

# **MASTER THESIS**

Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Engineering at the University of Applied Sciences Technikum Wien - Degree Program Medical Engineering & eHealth

## **Unsupervised Deep Learning Approach for Seizure Onset Zone Localisation in Epilepsy**

By: Tereza Přidalová, Bc.

Student Number: 2110228028

Supervisors: FH-Prof. DI Dr. Lars Mehnen  
Ing. Mgr. Jan Cimbálník, Ph.D.

Vienna, May 18, 2022

# Declaration

“As author and creator of this work to hand, I confirm with my signature knowledge of the relevant copyright regulations governed by higher education acts (see Urheberrechtsgesetz /Austrian copyright law as amended as well as the Statute on Studies Act Provisions / Examination Regulations of the UAS Technikum Wien as amended).

I hereby declare that I completed the present work independently and that any ideas, whether written by others or by myself, have been fully sourced and referenced. I am aware of any consequences I may face on the part of the degree program director if there should be evidence of missing autonomy and independence or evidence of any intent to fraudulently achieve a pass mark for this work (see Statute on Studies Act Provisions / Examination Regulations of the UAS Technikum Wien as amended).

I further declare that up to this date I have not published the work to hand nor have I presented it to another examination board in the same or similar form. I affirm that the version submitted matches the version in the upload tool.“

Vienna, May 18, 2022

Signature

# Abstract

Epilepsy affects about 50 million people worldwide, with one-third of patients being drug-resistant and therefore candidates for an invasive brain resection surgery. Brain resection surgery candidates undergo invasive intracranial encephalography (iEEG) monitoring to determine the seizure onset zone (SOZ). Recorded data can span over weeks and need to be manually reviewed by a physician to assess SOZ. This process can be time-consuming and burdensome due to the vast amount of collected data. This work investigates utilisation of a deep autoencoder for unsupervised data exploration and specifically its ability to discriminate between SOZ and non-SOZ (NSOZ) iEEG channels. The data used in this thesis consists of iEEG collected from 33 patients in two institutes (Mayo Clinic, Rochester, Minnesota, USA and St. Anne's University Hospital, Brno, Czech Republic - FNUSA) who underwent invasive pre-surgical monitoring. The autoencoder's capability to discriminate between SOZ and NSOZ was evaluated using a self-learned embedded feature space representation of the autoencoder network. Autoencoder features were compared to previously established biomarkers for SOZ determination. Discrimination capability was evaluated for both autoencoder features and biomarkers using a Naive Bayes classifier and leave-one-out cross-validation. The achieved area under receiver operating characteristic curve (AUROC) was 0.68 for the FNUSA and 0.56 for the Mayo dataset. Performance in discriminating between SOZ and NSOZ electrodes was not significantly different between the investigated autoencoder features and previously established biomarkers. Selecting the better performing classifier for each patient increased the AUROC to 0.75 and 0.64 for the FNUSA and Mayo dataset, respectively. The results suggest that future approaches combining biomarkers and self-learning methods have a potential to improve the SOZ vs NSOZ discrimination capability of unsupervised iEEG exploration systems, and thus to enhance the surgical management of epilepsy.

**Keywords:** epilepsy, seizure onset zone, deep learning, autoencoder, iEEG

# Acknowledgements

I would like to express my deep gratitude to my supervisor Ing. Mgr. Jan Cimbálník, Ph.D., for his time, expert opinions and advice dedicated to my work. His positive attitude toward exploring the scientific issues encouraged me along the way of completing my thesis. I would like to extend my sincere thanks to my supervisor FH-Prof. DI Dr. Lars Mehnen for his constructive comments and valuable guidance that helped me proceed and properly shape the form and content of my thesis. Many thanks to my colleagues from the Bioelectronics Neurophysiology and Engineering Lab at Mayo Clinic who provided me with enriching insights and a pleasant working environment. Special thanks go to Ing. Filip Mívalt, M.Sc. who shared his invaluable technical expertise and motivated me in many aspects. Last but not least, I would like to thank my family and friends for their unlimited support and patience throughout my entire studies.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Epilepsy . . . . .	1
1.2	Electroencephalography (EEG) in Epilepsy . . . . .	2
1.3	SOZ Localization . . . . .	4
1.4	Machine Learning . . . . .	4
1.5	Deep Learning . . . . .	6
1.6	Problem Statement . . . . .	9
1.7	Aims & Contributions . . . . .	11
<b>2</b>	<b>Methods</b>	<b>12</b>
2.1	Data . . . . .	12
2.2	Pre-processing . . . . .	13
2.3	SOZ Biomarkers . . . . .	14
2.4	Unsupervised iEEG Exploration . . . . .	14
2.5	Dataflow & Visualisation . . . . .	16
2.6	Statistical Evaluation . . . . .	16
2.7	Experimental Framework . . . . .	17
2.7.1	Experiment 1 . . . . .	17
2.7.2	Experiment 2 . . . . .	18
2.7.3	Experiment 3 . . . . .	19
2.7.4	Experiment 4 . . . . .	22
<b>3</b>	<b>Results</b>	<b>24</b>
3.1	Experiment 1 . . . . .	24
3.2	Experiment 2 . . . . .	26
3.3	Experiment 3 . . . . .	29
3.4	Experiment 4 . . . . .	35
<b>4</b>	<b>Discussion and Conclusions</b>	<b>37</b>
	<b>Bibliography</b>	<b>43</b>
	<b>List of Figures</b>	<b>48</b>
	<b>List of Tables</b>	<b>49</b>

**List of Abbreviations**

**51**

**A Appendix**

**53**

# 1 Introduction

The recent trends in neurology and the ongoing development of state-of-the-art technologies result in collecting of an enormous amount of neurophysiological data spanning over days and even years [1]. Long neurophysiological data needs to be manually reviewed and analysed in the research and clinical practice. However, manual analysis of the increasing amount of data is becoming lengthy and burdensome. Deep learning (DL) is a powerful artificial intelligence (AI) tool which can computationally solve complex multidimensional tasks associated with the enormous data inflow [2]. It has been successfully used in various areas, including biomedical applications such as clinical image segmentation, genomics, or seizure prediction [3–5]. This thesis aims to investigate the feasibility of utilising DL models for unsupervised exploration of iEEG data in epilepsy, more specifically, the capability to distinguish between iEEG signals recorded from the seizure onset brain structures and brain regions otherwise healthy in terms of epilepsy.

## 1.1 Epilepsy

Epilepsy is a neurological disease affecting around 50 million people worldwide, characterised by abnormal electrical brain activity resulting in epileptic seizures [6]. The first notes of the disease date back to ancient Greece. At that time, people suffering from epilepsy were considered to be possessed by supernatural forces and were therefore ostracised [7]. Nowadays, the true circumstances of epilepsy are well known, nevertheless, people with epilepsy carry the burden of short-term motor and perceptual dysfunctions that can cause social disadvantage and decrease the quality of daily life [7].

### Seizures

An epileptic seizure is described as a sudden, excessive electrical discharge that can have a variety of aetiologies. The genesis of seizures is attributed to genetic mutations, trauma, neurochemical mechanisms, inflammation or as a side-effect of other diseases [7]. Factors that influence seizure occurrence might be brain maturity, onset location, sleep-wake cycle, medication use and many others. The body functions that seizures can impact include memory, cognition, emotional state, consciousness, sensory, motor or autonomic system [8]. Seizures are classified according to their onset. Onset can be focal, generalized, unknown, or unclassified. Focal seizures have a single focus, whereas generalized seizures can originate in both

hemispheres. Unknown seizures express the uncertainty of the onset zone, but known manifestations [9]. Focal features are further categorized according to the state of consciousness as “impaired awareness” or “aware”. The next classification level for both focal and generalized seizures is defining their “motor” or “non-motor” characteristics [10].

## Treatment

The typical treatment option is antiepileptic drugs, which successfully suppress seizure occurrence in about 60-70 % of people with epilepsy. However, the remaining 20-30 % of patients who fail to achieve seizure freedom over a time scale of one year are considered to be drug-resistant [11]. Alternative treatment options for people with refractory epilepsy can involve diet therapies, electrical nerve stimulation, or a brain resection surgery [12]. Candidates for brain resection surgery undergo several procedures, including a thorough medical anamnesis, scalp electroencephalography (EEG) with video-monitoring, magnetic resonance imaging (MRI), and neuropsychological testing [13]. If brain resection surgery is endorsed, patients need invasive intracranial electroencephalography (iEEG) monitoring to more accurately distinguish the nature of the electric field potentials to determine the seizure onset zone (SOZ) [14].

