randomly drawn and total number of samples are indicated in the title and legend. Clusters with the following labels were kept: 1, 2, 4, 5.



(a) Average of spikes with standard error.

(b) Scatter plot of a subset of spikes with first three PCs.

(c)

(d)

(e)

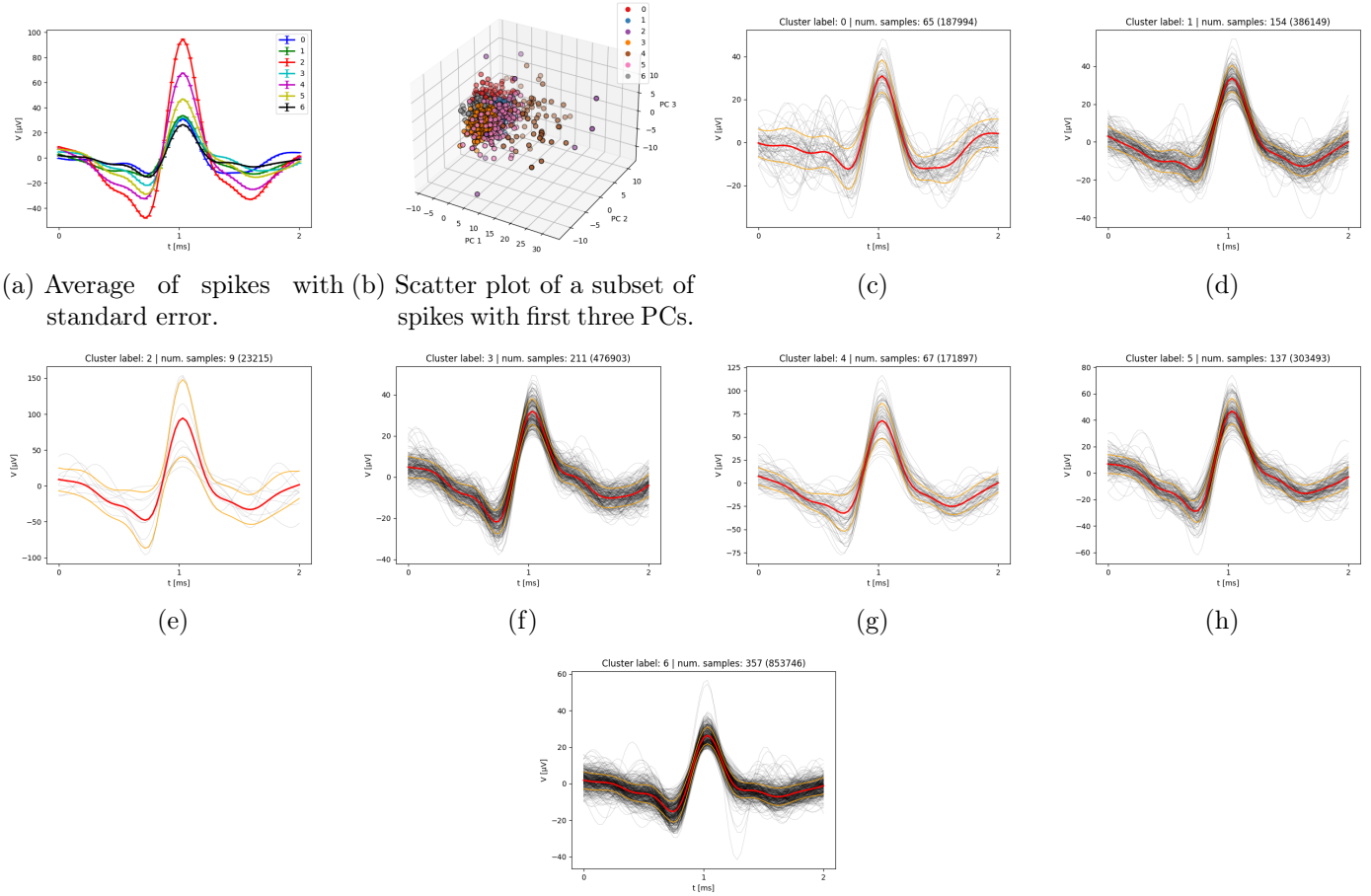
(f)

(g)

(h)

(i)

Figure A.6.: Clusters of spikes of patient 6. Assigned labels, randomly drawn and total number of samples are indicated in the title and legend. Clusters with the following labels were kept: 0, 3, 4, 5, 6.



(a) Average of spikes with standard error.

(b) Scatter plot of a subset of spikes with first three PCs.

(c)

(d)

(e)

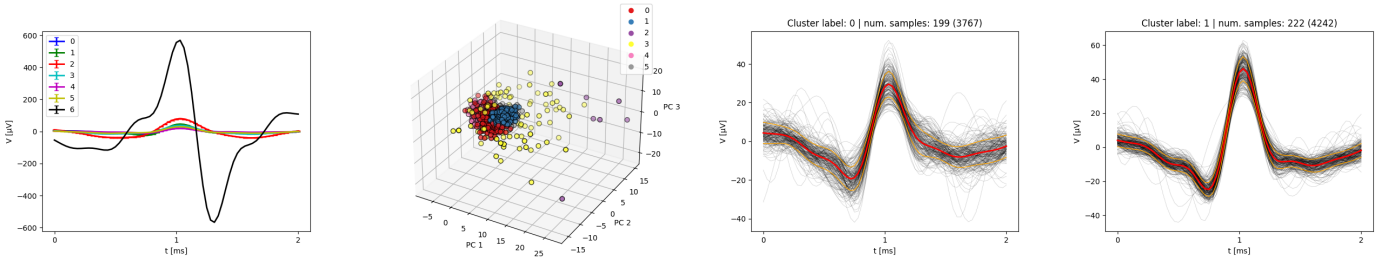
(f)

(g)

(h)

(i)

Figure A.7.: Clusters of spikes of patient 7. Assigned labels, randomly drawn and total number of samples are indicated in the title and legend. Clusters with the following labels were kept: 1, 3, 4, 5, 6.

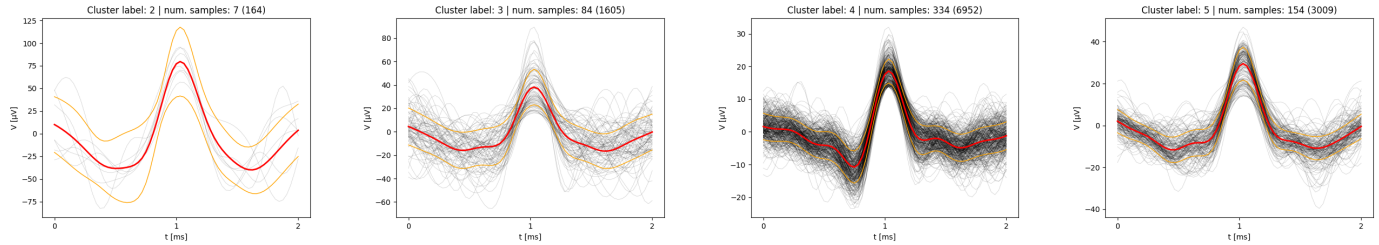


(a) Average of spikes with standard error.

(b) Scatter plot of a subset of spikes with first three PCs.

(c)

(d)



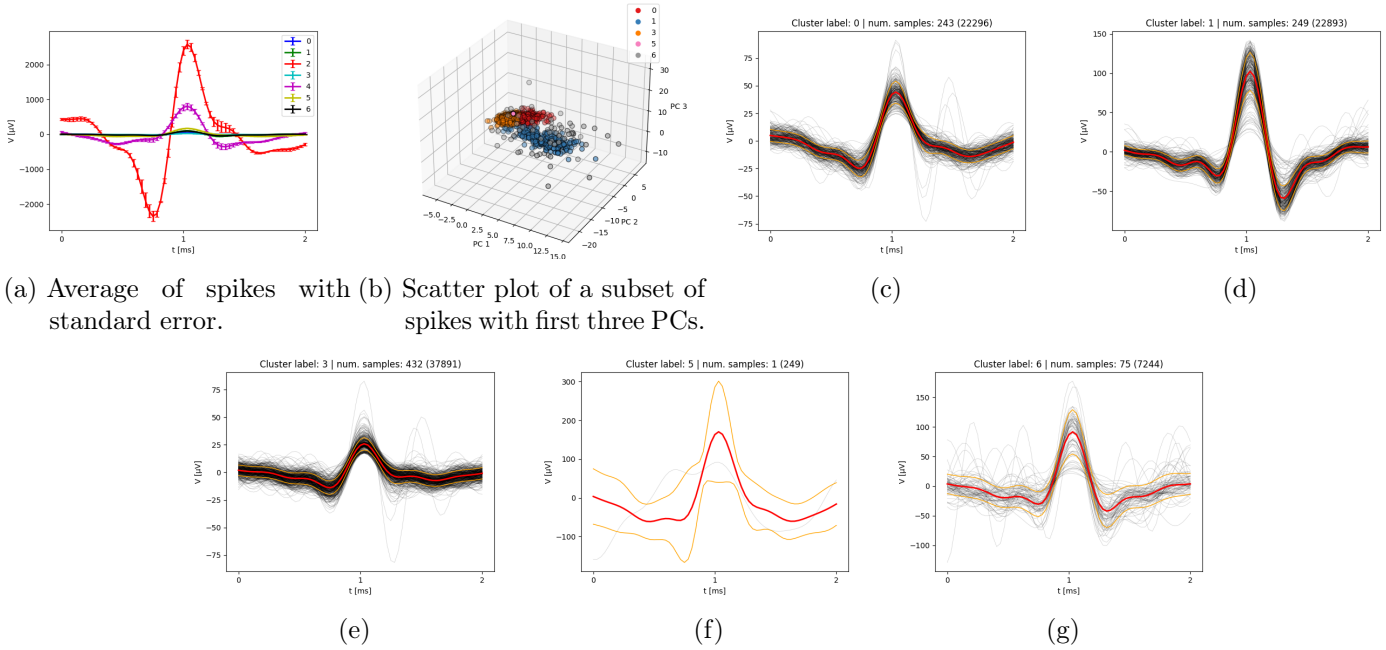
(e)

(f)

(g)

(h)

Figure A.8.: Clusters of spikes of patient 8. Assigned labels, randomly drawn and total number of samples are indicated in the title and legend. Clusters with the following labels were kept: 0, 1, 4.



(a) Average of spikes with standard error.

(b) Scatter plot of a subset of spikes with first three PCs.

(c)

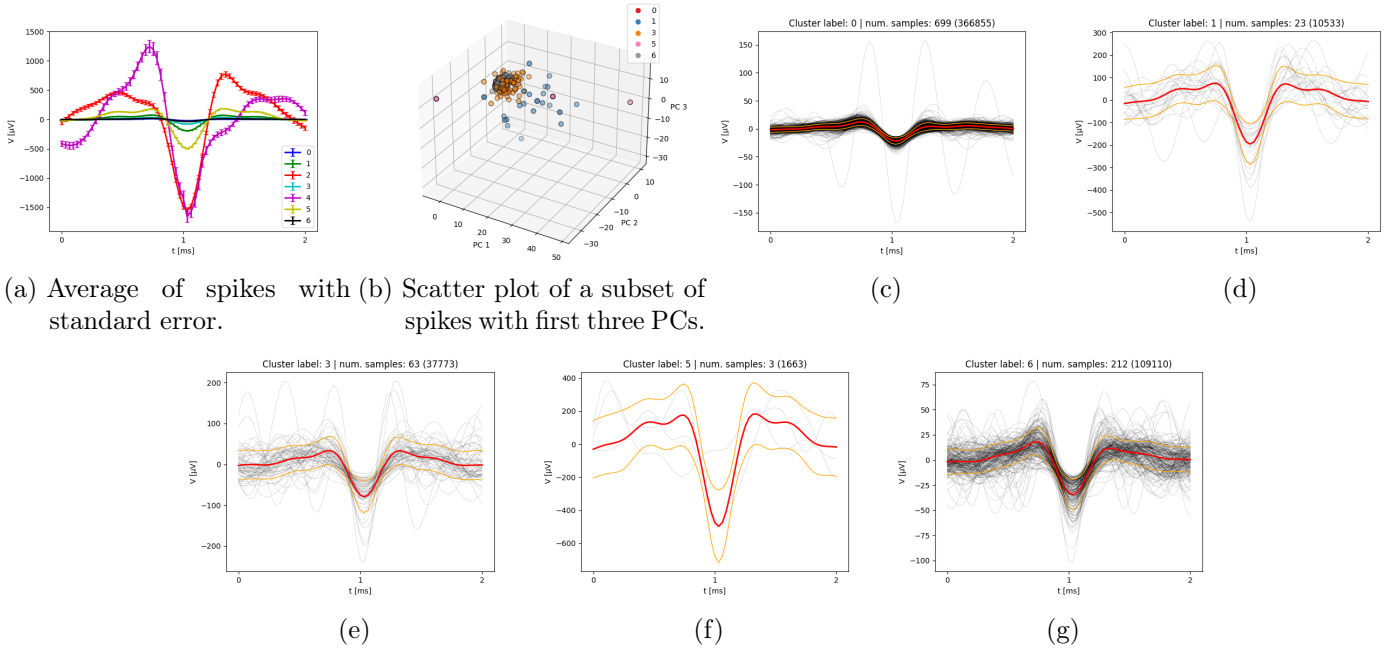
(d)

(e)

(f)

(g)

Figure A.9.: Clusters of spikes of patient 9. Assigned labels, randomly drawn and total number of samples are indicated in the title and legend. Clusters with the following labels were kept: 0, 1, 3.