## Seizure Onset Zone (SOZ)

The epileptic brain consists of several cortical regions related to seizure initiation and propagation. The seizure onset zone corresponds to one or more brain locations responsible for seizure generation. However, the precise definition of SOZ remains challenging. The epileptogenic zone covers the brain area manifesting abnormal electrical activity. The epileptogenic zone is usually more extensive than the SOZ and often overlaps with the SOZ, making the delineation of the boundaries between the two areas ambiguous. Patients with multiple epileptogenic foci with different thresholds may continue to have seizures after brain resection surgery because they will manifest another SOZ with a higher threshold. Due to these practical barriers, current workflows do not allow definite SOZ localization [15].

## 1.2 Electroencephalography (EEG) in Epilepsy

The EEG signal is generated by summing the excitatory and inhibitory postsynaptic potentials of simultaneously activated neurons located near the recording electrode. The EEG acquisition measures differences in electrical potentials between two electrodes. The maximal electrical activity is sensed when the electrode is placed perpendicularly to active neurons [16]. However, scalp EEG may not represent the neural activity precisely. Before a neuronal signal is detected at the electrode, it needs to pass through multiple tissues, decreasing its amplitude and spatial resolution. In addition, the recording may be distorted by unrelated activity of other biological

systems or electrical devices in the recording room. These effects can significantly reduce the ability to correctly interpret the EEG recording and assess the epileptogenic focus [16].

Invasive iEEG recording is an option to surpass the scalp EEG insufficiency. It can more accurately define the epileptogenic zone, and its connection to eloquent brain regions [13]. The most common techniques for iEEG monitoring involve electrocorticography (ECoG) and stereoelectroencephalography (SEEG). ECoG uses subdural electrodes, which are grid or strip arrays of disc-shaped contacts embedded in the silicon surface. The contact size is about 4-5 mm and the spacing between individual contacts ranges between 5mm and 1 cm. Implantation of subdural electrodes is usually performed by craniotomy [17]. SEEG recording utilises depth electrodes anchored on a hollow plastic tube. One tube usually contains 4-18 contacts spaced 2-10 mm apart. Depth electrodes are implanted during stereotactic surgery through burr holes [18]. The advantage of grid electrodes is their high density in continuous cerebrum areas allowing for good spatial resolution and tracing of seizure expansion. However, subdural electrodes cannot be inserted into deeper brain structures without a high risk of tissue damage [17]. In contrast, implantation of SEEG electrodes is safer than craniotomy and is feasible for recording from deep brain structures [19].

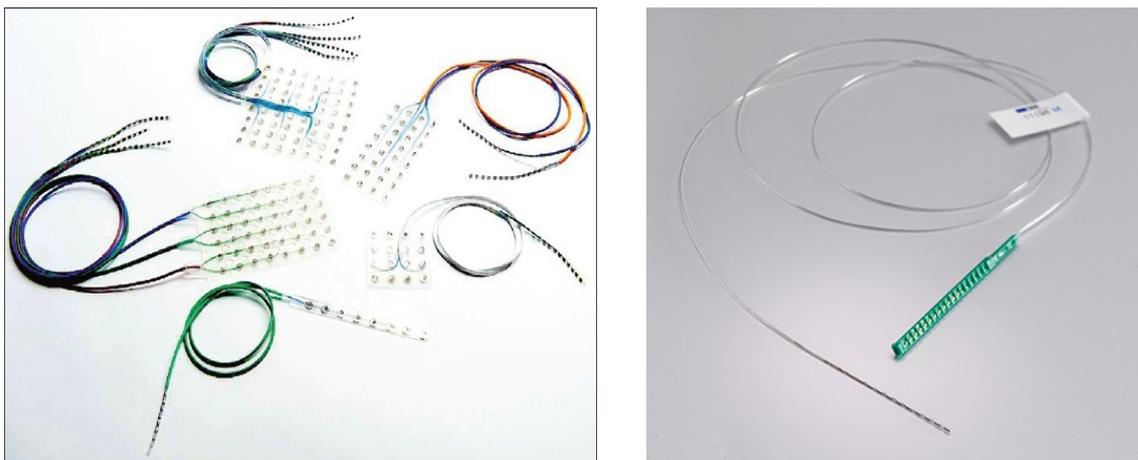


Figure 1: Electrodes used for iEEG recording. Left: Subdural electrodes (grids and strips). Right: Depth electrodes. (Source: [20, 21]).

## Brain Electrophysiology Manifestation

Brain neural activity is described by amplitude, frequency, morphology, continuity, localisation and synchronisation. Specific EEG patterns relate to the manifestation of physiological or pathological conditions [22].

Frequency is the most common characteristic for distinguishing individual EEG waveforms. The physiological scalp EEG primarily include four frequency bands: alpha, beta, theta, and delta. Delta (0.5 – 4 Hz) and theta (4 - 7 Hz) bands physiologically represent deep and early sleep stages. Alpha rhythm (8 – 12 Hz) characterises the background EEG, best acquired in

a relaxed mental state with eyes closed. Beta rhythm (13 – 30 Hz) represents normal wake EEG [22]. High frequency oscillations (HFO) indicate activity faster than 30 Hz. HFO are further divided into gamma (30 to 80 Hz), ripples (80 to 200 Hz), and fast ripples (200 to 500 Hz). Gamma waves are associated with sensory perception. Ripples and fast ripples can be physiologically observed in iEEG during cognitive tasks, or somatosensory stimulation [22].

HFO activity was also observed in relation to the epileptic brain in ictal and interictal states. However, various studies reported that ictal HFO frequency ranges overlap with physiological HFO, making them difficult to use as a single biomarker to identify epileptogenic tissue [23]. Other epilepsy-specific patterns involve interictal spikes, sharp waves, spike-wave complexes, seizure patterns, and status patterns [24]. Interictal spikes (IS) are defined as fast high-amplitude transients followed by a slow wave with a tendency to occur periodically. Although IS are established indicators of epilepsy, they do not necessarily associate with seizure generation, as their occurrence is supported by different neuronal mechanisms [25].

### 1.3 SOZ Localization

This section presents current approaches to determining SOZ. Conventional techniques usually involve the calculation of a particular biomarker on iEEG data. The latest research showed that HFO is an important indicator of epileptic brain tissue [26]. However, distinguishing between physiological and pathological HFO remains challenging. Therefore, the employment of HFO as the only biomarker seems to be insufficient [27].

A previous study showed that the multi-feature approach improves the SOZ localisation compared to classifiers utilising a single feature [28]. Moreover, another study suggests that SOZ localisation using iEEG signals recorded during non-rapid-eye-movement sleep improves SOZ classification performance compared to the awake state [29]. It was also shown that cognitive tasks alter HFO properties in the epileptic and non-epileptic hippocampus [26, 30].

Another study suggests using Bayesian filtering of power in band features to cluster iEEG channels exhibiting similar activity as an alternative method for SOZ localisation [31]. Relative entropy (REN), which attributes to the measurement of functional connectivity of the epileptic brain, was used as another biomarker for SOZ differentiability [32].

### 1.4 Machine Learning

Machine learning is a set of techniques including regression, classification, dimensionality reduction, and others that are widely used for data mining, object recognition, and other tasks requiring automation [33]. Dimensionality reduction techniques involve transformations from a high-dimensional space to a low-dimensional space while preserving relevant information from the original data. Dimensionality reduction techniques are widely utilised with emerging big data acquisition to explore huge datasets [34]

In this section, the dimensionality reduction methods and the classification model used in this work are introduced.

## Principal component Analysis (PCA)

Principal component analysis (PCA) is a widely used linear technique to reduce dimensionality. The goal of implementing this method is to find new features transformed from the original data while maintaining maximum variance. The new features, called principal components, are uncorrelated with each other. The PCA transformation is based on the mathematical concept of eigenvalue decomposition [35]. A covariance matrix is estimated from the original  $n$ -dimensional data to identify the variance of each feature and the internal correlations. The eigenvectors of the covariance matrix represent the directions of the axes with the largest variance, numerically explained by the eigenvalues. The sorted eigenvalues determine the order of the principal components based on the variance. The corresponding eigenvectors multiplied with the transposed original dataset then transform the original features into PCA components [36].

## Uniform Manifold Approximation and Projection (UMAP)

UMAP is a manifold learning method for dimensionality reduction. The theoretical background of UMAP involves manifold theory and topological data analysis. Briefly, the UMAP reduction technique keeps the local distances and still preserves the global structure. It can scale large dataset volumes without computational constraints on embedding dimensions. The purpose of UMAP is to create a topological representation of high-dimensional data. That consists of two steps: 1) approximation of manifold on which the data is estimated to lie, 2) creating a fuzzy simplified set representation of the approximated manifold. Next, the layout of data representation in low dimensions is optimised to achieve the best fit for high-dimensional representation [37].