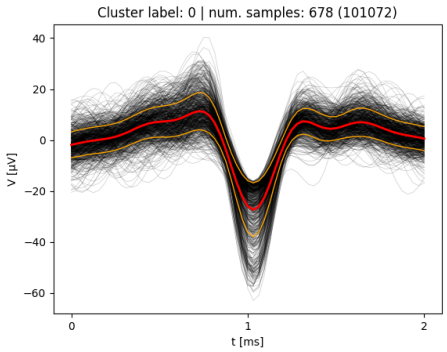


(a) Average of spikes with standard error. (b) Scatter plot of a subset of spikes with first three PCs. (c) (d)

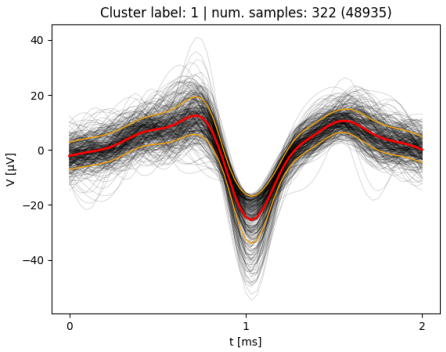
(e) (f) (g)

Figure A.10.: Clusters of spikes of patient 10. Assigned labels, randomly drawn and total number of samples are indicated in the title and legend. Clusters with the following labels were kept: 0.

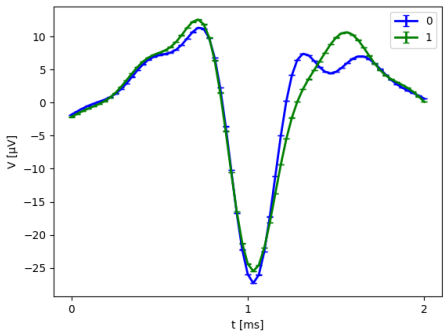
B. Clustering results



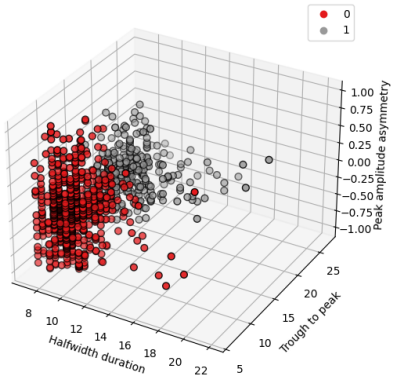
(a) Putative interneurons (label 0).



(b) Putative pyramidal cells (label 1).

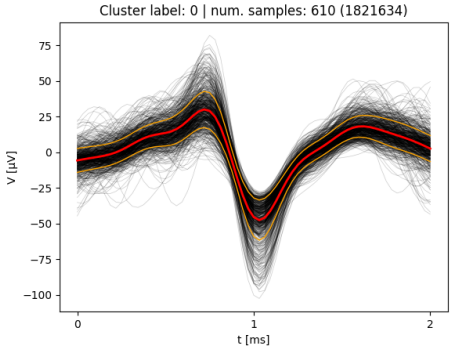


(c) Average of spikes with standard error.

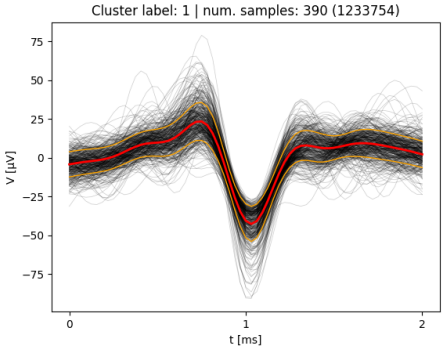


(d) Scatter plot of a subset of spikes represented by three features.

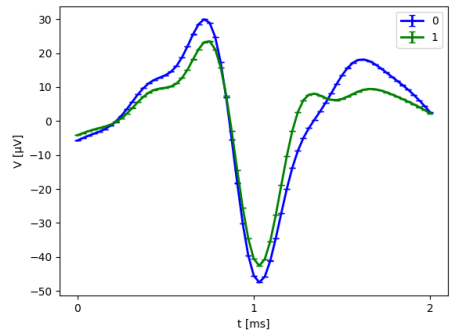
Figure B.1.: Clusters of the two neuronal types of patient 1. a-b) Assigned labels, randomly drawn and total number of samples are indicated in the title. Red curve: average of spikes, orange curve: one standard deviation above and below the mean.



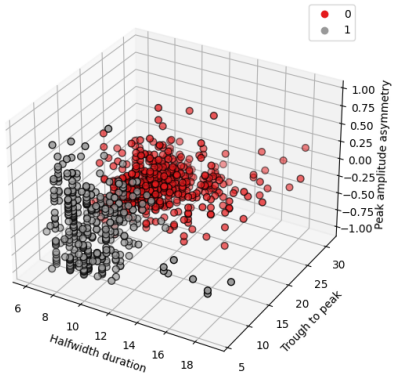
(a) Putative pyramidal cells (label 0).



(b) Putative interneurons (label 1).

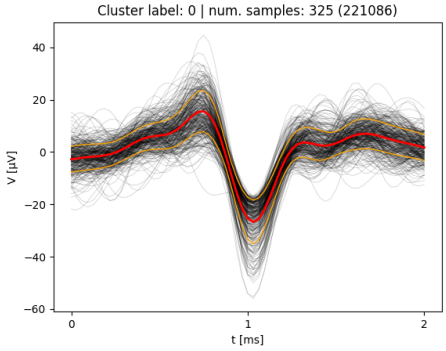


(c) Average of spikes with standard error.

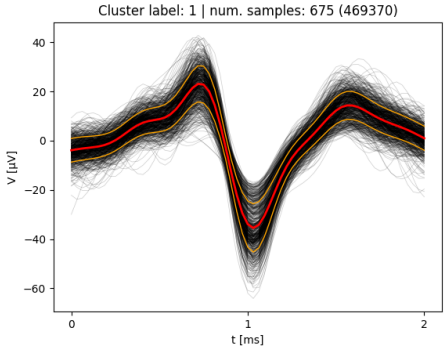


(d) Scatter plot of a subset of spikes represented by three features.

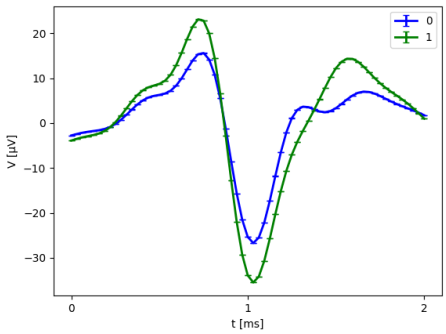
Figure B.2.: Clusters of the two neuronal types of patient 2. a-b) Assigned labels, randomly drawn and total number of samples are indicated in the title. Red curve: average of spikes, orange curve: one standard deviation above and below the mean.



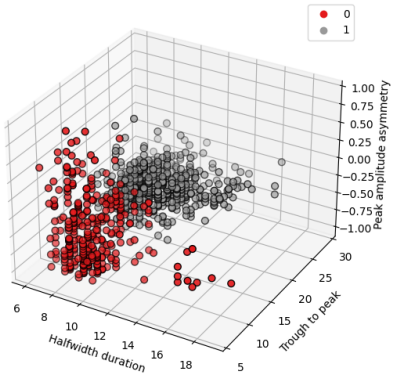
(a) Putative interneurons (label 0).



(b) Putative pyramidal cells (label 1).

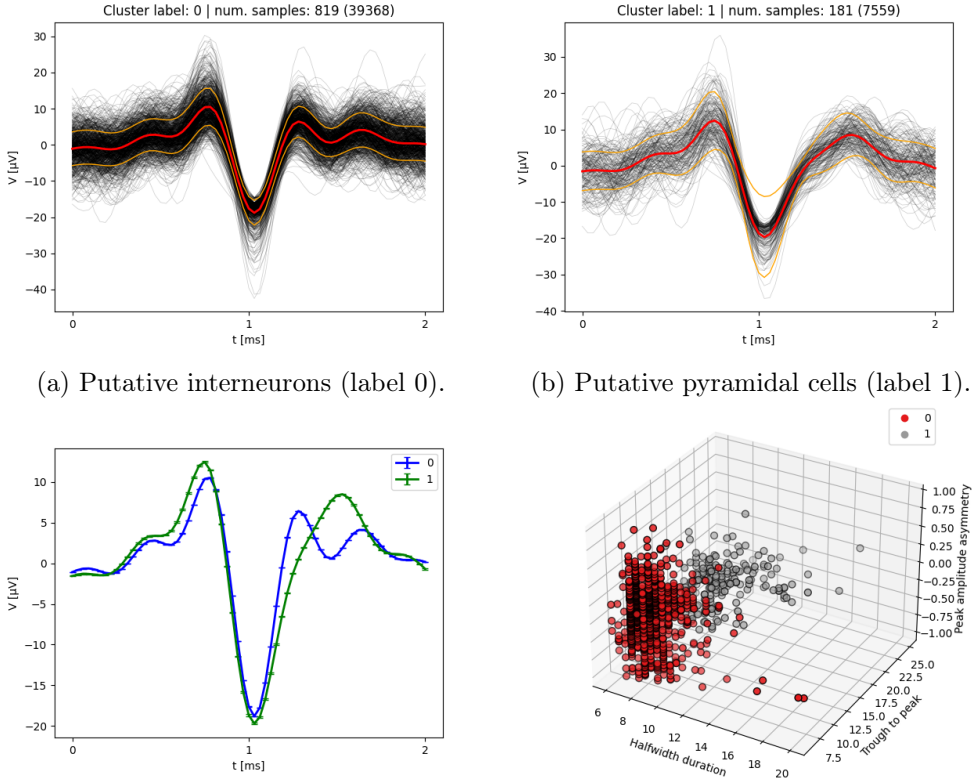


(c) Average of spikes with standard error.



(d) Scatter plot of a subset of spikes represented by three features.

Figure B.3.: Clusters of the two neuronal types of patient 3. a-b) Assigned labels, randomly drawn and total number of samples are indicated in the title. Red curve: average of spikes, orange curve: one standard deviation above and below the mean.



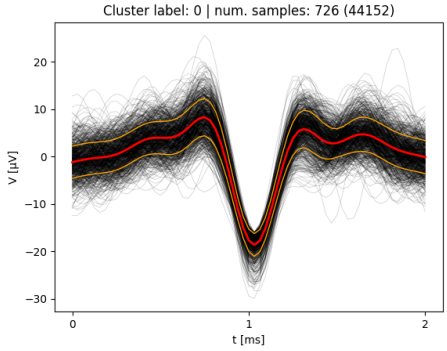
(a) Putative interneurons (label 0).

(b) Putative pyramidal cells (label 1).

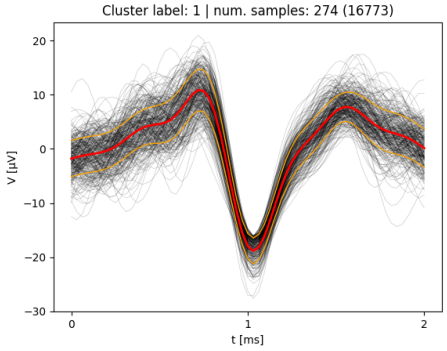
(c) Average of spikes with standard error.

(d) Scatter plot of a subset of spikes represented by three features.

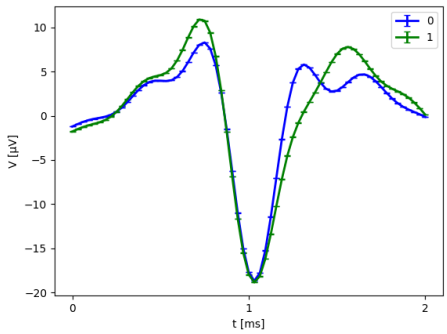
Figure B.4.: Clusters of the two neuronal types of patient 4. a-b) Assigned labels, randomly drawn and total number of samples are indicated in the title. Red curve: average of spikes, orange curve: one standard deviation above and below the mean.