## Naive Bayes Classifier

The Naive Bayes classifier is built upon a Bayes' Theorem (Equation 1) predicting a conditional probability  $P(C|X)$  of the class  $C$  given the feature vector  $X$ . Class-conditional probability distribution  $P(X|C)$  is obtained as a probability density function of the given feature vector  $X$  for each class  $C$  [38]. A probability density function for individual classes is estimated by employing kernel density estimation (KDE) technique [39].

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)} \quad (1)$$

## 1.5 Deep Learning

Deep learning (DL) is a set of machine learning techniques characterised by multi-layer transformations that nowadays supports various machine tasks in our daily routines. Its main benefit is the ability to process a vast amount of data within a relatively short period. Unlike machine learning techniques, most DL algorithms can work directly with raw data without the need for predefined features. Typical problems solved by DL include face recognition, object detection, image segmentation, speech recognition, language processing, and many others. Such tasks would be impossible to handle in a reasonable time frame solely by a human [40].

In deep learning applications, there are usually two basic approaches: supervised and unsupervised. Supervised learning requires labels assigning the training data into specific groups. The algorithms then learn to discriminate a pattern in the features based on the label. Typical tasks for supervised learning are classification and regression. Examples of supervised models are convolutional neural networks (CNN) or recurrent neural networks (RNN). On the other hand, unsupervised learning methods do not need labels and learn a pattern directly from the provided data. Common unsupervised learning algorithms involve autoencoder or deep Boltzmann machines models. Depending on the task, both approaches can be combined as so-called hybrid or semi-supervised approaches [41].

### Convolutional Neural Network (CNN)

A convolutional neural network (CNN) forwards the input training sample through multiple hidden layers applying convolution and usually non-linear functions to create feature representations. The feature representations are then commonly used for various regression or classification tasks. CNN can process one-dimensional, or image data [42]. The convolutional layer convolves the input with a chosen kernel. The mathematical formula of discrete 2D convolution is shown in Equation 2, where  $h$  represents the kernel matrix, and  $x$  represents the input image. The asterisk denotes convolution operation. The indices  $m, n$  relate to the kernel matrix, and the indices  $i, j$  denote the image pixels.

$$y(i, j) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} h(m, n) * x(i - m, j - n) \quad (2)$$

The result of convolving the input image is a feature map that keeps the information extracted from the previous layers. The advantage of a convolutional feature map is that it can reduce the size of the original image and store the information locally as the kernel moves along the image [43]. A schematic illustration of the 2D convolution is shown in Figure 2.

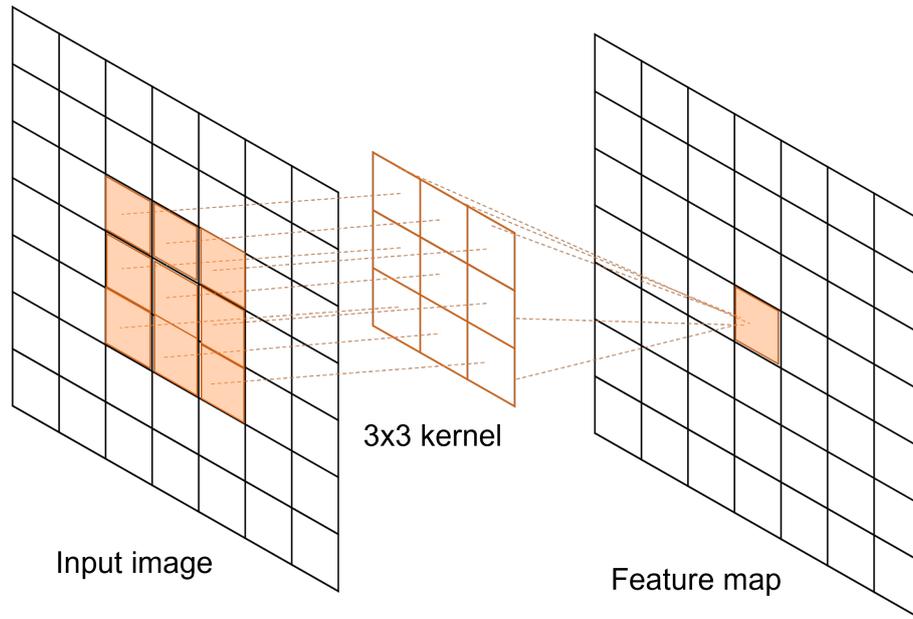


Figure 2: A schematic example of 2D convolution.  $3 \times 3$  kernel is convolved with a part of the input image matrix and produces a single value on the feature map.

## Recurrent Neural Network (RNN)

RNN is another type of neural network architecture. RNN contains a recurrent branch that allows the network to store and retain information over time. This makes the RNNs more suitable for processing time-series data such as speech, text, or videos [41]. The Gated Recurrent Units (GRU) network is a type of RNN proposed by Cho et al. [44] upon the previously established long short-term memory network [45]. An illustration of a hidden GRU unit is depicted in Figure 3.

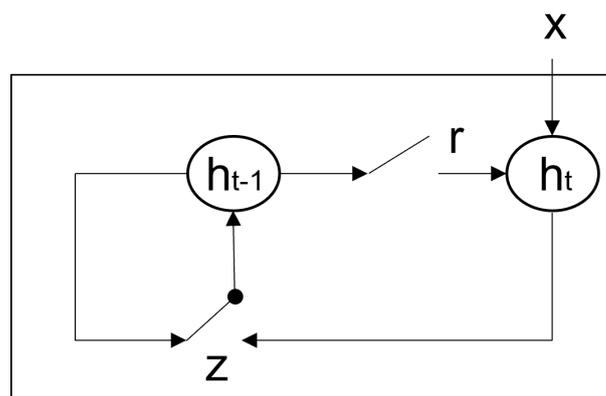


Figure 3: Structure of a hidden GRU unit. The new input sample  $x$  is passed to the unit and contributes to the new hidden state. The update gate  $z$  controls if the previous hidden state is replaced with the current hidden state. The reset gate  $r$  determines if it drops the last hidden state.

## Attention Mechanism

An attention mechanism was introduced to capture contextual information from the long sequence data by modelling interdependence between all input samples [46]. Later it was also applied in the image domain for segmentation [47]. The attention mechanism is inspired by cognitive processes in the human brain that select only a particular piece of information from the surroundings to focus on [48].

## Autoencoder

Autoencoder (AE) models belong to unsupervised deep learning methods. Training an autoencoder model involves learning a characteristic pattern from the data itself rather than from the labels. The autoencoder consists of an encoding and a decoding part. The encoding part transforms the input into low-dimensional embedded features, which are then forwarded through the decoding part to reconstruct the original input [49]. The following section describes the architecture of the in-house developed Conv-GRU autoencoder used in this work.

The in-house unsupervised temporal Conv-GRU autoencoder was originally designed to extract features from sequential iEEG data and cluster them into categories according to signal type. The autoencoder architecture can be seen in Figure 4. The input data is represented by a spectrogram generated from the iEEG signals. The encoding part contains convolutional, batch normalisation, GRU, and linear attention layer. The decoding part consists of GRU and convolutional layers. The reconstructed and original spectrogram difference is evaluated by the mean absolute error (MAE). The error is minimised using the ADAM optimiser [50]. The self-learned embedded features produced by the encoding part are evaluated within this thesis.

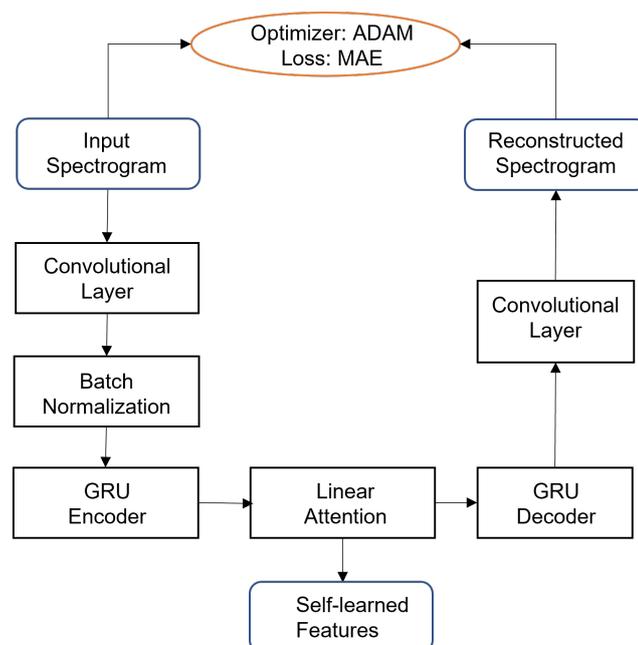


Figure 4: Architecture of Conv-GRU autoencoder network for unsupervised iEEG feature extraction.