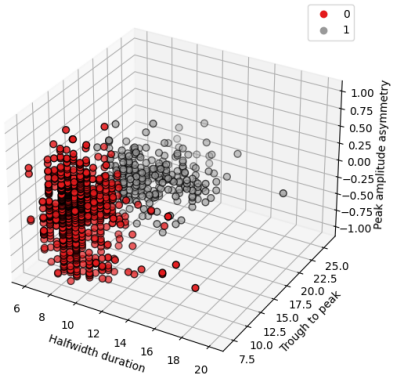
(a) Putative interneurons (label 0).



(b) Putative pyramidal cells (label 1).

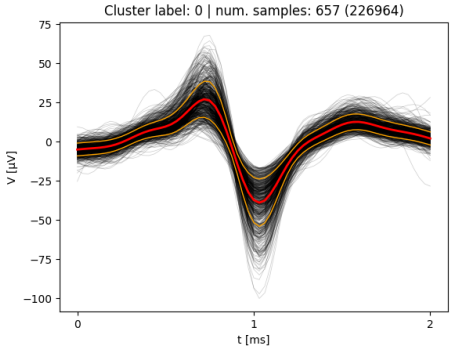


(c) Average of spikes with standard error.

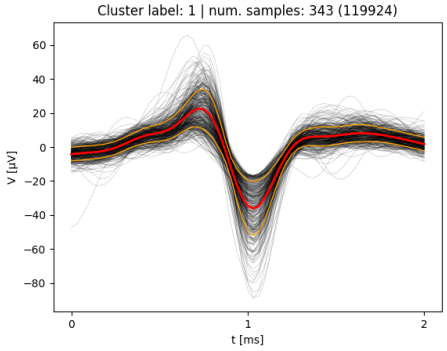


(d) Scatter plot of a subset of spikes represented by three features.

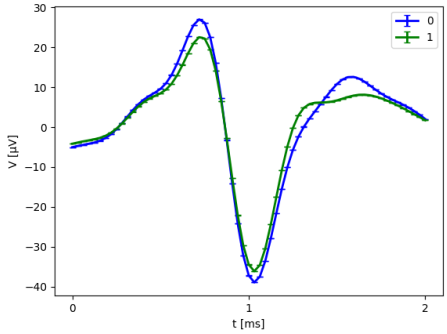
Figure B.5.: Clusters of the two neuronal types of patient 5. a-b) Assigned labels, randomly drawn and total number of samples are indicated in the title. Red curve: average of spikes, orange curve: one standard deviation above and below the mean.



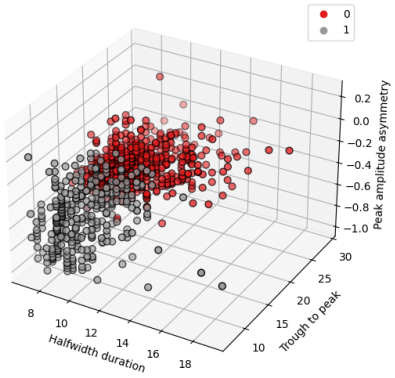
(a) Putative pyramidal cells (label 0).



(b) Putative interneurons (label 1).

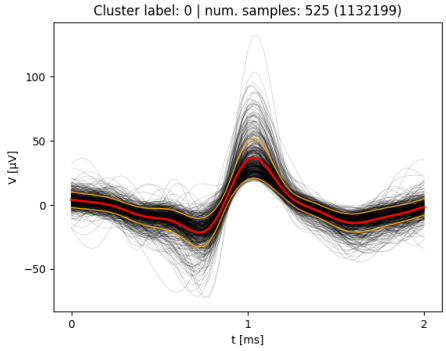


(c) Average of spikes with standard error.

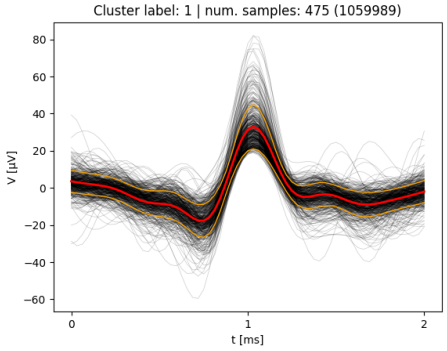


(d) Scatter plot of a subset of spikes represented by three features.

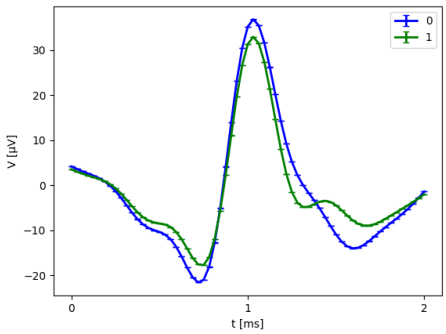
Figure B.6.: Clusters of the two neuronal types of patient 6. a-b) Assigned labels, randomly drawn and total number of samples are indicated in the title. Red curve: average of spikes, orange curve: one standard deviation above and below the mean.



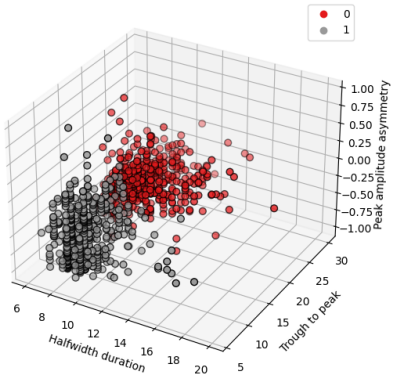
(a) Putative pyramidal cells (label 0).



(b) Putative interneurons (label 1).

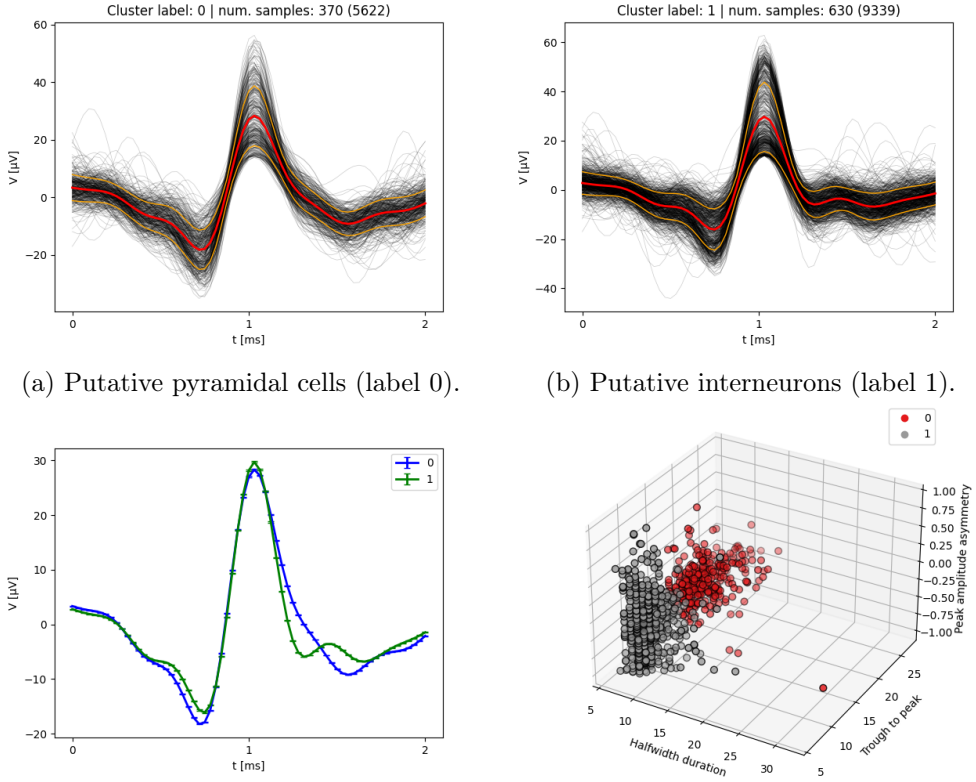


(c) Average of spikes with standard error.



(d) Scatter plot of a subset of spikes represented by three features.

Figure B.7.: Clusters of the two neuronal types of patient 7. a-b) Assigned labels, randomly drawn and total number of samples are indicated in the title. Red curve: average of spikes, orange curve: one standard deviation above and below the mean.



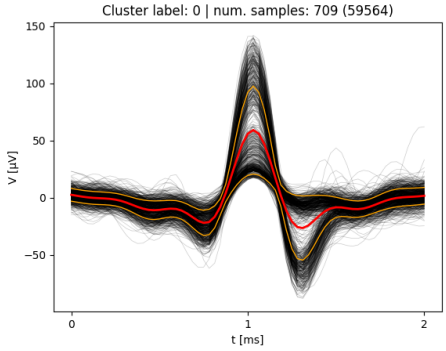
(a) Putative pyramidal cells (label 0).

(b) Putative interneurons (label 1).

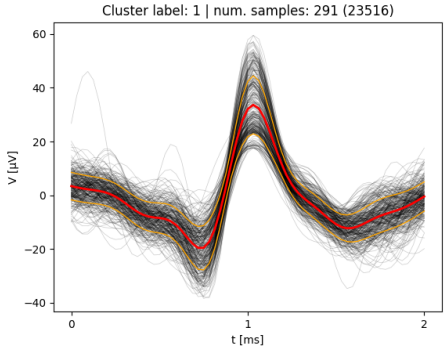
(c) Average of spikes with standard error.

(d) Scatter plot of a subset of spikes represented by three features.

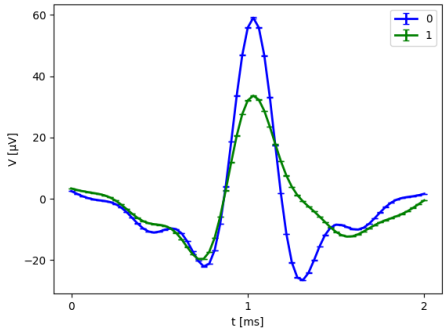
Figure B.8.: Clusters of the two neuronal types of patient 8. a-b) Assigned labels, randomly drawn and total number of samples are indicated in the title. Red curve: average of spikes, orange curve: one standard deviation above and below the mean.



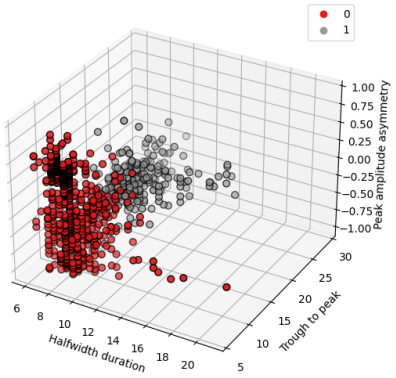
(a) Putative interneurons (label 0).



(b) Putative pyramidal cells (label 1).

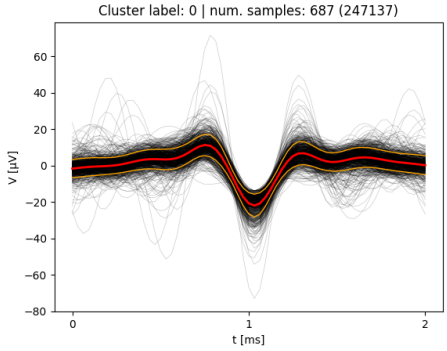


(c) Average of spikes with standard error.

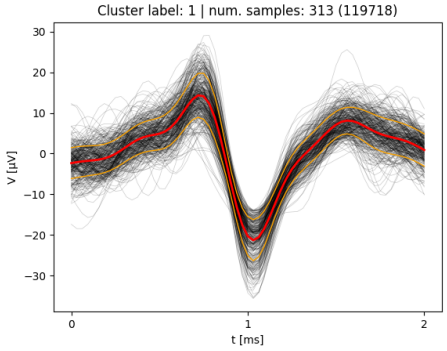


(d) Scatter plot of a subset of spikes represented by three features.

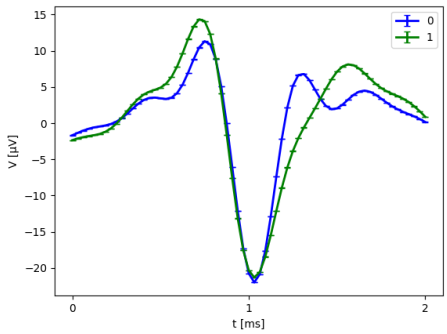
Figure B.9.: Clusters of the two neuronal types of patient 9. a-b) Assigned labels, randomly drawn and total number of samples are indicated in the title. Red curve: average of spikes, orange curve: one standard deviation above and below the mean.



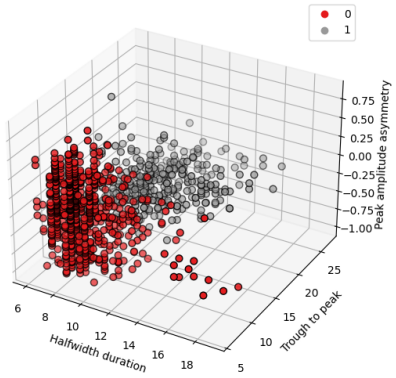
(a) Putative interneurons (label 0).



(b) Putative pyramidal cells (label 1).



(c) Average of spikes with standard error.



(d) Scatter plot of a subset of spikes represented by three features.

Figure B.10.: Clusters of the two neuronal types of patient 10. a-b) Assigned labels, randomly drawn and total number of samples are indicated in the title. Red curve: average of spikes, orange curve: one standard deviation above and below the mean.

C. Benchmark results

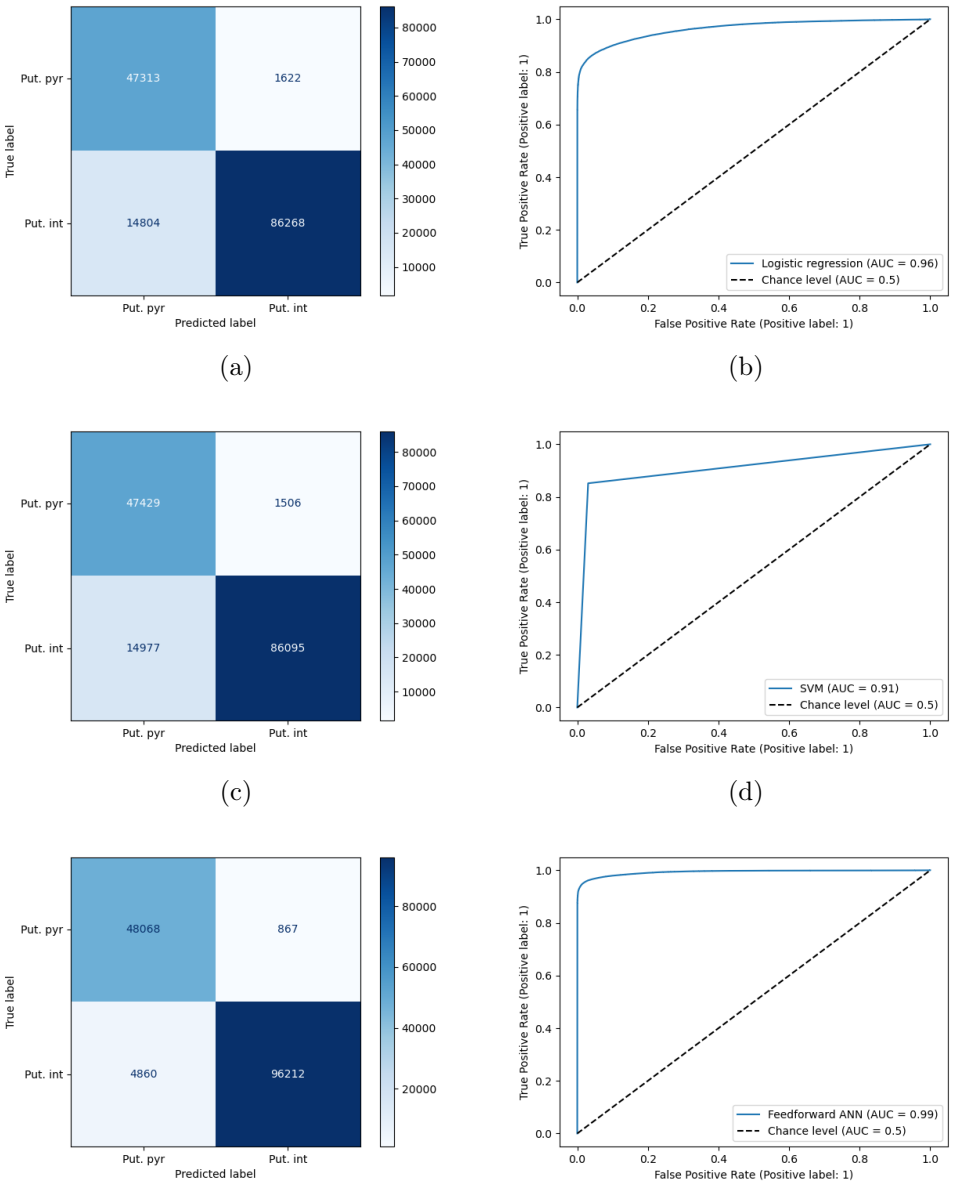
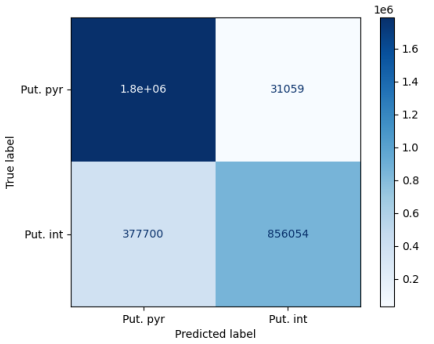
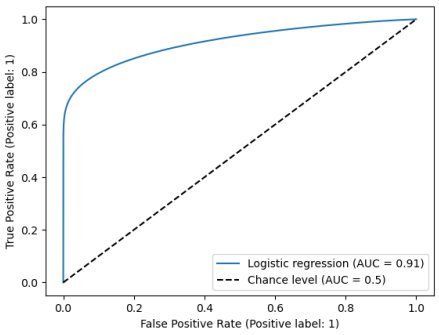


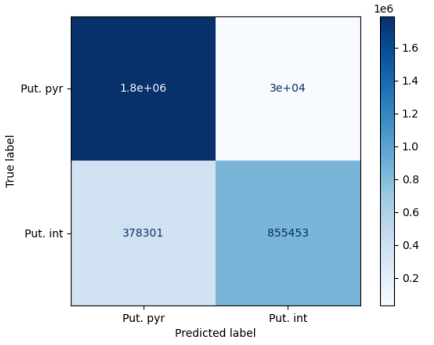
Figure C.1.: Visualization of the benchmark for patient 1.



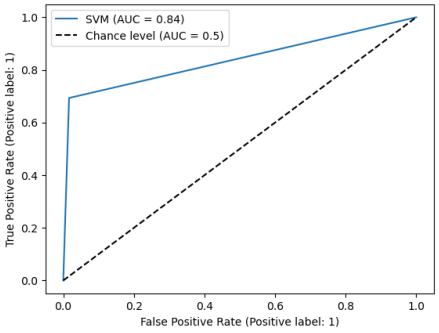
(a)



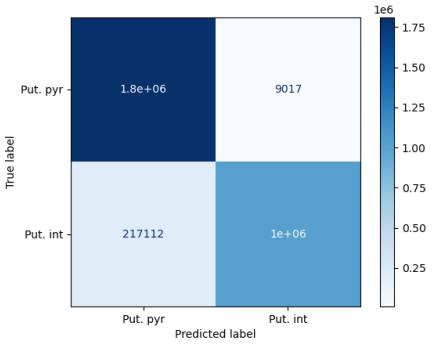
(b)



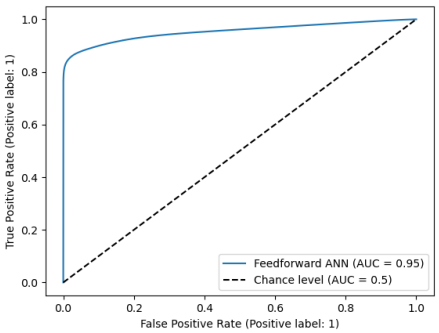
(c)



(d)



(e)



(f)

Figure C.2.: Visualization of the benchmark for patient 2.

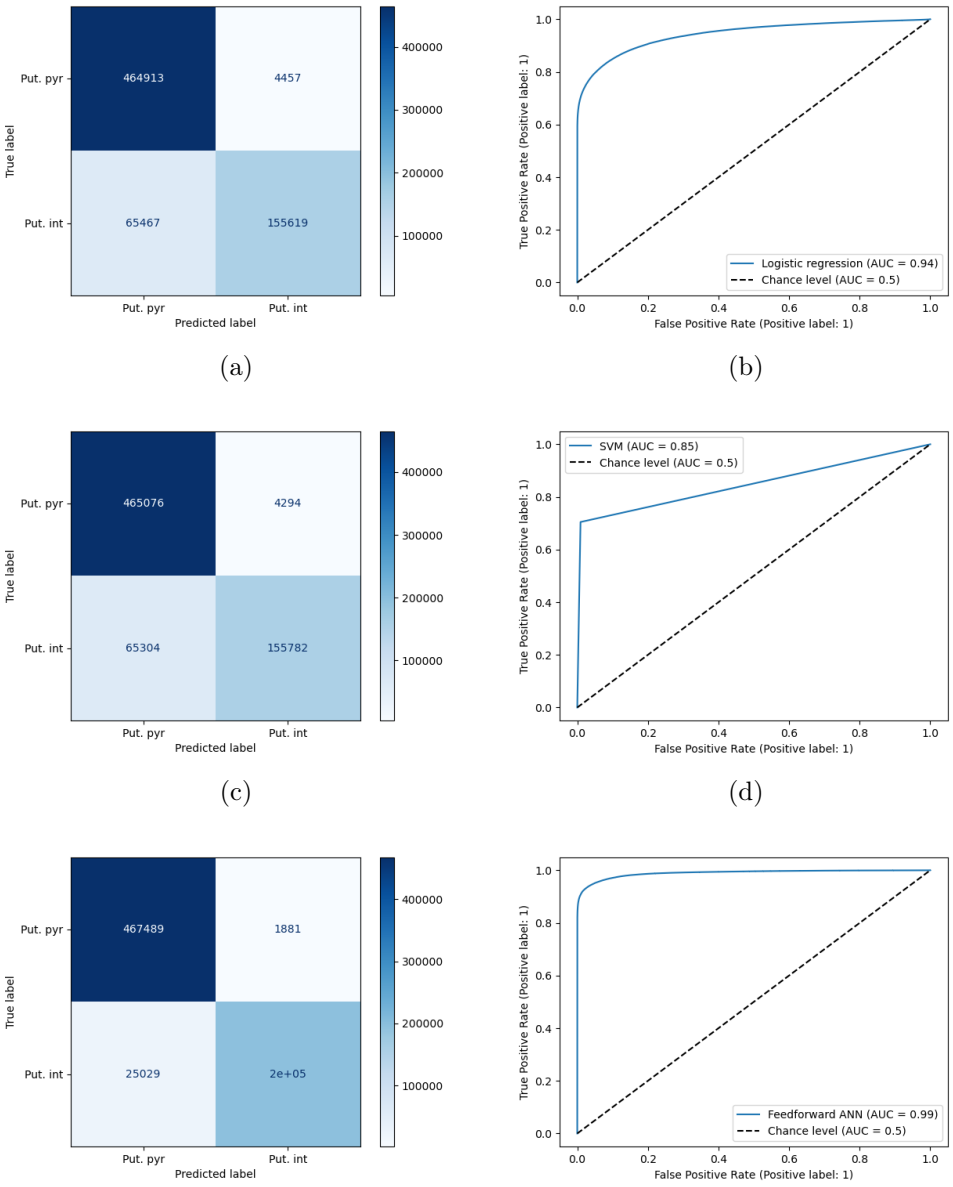


Figure C.3.: Visualization of the benchmark for patient 3.