## 1.6 Problem Statement

SOZ localisation is a necessary task prior to a surgical procedure for invasive seizure management. The area of the brain to be surgically removed is routinely determined by a physician consortium using iEEG recordings. Even though the clinical practice might differ across countries, the trend of increasing data volume is common. The massively increasing amount of collected iEEG data makes manual analysis by clinical staff unbearable [51]. The big amount of acquired iEEG signals presents an opportunity to investigate algorithms for unsupervised exploration of iEEG data in epilepsy [52]. Even though multiple algorithms for SOZ detection have been developed in the past years [26,28,32], their use in clinical practice has not yet been established. Similarly, many machine learning algorithms focused on iEEG data mining [53], seizure detection [54], or prediction [55,56] were introduced. However, these algorithms were not designed to identify brain regions initiating seizures. To the best of my knowledge, unsupervised iEEG analysis algorithms and their ability to extract features capable of distinguishing SOZ brain tissue from NSOZ brain tissue have not been investigated so far.

The ultimate vision to be achieved in enhancing the clinical data review in epilepsy is illustrated in Figure 5. A candidate for epilepsy surgery undergoes invasive iEEG monitoring, resulting in a large amount of recorded data. The vision is to dispense a system that will automatically analyse the recorded data in an unsupervised manner and sort the iEEG signals into the categories of interest. This work investigates the feasibility of employing an unsupervised DL model that would encode the iEEG signals into embedded feature representation (EFR) based on the characteristic of recorded data. The EFR would be clustered using unsupervised approaches providing information about the iEEG signals and presented to the physician. This would allow focusing the human expert's attention only on the regions of interest and, therefore, enhance the whole data review.

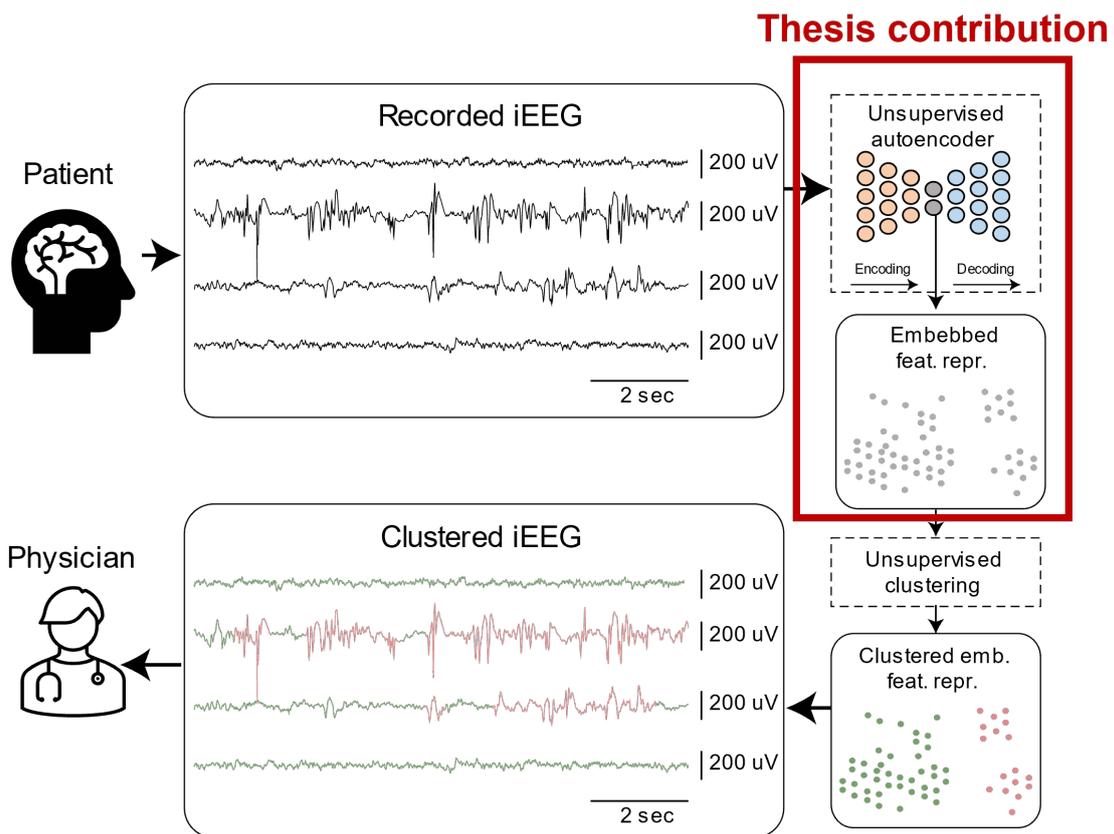


Figure 5: Unsupervised intracranial EEG (iEEG) clustering paradigm for aiding clinical staff in intracranial EEG assessment prior to epilepsy surgery. Long-term pre-surgical iEEG is reduced into embedded feature representation (EFR) using unsupervised autoencoder and processed using unsupervised clustering algorithms. The clustering results are presented to clinical staff emphasizing prominent iEEG segments. This thesis aims to explore the feasibility of autoencoder utilization to discriminate the SOZ channels from NSOZ channels and to compare the approach with established biomarkers.

## 1.7 Aims & Contributions

This work lays the first foundations for the vision described above by investigating the feasibility of an unsupervised DL autoencoder to discriminate between SOZ and NSOZ channels as shown in Figure 5. Specifically, the thesis explores whether a previously in-house developed unsupervised Conv-GRU autoencoder can differentiate among signals of various nature and compares the autoencoder to established biomarkers. Four specific questions were stated to address this scientific problem.

The specific question 1 was defined to validate the capability of the Conv-GRU autoencoder to distinguish among different iEEG categories including short iEEG segments of physiological, pathological, noise, and powerline interference signals. Within the specific question 2, the thesis explored potential use of the Conv-GRU autoencoder to differentiate between short segments of SOZ and NSOZ iEEG. Specific question 3 was addressed to closely analyse features derived from established biomarkers (HFO, IS, and REN) and their ability to discriminate between SOZ and NSOZ iEEG channels on full-length clinical recordings. The purpose of Specific question 3 was to extend the knowledge on the discriminative potential of individual features as a follow-up on previously published studies and to determine the baseline for performance comparison with autoencoder-extracted features. Finally, Specific question 4 was aimed at comparing the performance of features extracted by the unsupervised autoencoder against the features derived from established biomarkers to assess whether unsupervised AE can be employed in the data review process for SOZ localisation, and therefore become a potential subject of future research.

**Specific question 1:** “Are self-learned embedded features extracted by AE trained using 1-second iEEG segments capable of distinguishing between physiological, pathological, noise and powerline interference segments?”

**Specific question 2:** “Are self-learned embedded features extracted by AE trained using 1-second iEEG segments capable of distinguishing between SOZ and NSOZ segments?”

**Specific question 3:** “What is the performance of established biomarker-derived features for SOZ vs NSOZ iEEG discrimination in full-length clinical data (30-120 min) on a) a group level; b) a subject-level?”

**Specific question 4:** “Do self-learned embedded features extracted by AE outperform conventional biomarker-derived features in discriminating SOZ and NSOZ channels?”

## 2 Methods

The methods chapter is organised in the following way. The first part introduces a validation of the in-house developed unsupervised autoencoder for iEEG classification. The second part is focused on investigating the autoencoder feasibility for SOZ channels distinction on full-length clinical data and its comparison with established biomarkers. All data were processed and analysed in a Python programming language.

### 2.1 Data

#### a) Full-length clinical iEEG data

Full-length iEEG data were recorded during pre-surgical monitoring for SOZ localisation at St Anne's University Hospital (Brno, Czech Republic) and at the Mayo Clinic (Rochester, Minnesota, United States of America) in patients diagnosed with drug-resistant epilepsy (DRE). FNUSA data were collected in 18 patients implanted with intracranial depth electrodes using the BrainScope acquisition system. The recording time was 30 minutes with a sampling frequency of 25 kHz. The Mayo Clinic dataset contains 2 hours long iEEG from 15 patients with implanted subdural or intracranial depth electrodes or a combination of both. For the purposes of this thesis, only signals recorded from the depth electrodes were included. The used recording device was the Neuralynx Cheetah system with a 25 kHz sampling frequency. All data were resampled a sampling rate of 5000 Hz. The recorded channels were manually labelled as SOZ, NSOZ, or irritative zone (IZ), reflecting area generating IS, whose classification as SOZ or NSOZ is ambiguous. Channels assigned as IZ were excluded from the analysis. The overview of the information on iEEG recordings from both institutes is available in Appendix in Tables 16 and 17.