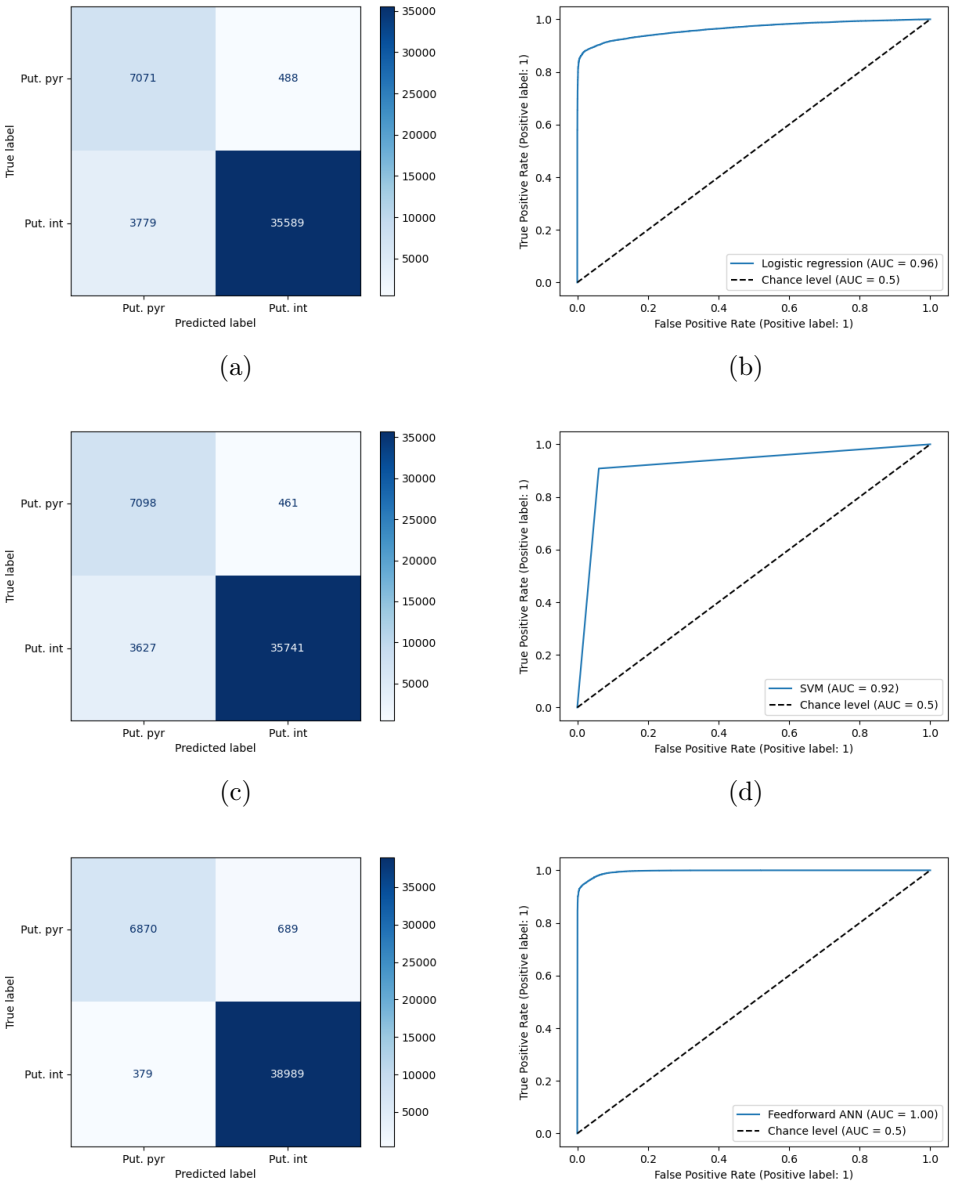


Figure C.4.: Visualization of the benchmark for patient 4.

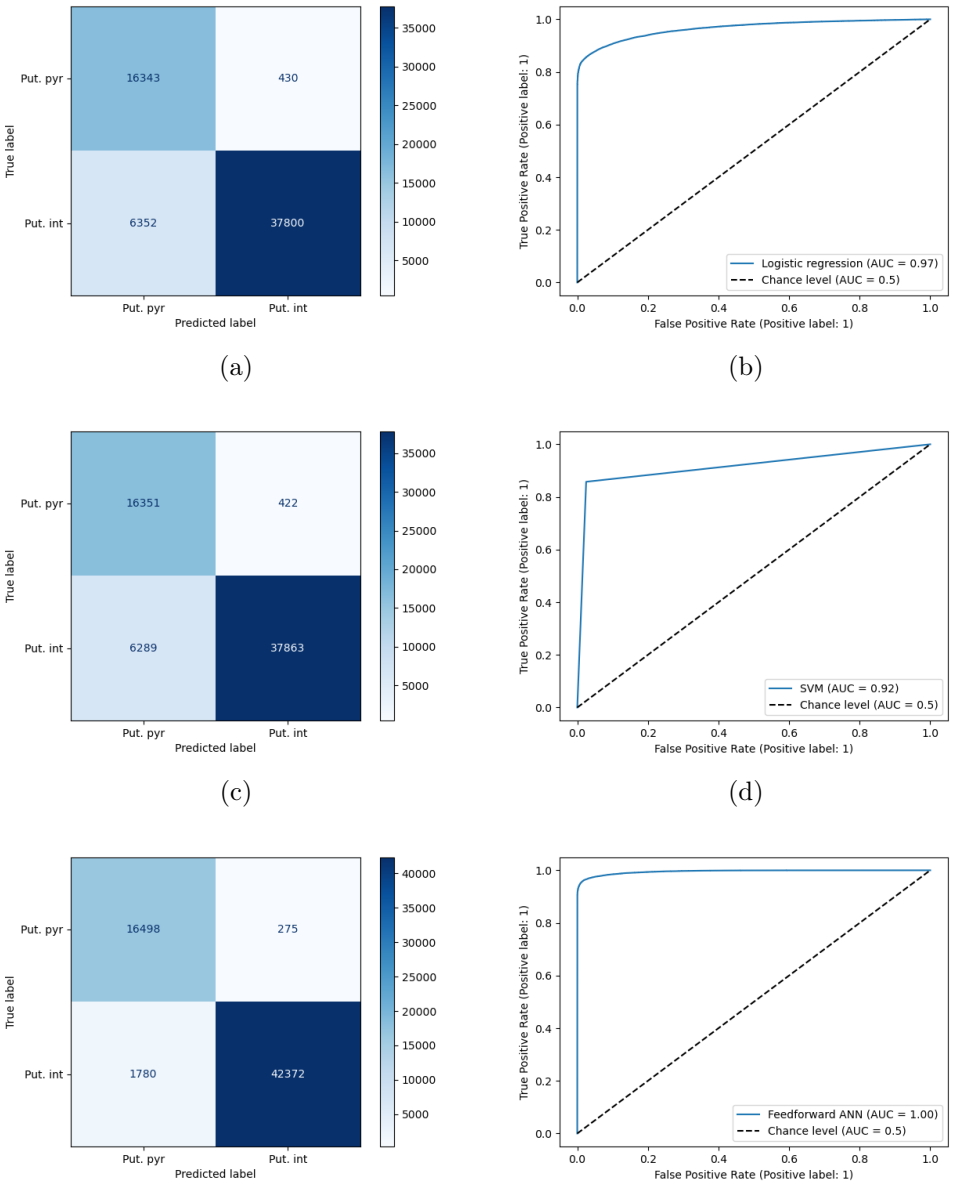
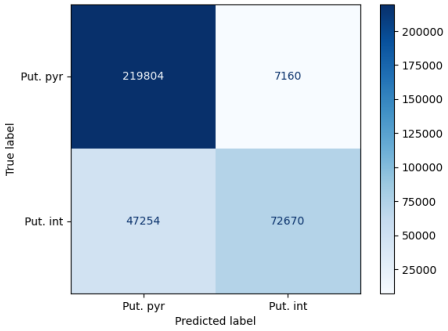
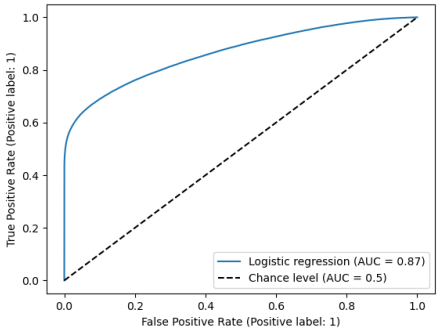


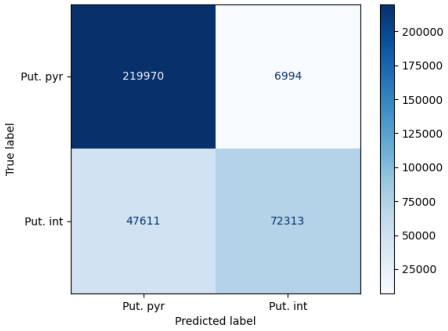
Figure C.5.: Visualization of the benchmark for patient 5.



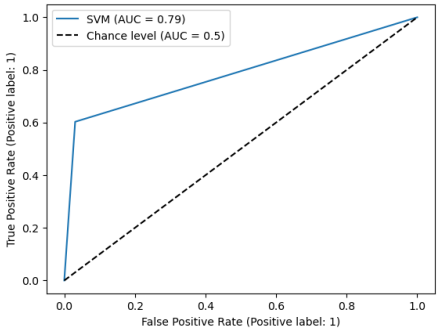
(a)



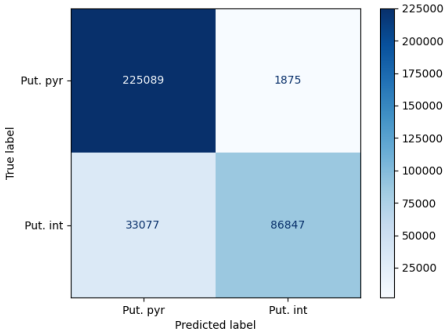
(b)



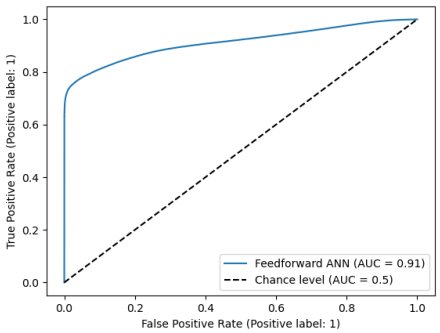
(c)



(d)



(e)



(f)

Figure C.6.: Visualization of the benchmark for patient 6.

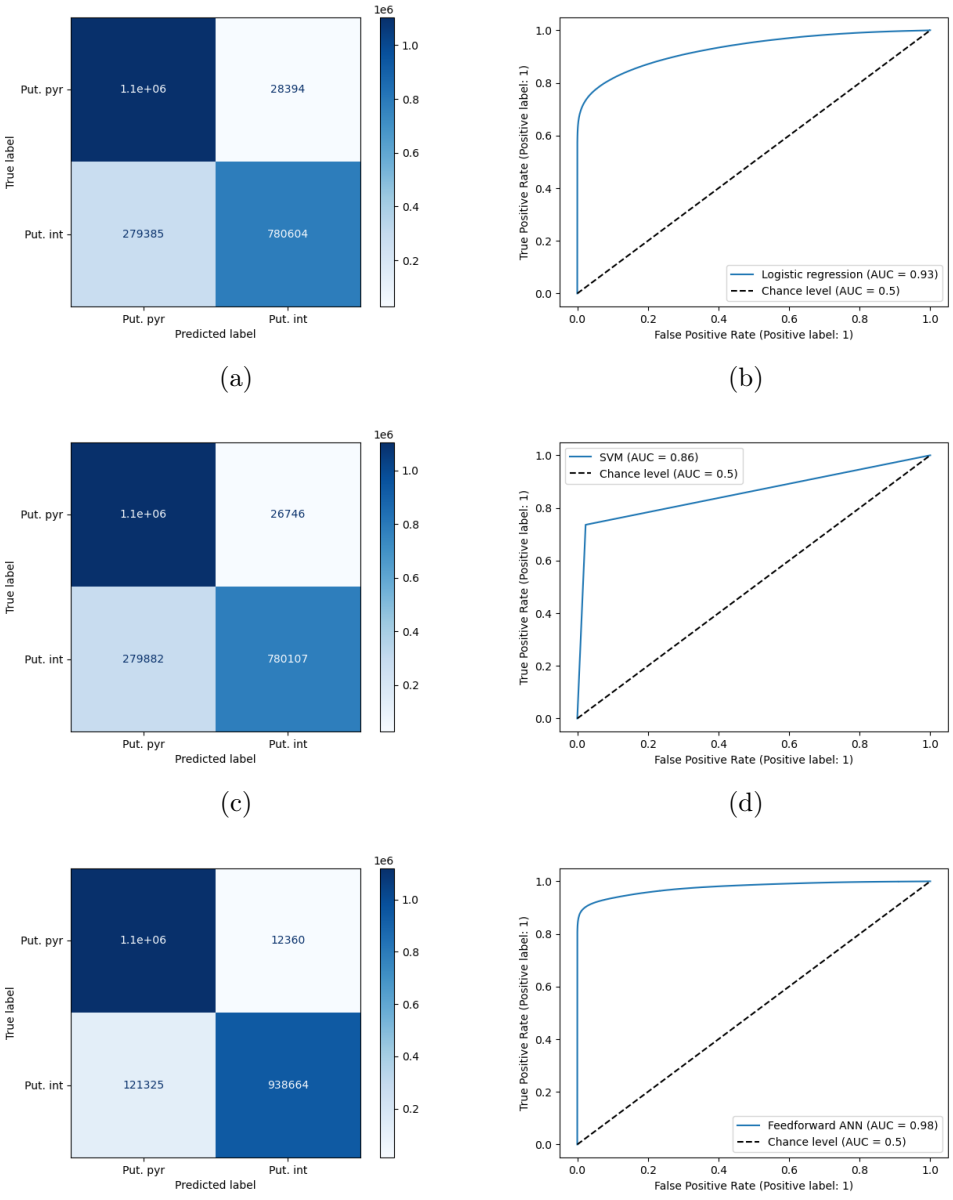
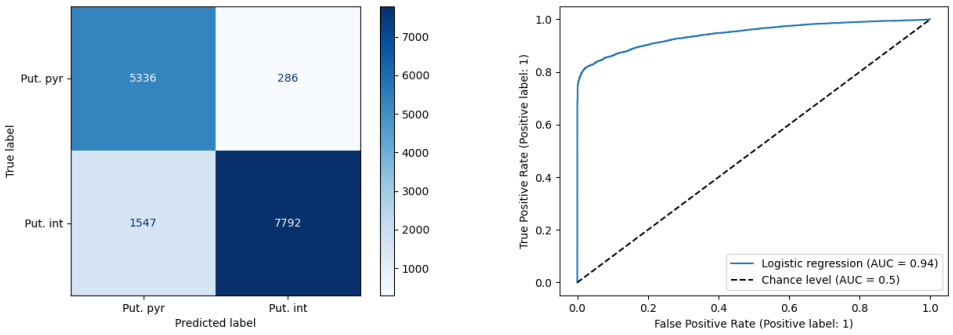
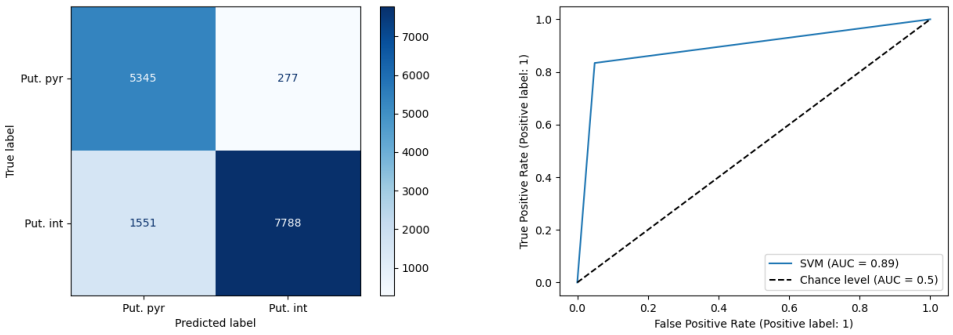