#### b) Derived iEEG-segments dataset

The publicly available dataset [57] was derived from iEEG recordings presented beforehand. The dataset consists of 3-second iEEG signal segments collected from St. Anne's University Hospital (FNUSA) and the Mayo Clinic. The FNUSA dataset and the Mayo Clinic dataset comprise 193,118 and 155,182 signal fragments, respectively. The sampling frequency was 5000 Hz. The iEEG segments contain four groups of signals: physiological activity, pathological activity, powerline interference, and noise. Examples of each signal type can be seen in Figure 6. All data is annotated for its respective signal group and SOZ information.

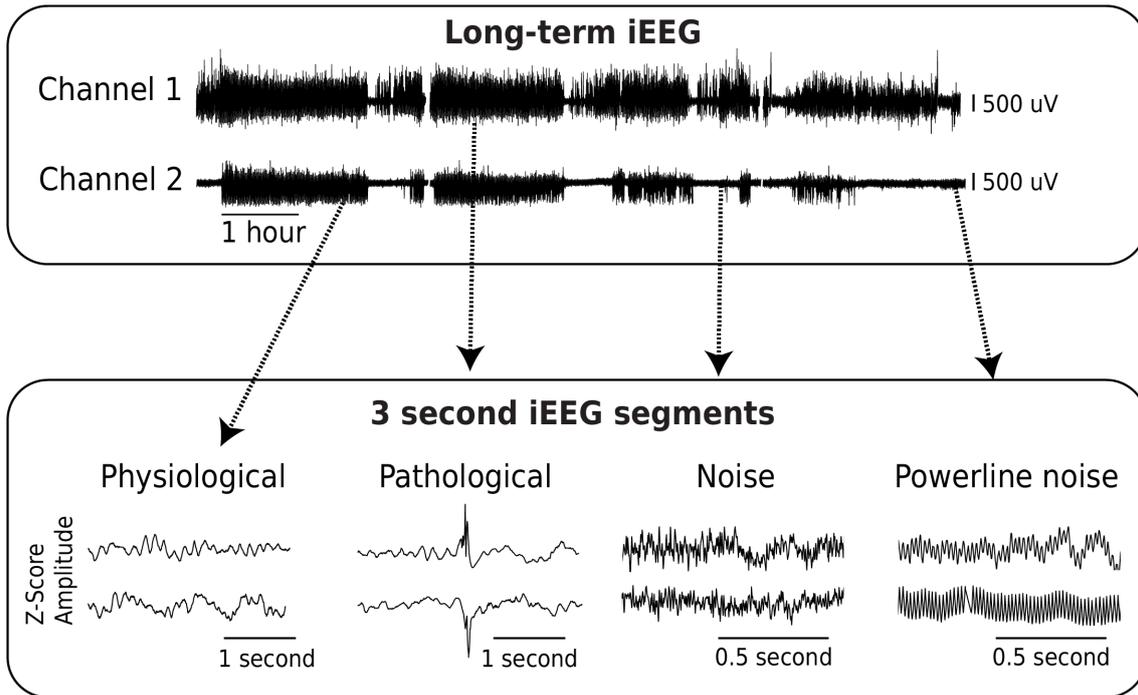


Figure 6: Illustrative examples of signals from the used datasets. Top row: long-term iEEG signals from two channels employed in Experiment 3 and Experiment 4 to explore the potential of iEEG-extracted features to discriminate the SOZ. Bottom row: dataset containing 3-second segments derived from the long-term data. The dataset includes physiological, pathological, powerline interference, and noise segments used in Experiment 1 and Experiment 2 to train and validate the Conv-GRU autoencoder.

## 2.2 Pre-processing

Data pre-processing differed depending on the form of feature extraction. The autoencoder-based feature extractor obtained already filtered data, whereas algorithms for established biomarkers had filtering implemented internally and therefore received raw data. The following section describes pre-processing techniques used before passing to the autoencoder. The iEEG signals were passed through a third-order Butterworth filter [58] with cut-off frequencies of 0.5 and 200 Hz. They were further filtered by a median filter with a window size of 5. Finally, the signals were normalised by subtracting the mean and divided by the standard deviation. Two spectrograms computed from the iEEG signals were used as input for the autoencoder feature extractor. The spectrograms were estimated by short-time Fourier transform in one-second segments with a window size of 299 samples and an overlap of 199 samples. The maximal frequency was cut down to 400 Hz. One spectrogram was subsequently transformed into a decadic logarithmic scale. The second spectrogram was z-score normalized along the time axis, highlighting the events in time rather than in frequency content. An example of a signal segment and derived spectrograms is depicted in Figure 7.

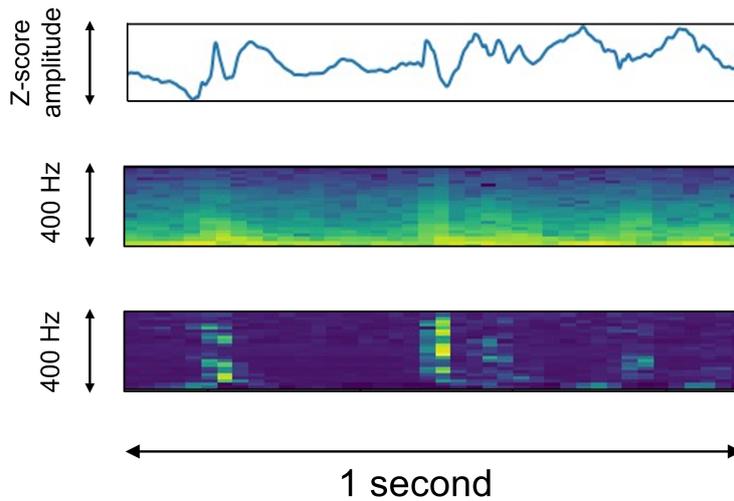


Figure 7: Example of an iEEG signal segment (top row) and its corresponding spectrograms. The middle row shows spectrogram in decadic logarithm scale. The bottom row displays the normalized spectrogram.

## 2.3 SOZ Biomarkers

Conventional biomarkers were used as a reference for the performance of autoencoder-based features. The algorithms applied for feature extraction are implemented in the EPYCOM library available at <https://gitlab.com/icrc-bme/epycm>. Specifically, the biomarkers include HFO, IS, and REN. HFO were detected in a 10-second statistical window using the CS detector [59]. The HFO detector produces amplitude, frequency dominance, frequency and amplitude dot product, and duration. The amplitude and duration features were used for the analysis. The IS features were also calculated in a 10-second moving window using the Barkmeier detector [60]. The IS features are described by amplitude (event, left, right) and duration (left, right). The mean rate of events per 10 minutes and the mean amplitude per 10 minutes were calculated for both HFO and IS features according to previously published studies [30,61]. REN was estimated in a one-second window between all channel pairs in each patient as  $REN(X, Y) = \sum P(X) \log \frac{P(X)}{P(Y)}$ , where  $P(X), P(Y)$  represent the probability distributions of the two channels. REN was computed only for electrodes that had adjacent contacts. For the purpose of distinguishing between the SOZ and NSOZ channels, the mean value from the adjacent electrodes was associated with a single channel.

## 2.4 Unsupervised iEEG Exploration

The ability of the Conv-GRU autoencoder described in section 1.5 to separate iEEG segments into categories was examined on the publicly available dataset.

Two autoencoders were trained separately on one-second iEEG segments from the FNUSA

dataset (AE-FNUSA model) and from the Mayo dataset (AE-Mayo model). The AE models received input in the form of two spectrograms which represent the fluctuating frequency content of the signal over time. The autoencoder training consisted of extracting features characterising the input spectrograms and then using the features for reconstruction. An example of the original and reconstructed spectrograms during training is shown in Figure 8. Convolutional and training parameters were set empirically (Table 1). Both models were trained for 10 epochs while minimising MAE by Adam optimiser. The outcome of the trained autoencoder was 128 self-learned encoding features for each second of input data.