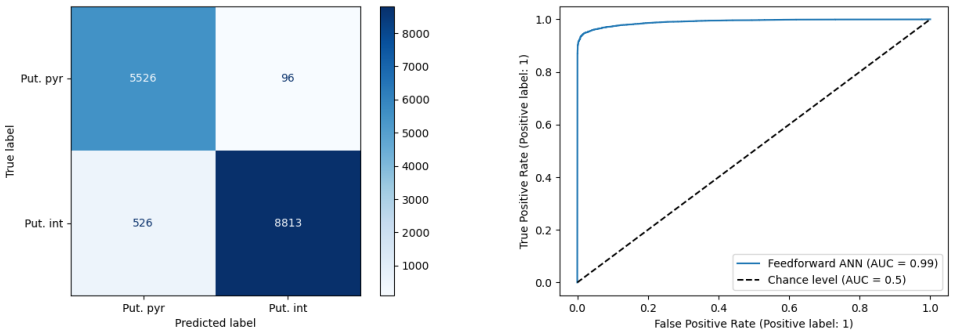
Figure C.7.: Visualization of the benchmark for patient 7.



(a) (b)



(c) (d)



(e) (f)

Figure C.8.: Visualization of the benchmark for patient 8.

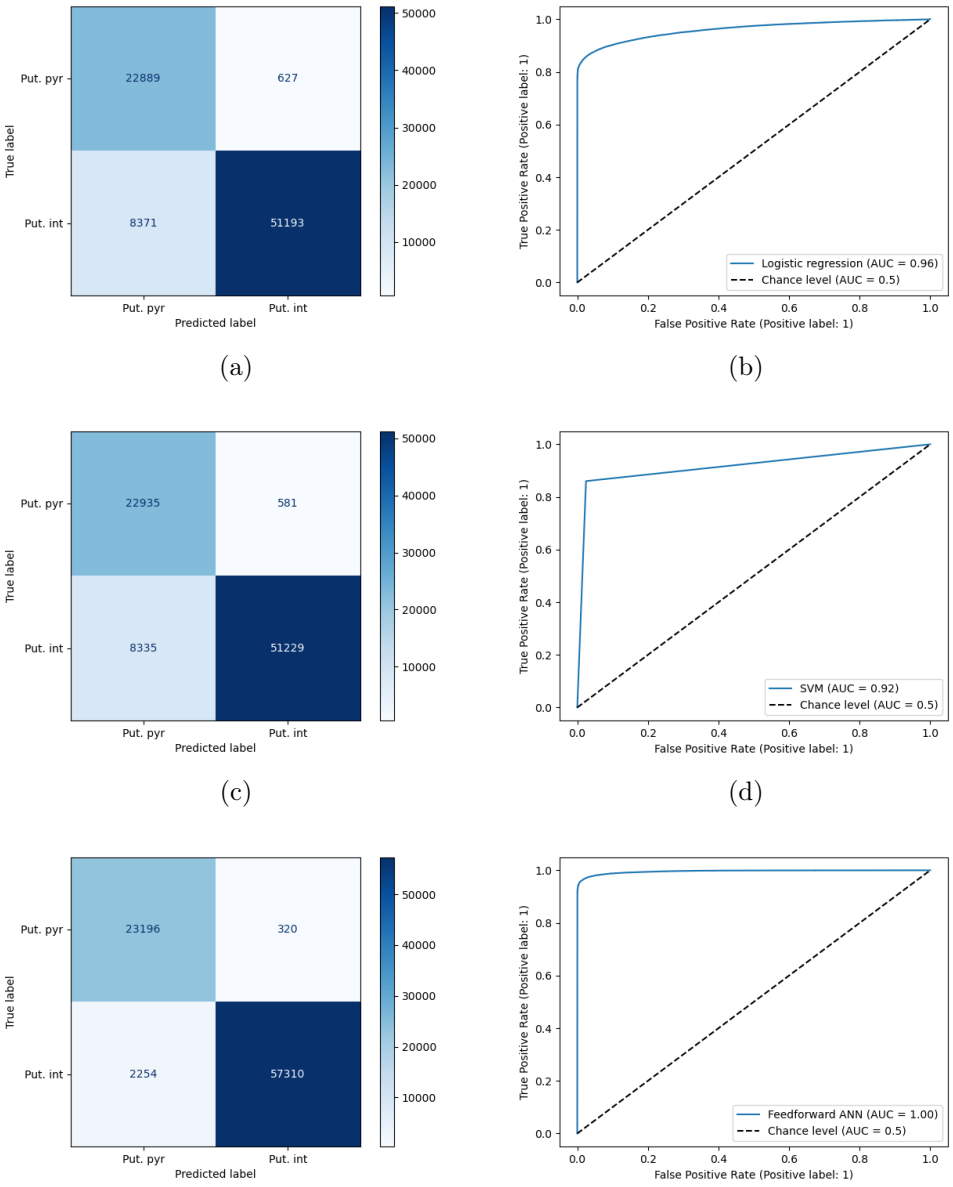
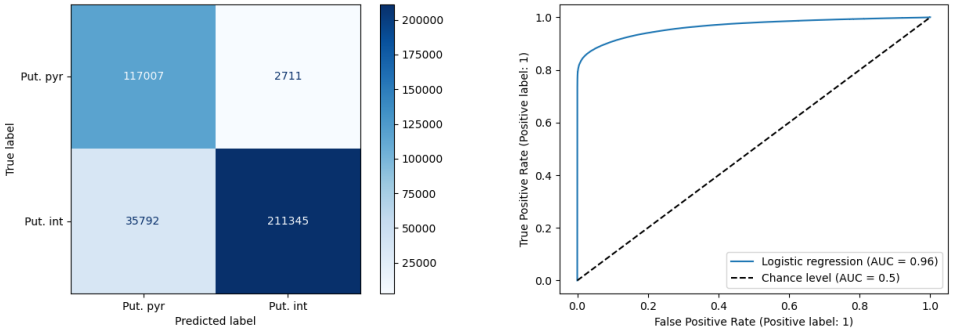
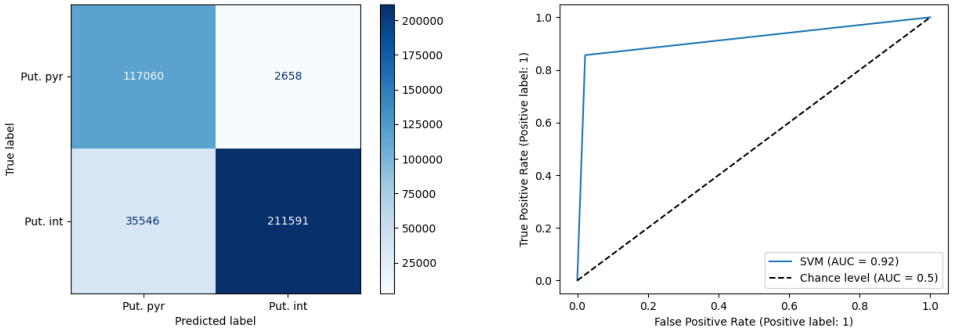


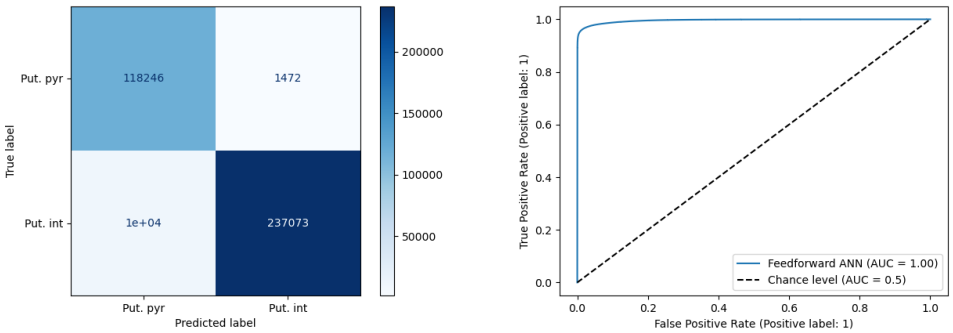
Figure C.9.: Visualization of the benchmark for patient 9.



(a) (b)



(c) (d)



(e) (f)

Figure C.10.: Visualization of the benchmark for patient 10.

D. IED-SUA correlation results

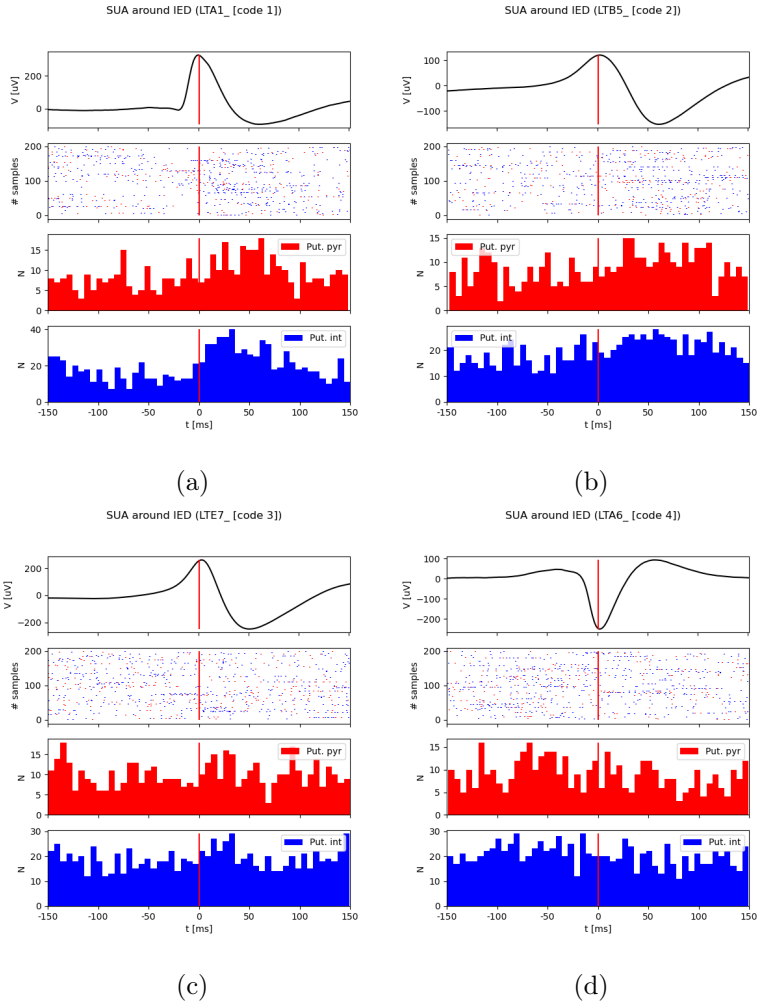


Figure D.1.: Event plots of patient 1 depicting SUA around IED samples.

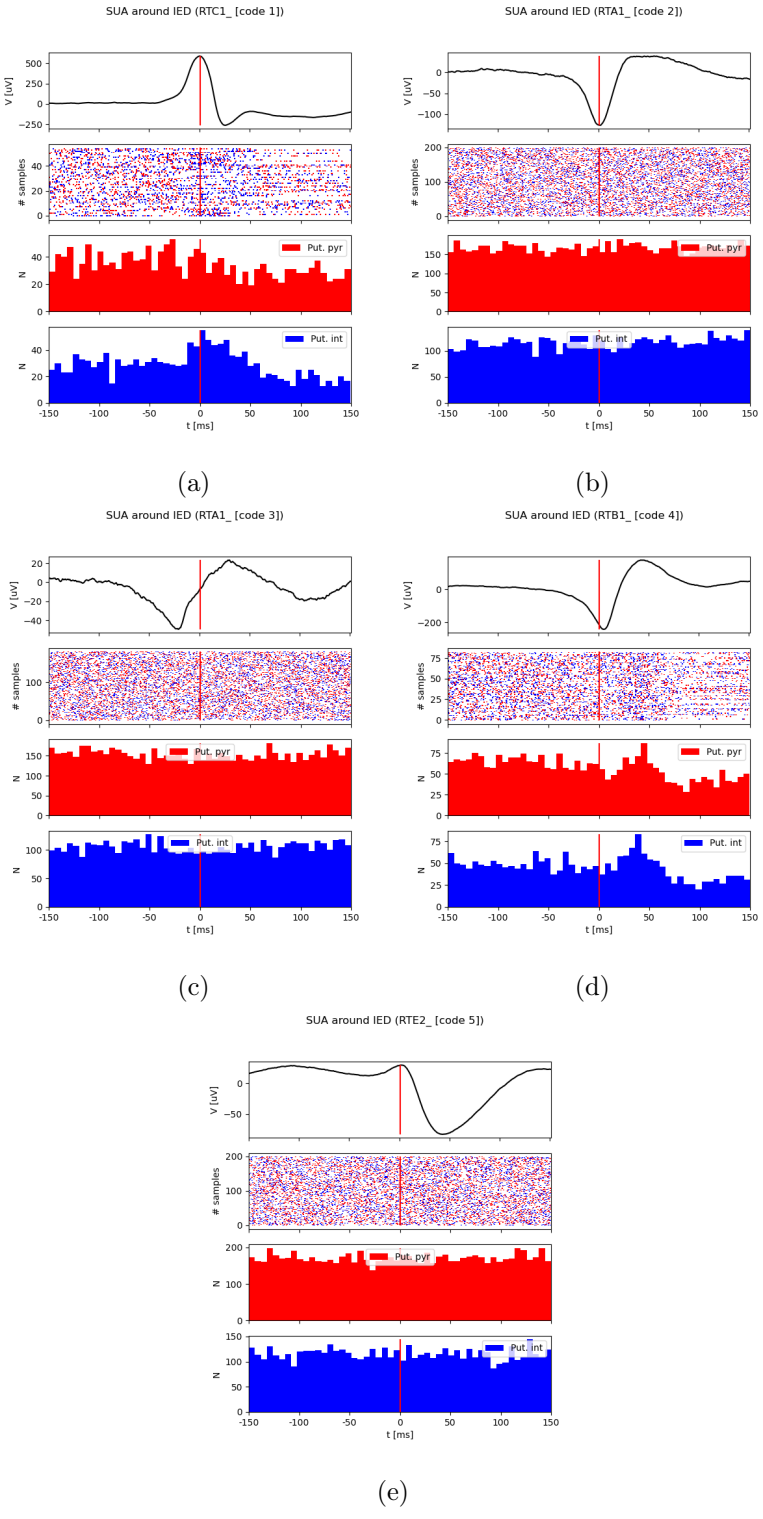


Figure D.2.: Event plots of patient 2 depicting SUA around IED samples.

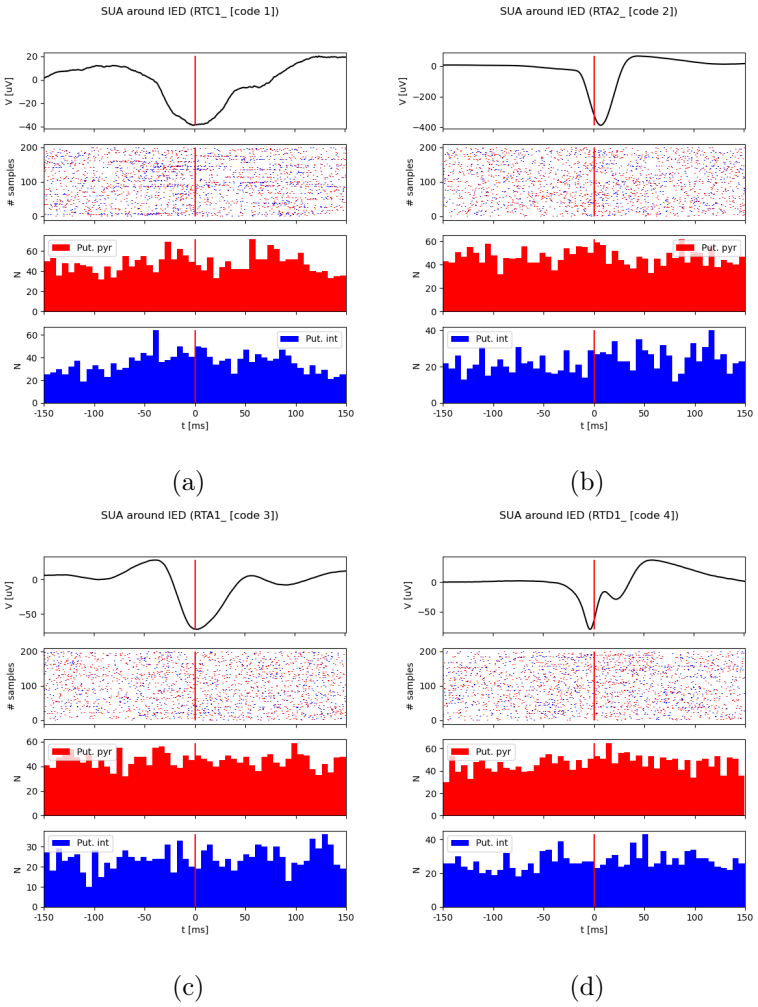


Figure D.3.: Event plots of patient 3 depicting SUA around IED samples.

D. IED-SUA correlation results

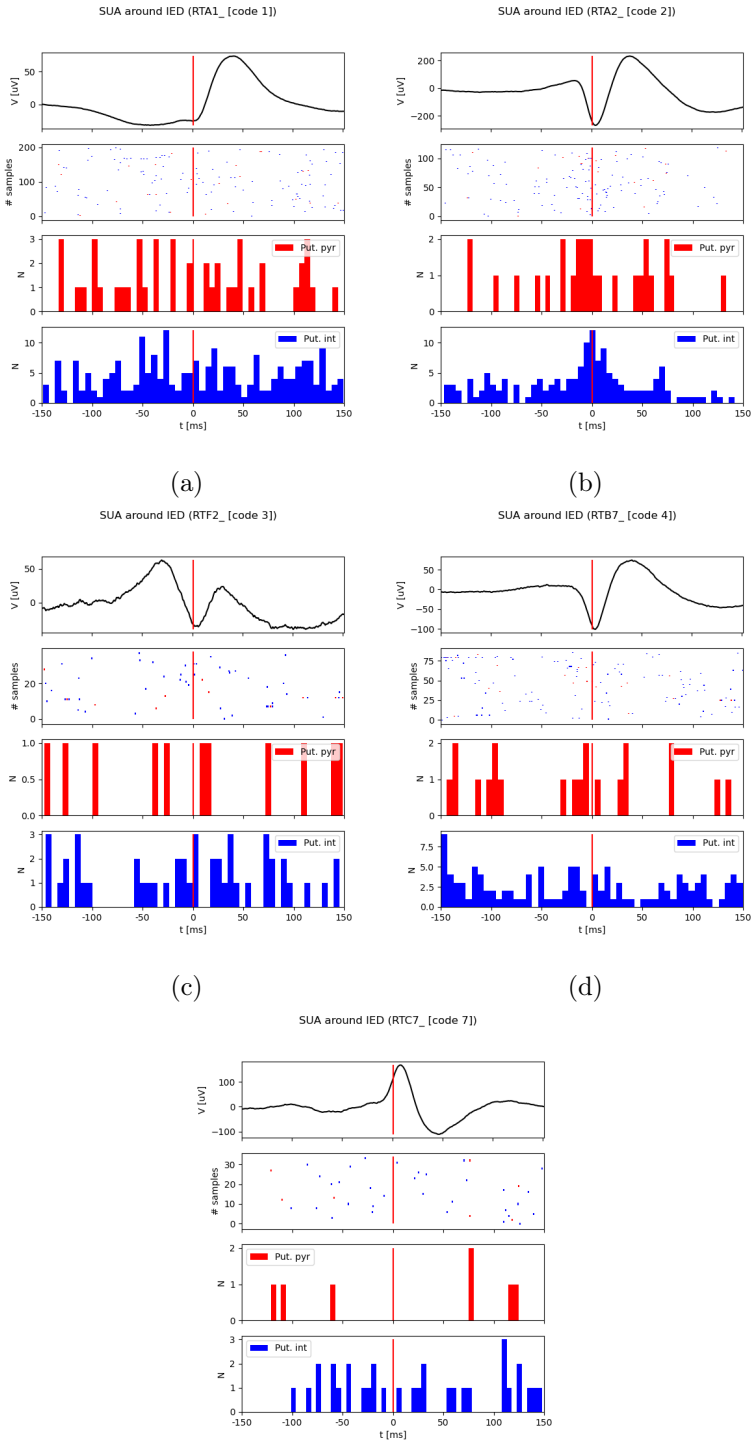


Figure D.4.: Event plots of patient 4 depicting SUA around IED samples.