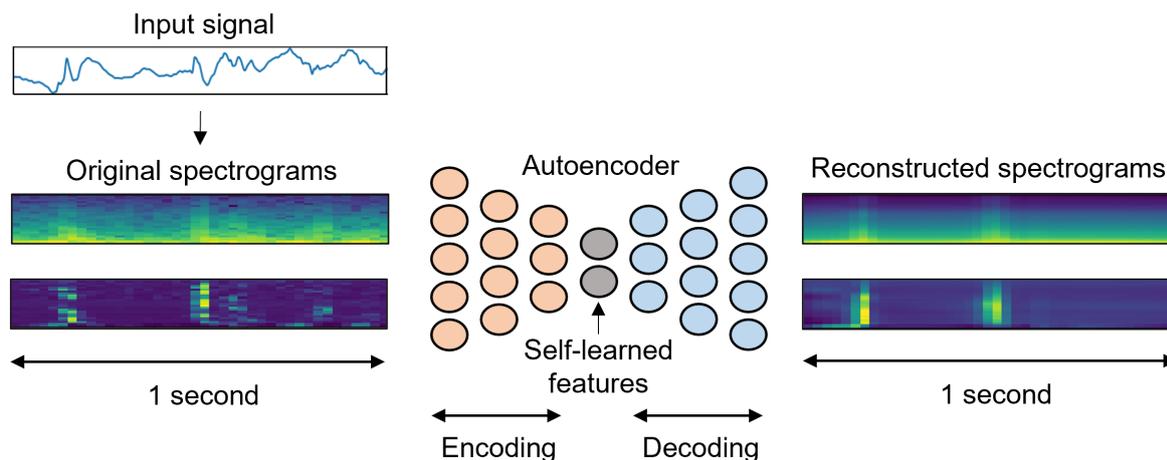


Figure 8: A schematic example of an autoencoder. Input data passes through the encoding layers, and the autoencoder extracts the self-learned features. In the reconstruction process, the features are propagated through decoding layers to produce an output.

Input kernel size	24 x 1
Stride	1
Padding	0
Batch size	128
Epochs	10
Learning rate	0.003
Iterations per epoch - FNUSA	1213
Iterations per epoch - Mayo	1509
Loss function	MAE
Optimizer	Adam
Gradient clipping	0.1

Table 1: Training parameters for Conv-GRU autoencoder training and optimization.

## 2.5 Dataflow & Visualisation

When processing the full-length data, the features were obtained within a non-overlapping moving window that split the iEEG signals into shorter segments. The particular extractor (autoencoder, HFO detector, IS detector) processed the input signal within the window along the whole signal in all channels per patient. The moving window length was 1 second for autoencoder-extracted and relative entropy features and 10 seconds for HFO and IS features. The scheme of feature extraction workflow is shown in Figure 9. Due to a high number of AE self-learned features, methods for dimensionality reduction were employed. For visualisation purposes of multidimensional data, the UMAP technique was applied to reduce the dimensionality down to 2 embedded features. For evaluation, the features were transformed into a lower number of components using PCA or UMAP.

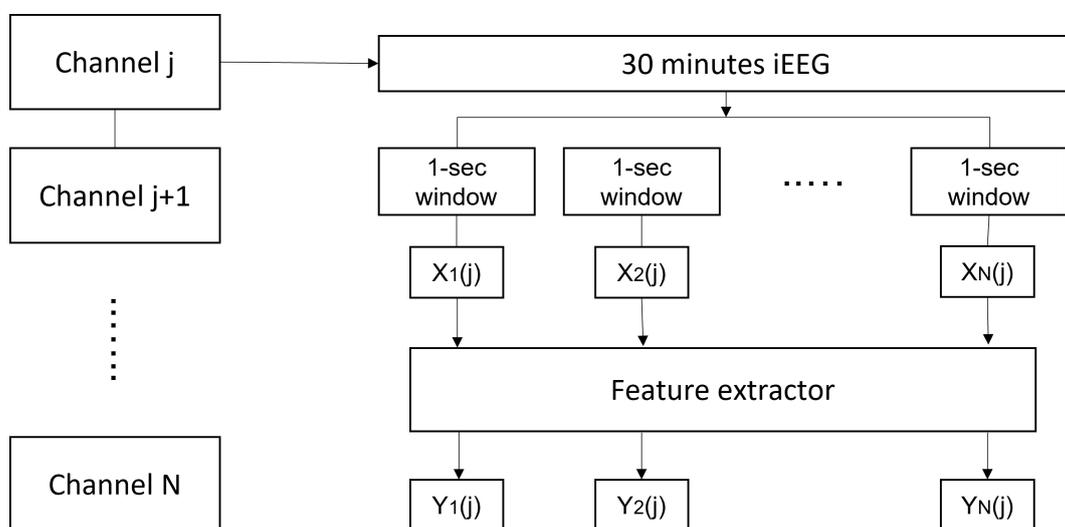


Figure 9: Data workflow during feature extraction. All channels for individual patients are consecutively divided into shorter segments  $X_{1..N}$ . The feature extractor produces the corresponding outputs  $Y_{1..N}$ . The length of a moving window relies on the type of computed feature.

## 2.6 Statistical Evaluation

In order to evaluate the performance of analysed features and answer the questions stated in section 1.7, statistical testing and classification scores were used as metrics.

The separation ability of AE embedded features to distinguish different signal segments was assessed based on their performance in the naive Bayes classifier using F1 score [62]. Statistical difference of biomarkers distributions between SOZ and NSOZ channels was examined using the non-parametric Mann-Whitney test [63]. The capability of individual biomarker features to discriminate between SOZ and NSOZ was evaluated by area under the receiver operating curve (AUROC) [64]. The performance of biomarkers was analysed for the aggregated group

of patients as well as for all patients separately. To compare performances of AE self-learned embedded features against biomarker features, AUROC, average precision (AP) [65], and F1 scores were evaluated on separate classifiers. The statistical difference between the scores obtained by each classifier was tested with the paired Wilcoxon signed-rank test [66].

## 2.7 Experimental Framework

This section introduces specific workflows that were followed in order to investigate the questions stated in section 1.7. Experiment 1 and Experiment 2 followed the same workflow except for the last part, which differed in the types of categories analysed. The graphical explanation of the framework of Experiment 1 and Experiment 2 can be seen in Figure 10.

AE-FNUSA and AE-Mayo models were trained on the second-long iEEG segments containing physiological, pathological, noise, and powerline interference fragments. Training of each model was performed on segments corresponding to the institution in which they were recorded. After training, the one-second iEEG segments were passed to their institution-respective model, and 128 self-learned features were extracted for each sample. In the next step, the self-learned features were transformed using PCA or UMAP into low-dimensional embedded components. The minimum 85% of explained variance determined the selected number of embedded components for PCA transformation. For the UMAP transformation, several numbers of embedded components were investigated in terms of classification score and overall duration. 85% of the embedded features was used to train the Naive Bayes classifier utilising the KDE probability density function. The remaining 15% of embedded features was employed as a test subset for the classifier to assess the index of separability.

### 2.7.1 Experiment 1

Experiment 1 was focused on exploring the potential of the unsupervised autoencoder to extract features capable of discriminating signals of different nature. The total number of samples in each category in both institutions is summarized in Table 2. Following the workflow outlined in the previous section 2.7, embedded self-learned features were extracted by AE models, reduced into lower dimensionality, and evaluated by a Naive Bayes classifier. The classifier was used in a manner of scoring metric to assess the separability index determined by the F1 score. The separability index was obtained for all pairwise combinations between categories. The whole experiment was repeated ten times for PCA-transformed features and five times for UMAP-transformed features. The mean F1 score and standard deviation were calculated. Based on the F1 score and processing time, only one method for dimensionality reduction was selected and employed in the subsequent experiments.

The aim of Experiment 1 was to obtain four non-overlapping clusters, each represented exclusively by features of physiological, pathological, noise, or powerline interference categories.

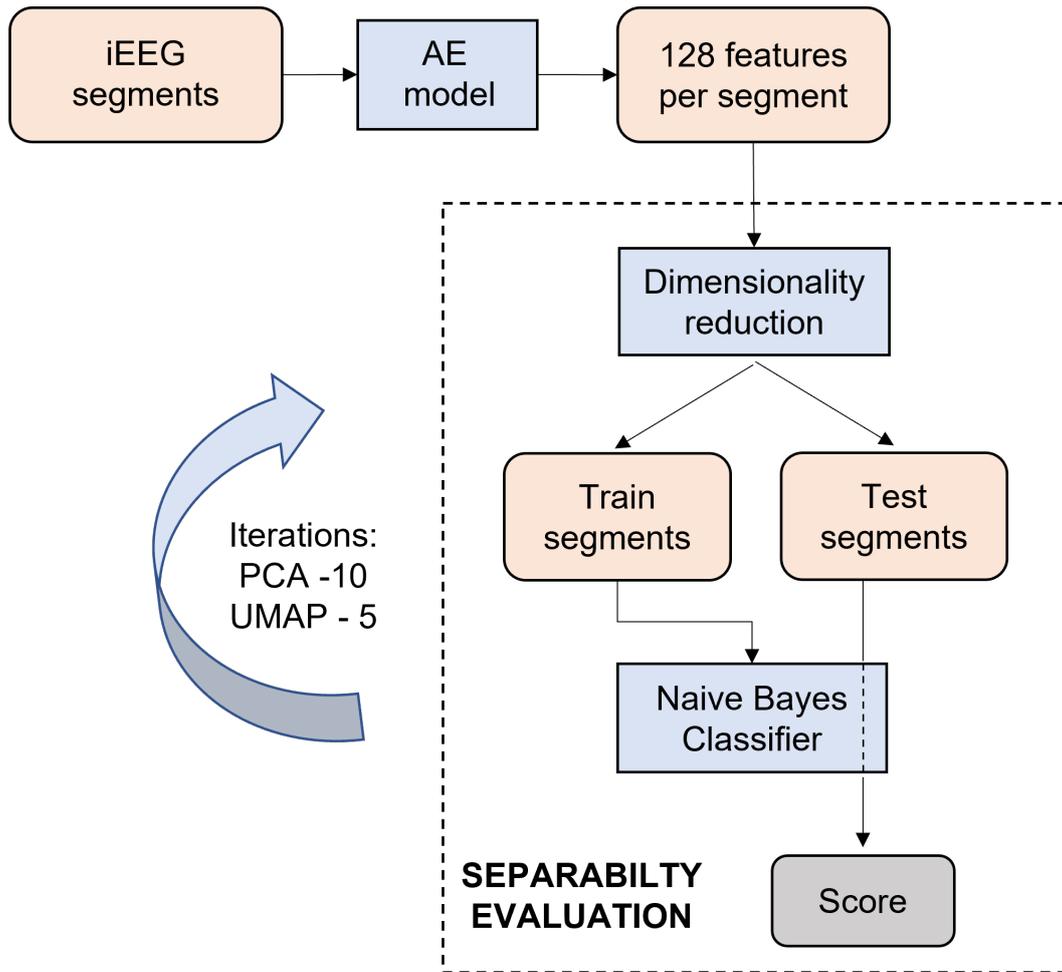


Figure 10: The general framework for determining the index of separability in Experiment 1 and Experiment 2.

The resulting clusters were plotted as two embedded UMAP components coloured according to their respective segment labels.

## 2.7.2 Experiment 2

Experiment 2 encompassed a slight modification of the first experiment. The investigated task was to assess the capability of the unsupervised autoencoder to differentiate iEEG data according to whether the segment originated from the SOZ or NSOZ. Since noise-corrupted iEEG signals are not considered in the visual assessment of the SOZ, only physiological and pathological segments were included in the analysis. The number of iEEG segments related to SOZ or NSOZ in each dataset is shown in Table 3. The same techniques as in the previous experiment were applied. The experiment was replicated ten times with the selected dimensionality reduction technique, and the mean F1 score and standard deviation were calculated. In Experiment 2, the target outcome was to retrieve two non-overlapping clusters, one representing the segments from SOZ and the other representing the segments from NSOZ. To visualise the

resulting clusters, the two UMAP transformed self-learned embedded features were displayed by a scatter plot and a contour plot expressing the density of cluster data distributions. The concept of Experiment 1 and Experiment 2 is illustrated in Figure 11.

Category	Number of samples in FNUSA dataset	Number of samples in Mayo dataset
Noise	32,599	41,303
Pathology	52,470	15,227
Physiology	94,560	56,730
Powerline interference	13,489	41,922

Table 2: Total number of iEEG segments in each category in FNUSA and Mayo datasets.

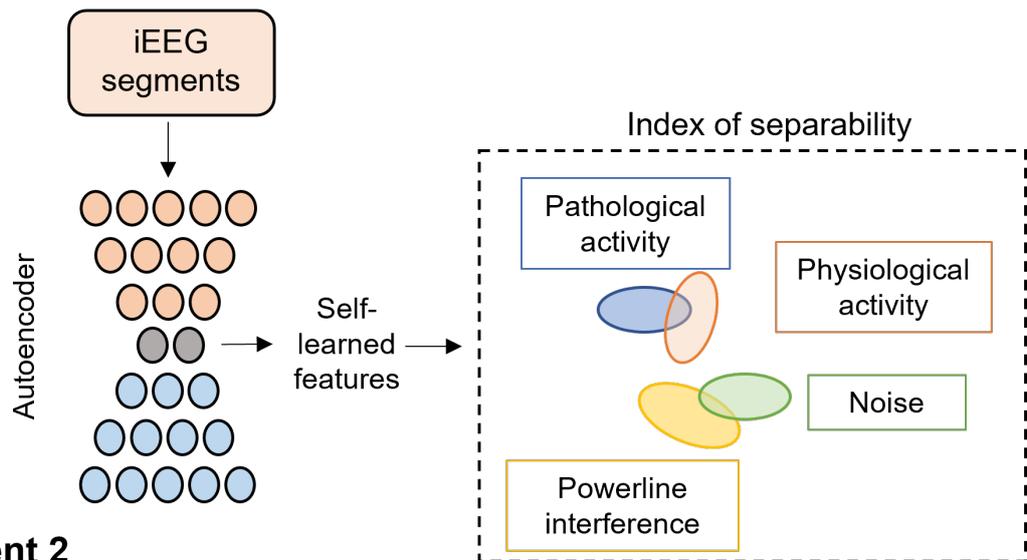
Category	Number of samples in FNUSA dataset	Number of samples in Mayo dataset
SOZ	60,844	25,006
NSOZ	86,186	46,951

Table 3: Number of SOZ and NSOZ samples in FNUSA and Mayo datasets, including only iEEG segments from physiological and pathological category.

### 2.7.3 Experiment 3

Experiment 3 involved statistical testing of the established biomarkers to assess their ability to discriminate the SOZ. All iEEG channels of full-length clinical data were processed by HFO and IS detector, and the related features were computed in 10-second statistical window. Relative entropy was estimated in every second only in channels with adjacent contact. The mean values per channel were used for statistical testing of all biomarkers. Experiment 3 involved analysing the potential of each feature to discriminate between SOZ and NSOZ from two perspectives. One approach evaluated feature statistics for all patients combined. The second approach statistically tested the feature distinguishing capability for each individual patient. The specific statistical tests used for evaluation are described in section 2.6. The evaluation scheme is shown in Figure 12. The AE embedded features do not have a clear physical interpretation. Therefore, they need to be evaluated as a whole set and were not statistically tested as individual features.

## Experiment 1



## Experiment 2

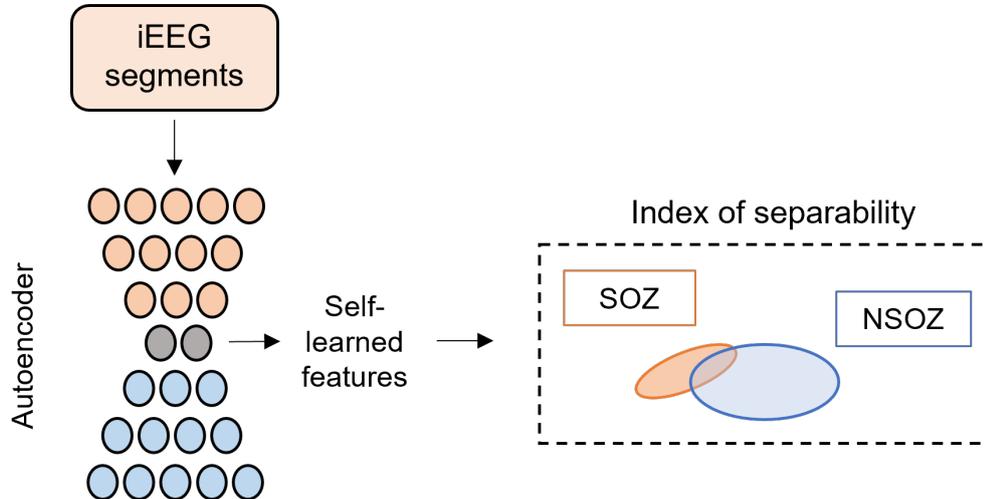
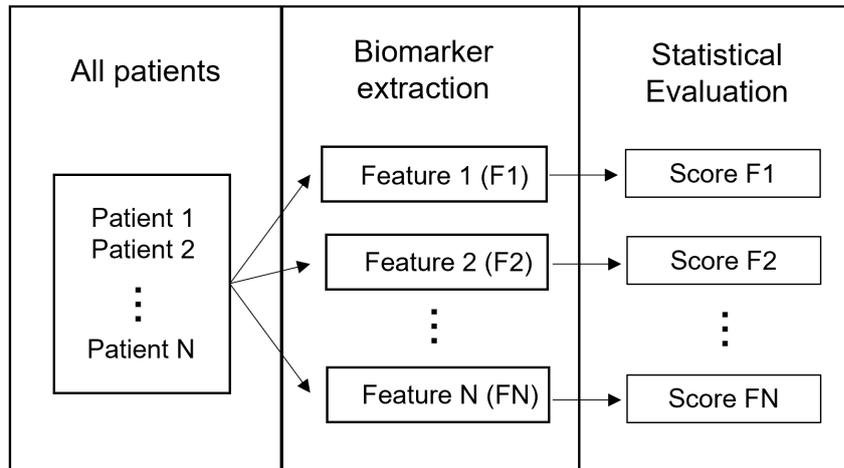