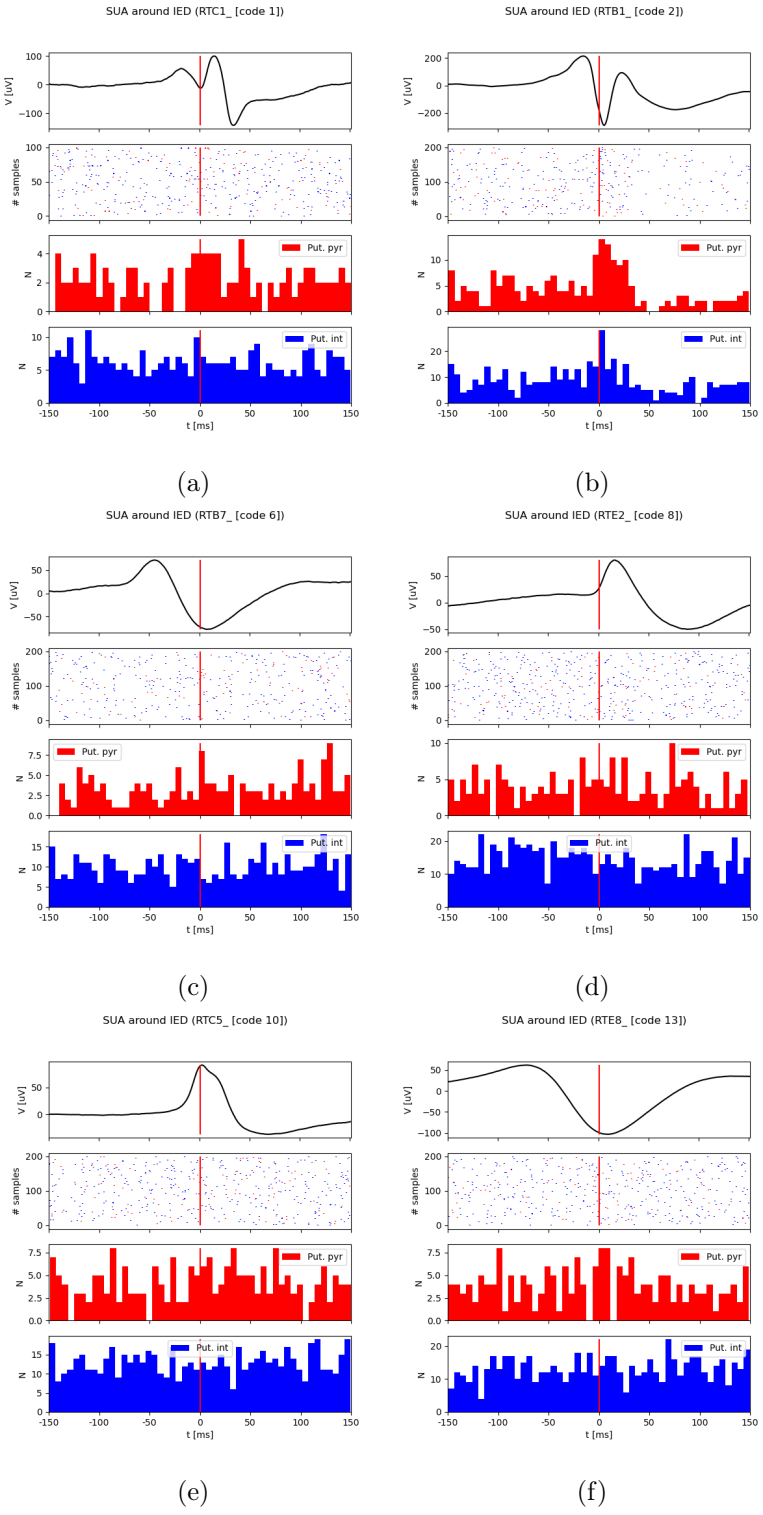
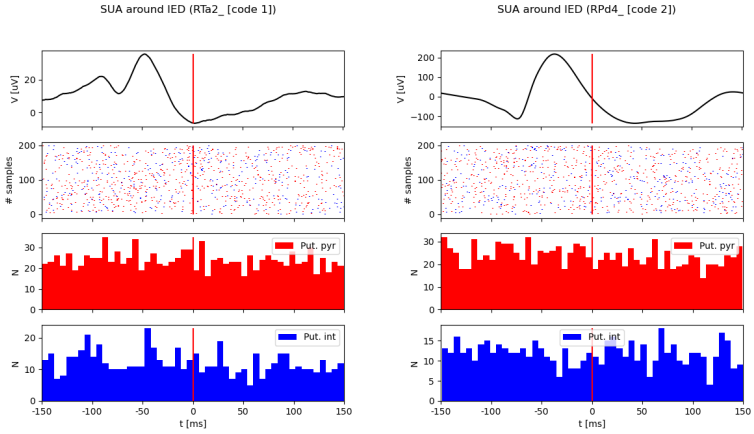
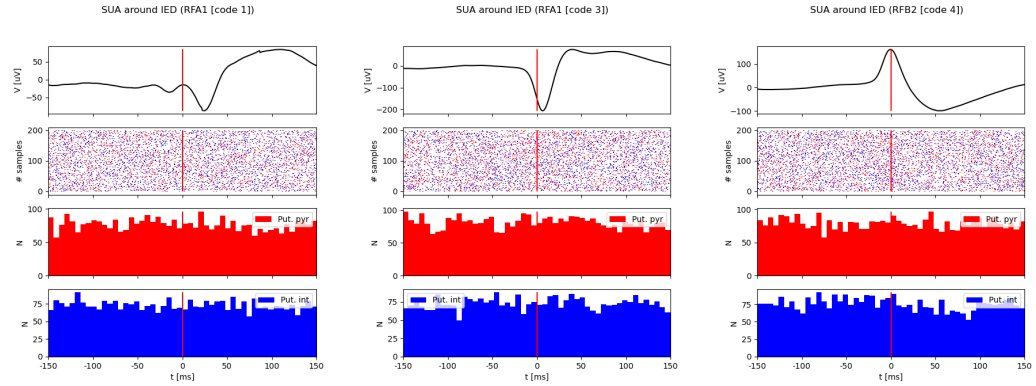


Figure D.5.: Event plots of patient 5 depicting SUA around IED samples.



(a) (b)

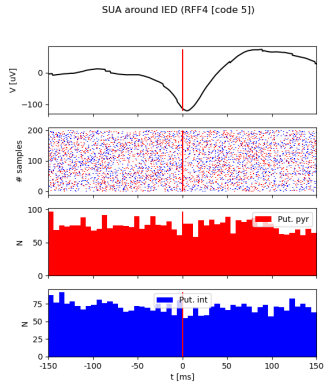
Figure D.6.: Event plots of patient 6 depicting SUA around IED samples.



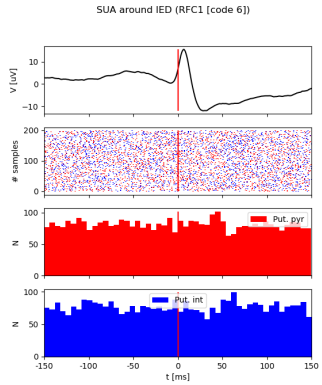
(a)

(b)

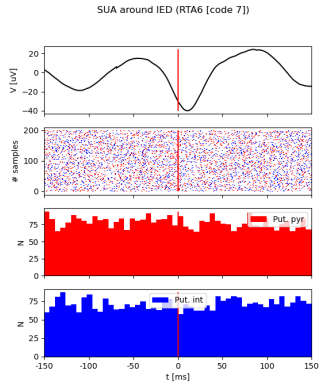
(c)



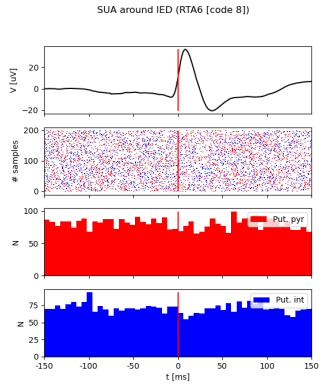
(d)



(e)

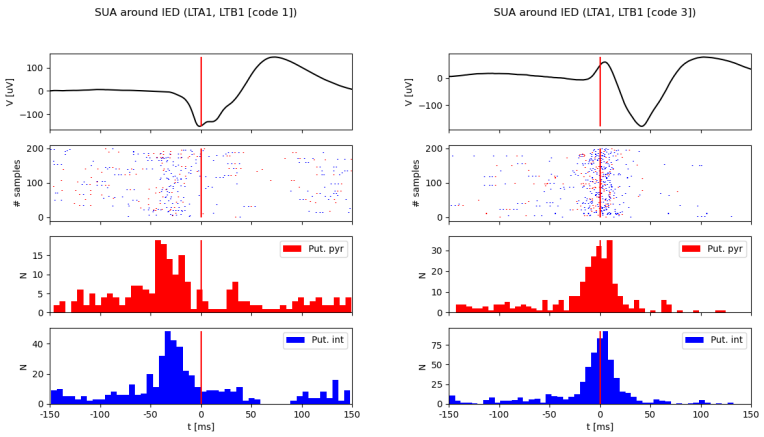


(f)



(g)

Figure D.7.: Event plots of patient 7 depicting SUA around IED samples.



(a) (b)

Figure D.8.: Event plots of patient 8 depicting SUA around IED samples.

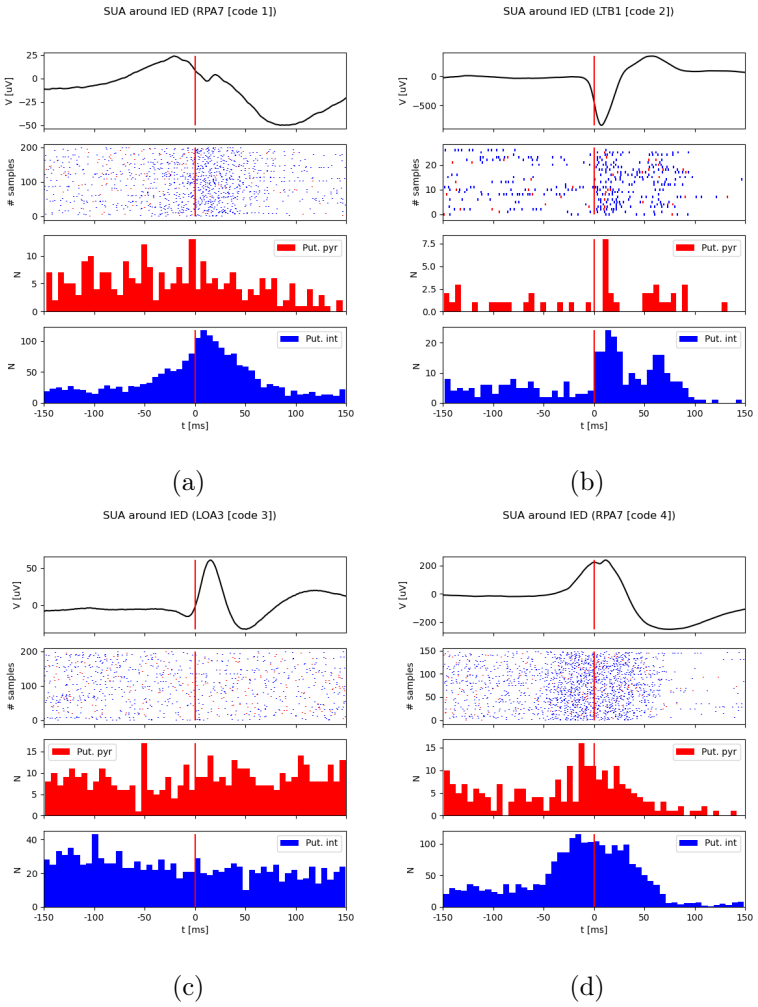


Figure D.9.: Event plots of patient 9 depicting SUA around IED samples.

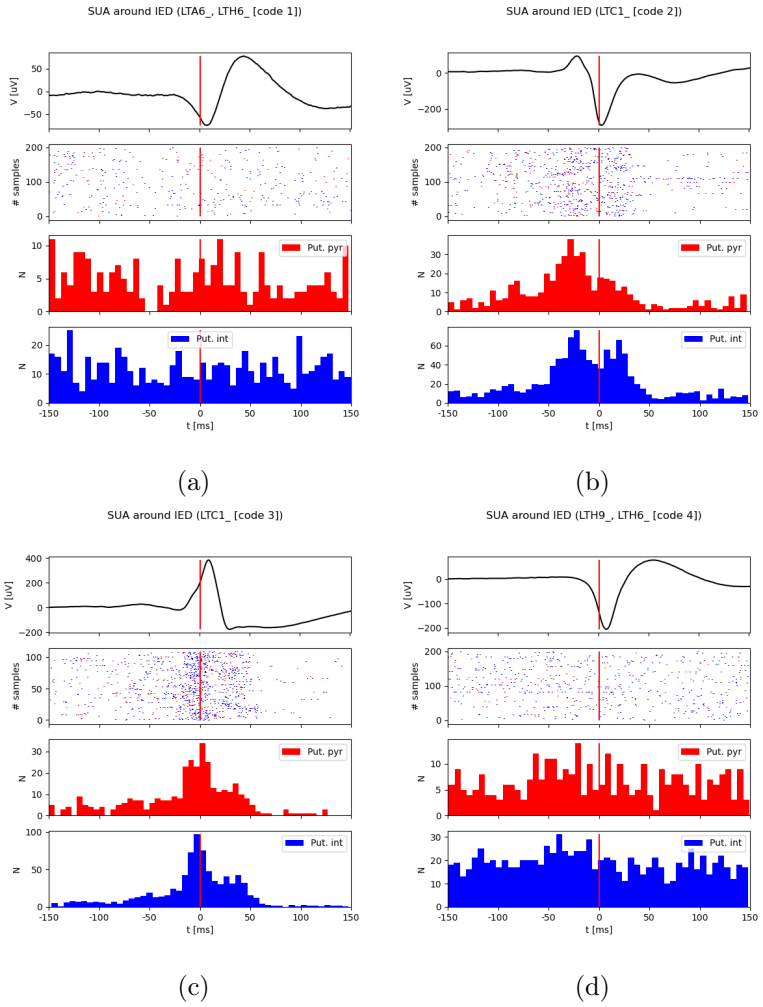


Figure D.10.: Event plots of patient 10 depicting SUA around IED samples.