Figure 11: Graphical illustration of Experiment 1 and Experiment 2. Autoencoder model processes the input iEEG segments extracting self-learned embedded features and visualising them with corresponding labels.

## A) Aggregate Analysis



## B) Individual Analysis

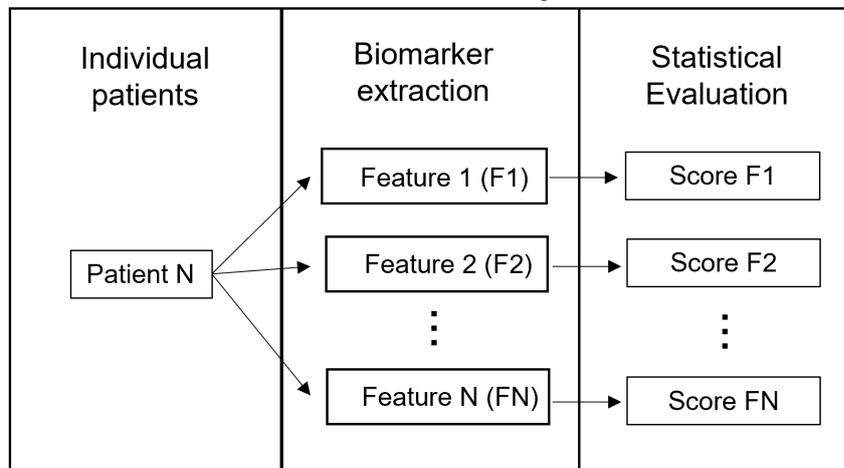


Figure 12: Schematic illustration of two perspectives in Experiment 3. The top scheme depicts a statistical analysis of each biomarker feature of all patients combined. The bottom scheme represents each biomarker feature statistically evaluated for every patient separately.

## 2.7.4 Experiment 4

Experiment 4 focused on comparing the set of AE embedded features against the established biomarkers to differentiate between SOZ and NSOZ channels on full-length iEEG data. For comparison purposes, three separate Naive Bayes classification models were trained with either AE embedded features, biomarkers, or a combination of both. The pipeline of Experiment 4 is shown in Figure 13. All signals from the full-length iEEG dataset were subjected to feature extraction by the autoencoder and biomarker detector. Since some of the extracted features were computed within different window sizes, all features were averaged to correspond to a 10-minute segment to ensure shape uniformity for all analysed channels. Extracted features were transformed into PCA components explaining at least 85% of variance to reduce dimensionality. Each Naive Bayes classifier was trained with segments from all but one patient, and the segments for the remaining patient were used as a test set. Remaining cognizant of the imbalanced dataset, the train and test sets were separately balanced to compensate for the unequal representation of SOZ and NSOZ channels. The performance of each model was evaluated with F1, AUROC, and AP scores. The Wilcoxon paired signed-rank test was used to determine whether models' performances were statistically different from each other.

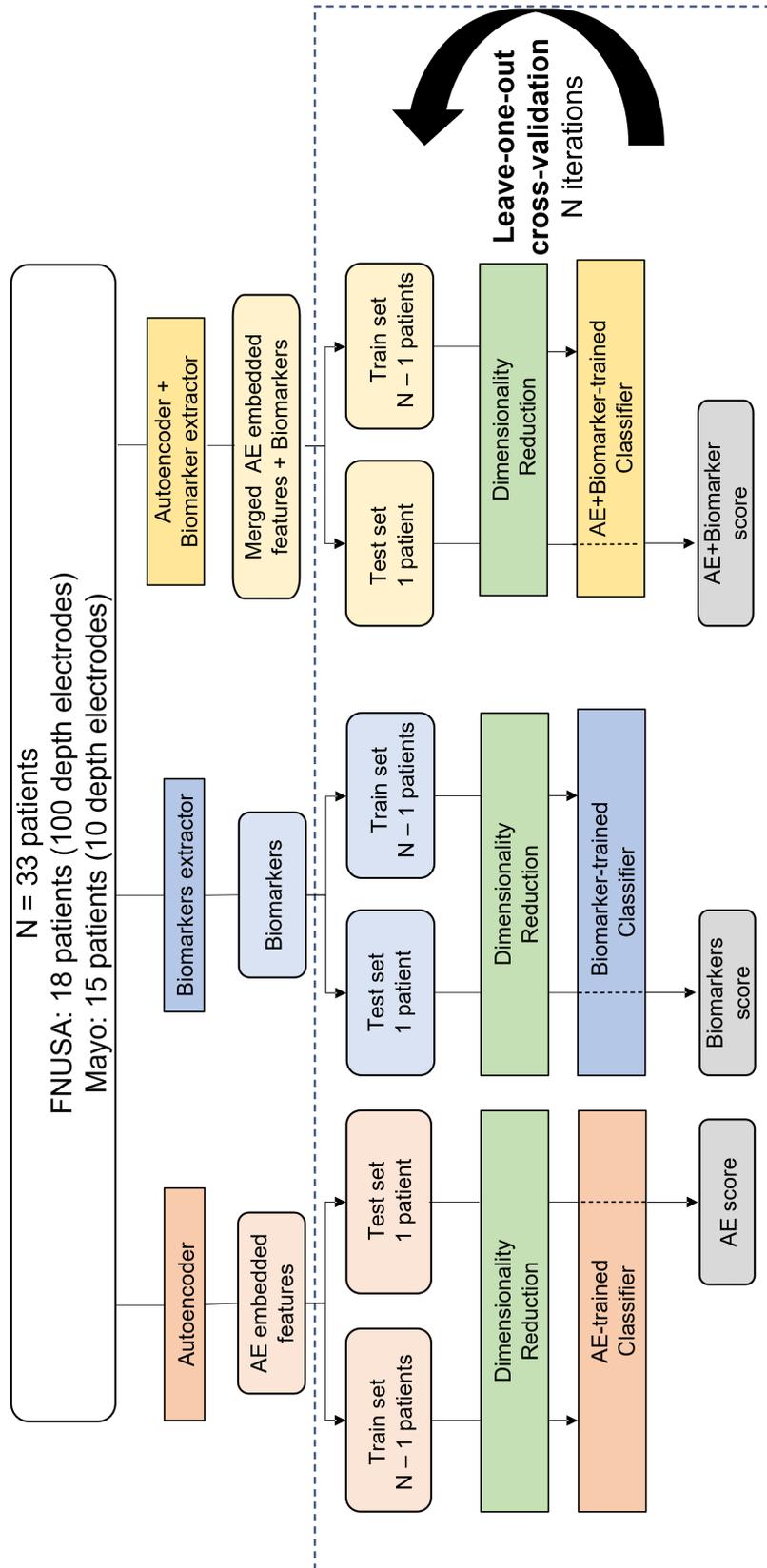


Figure 13: Framework of Experiment 4 for evaluation of performance of AE-trained and Biomarker-trained models.

## 3 Results

The following section presents the results of the experiments conducted in this thesis. The findings on each specific question are stated in separate subsections and further described in the Discussion.

### 3.1 Experiment 1

**Specific question 1: “Are self-learned embedded features extracted by AE trained using 1-second iEEG segments capable of distinguishing between physiological, pathological, noise and powerline interference segments?”**

Experiment 1 investigated the capability of AE embedded features to separate iEEG segments with different characteristics.

The experiment was conducted separately for the AE-Mayo and AE-FNUSA models. Two dimensionality reduction techniques (PCA and UMAP) were implemented within the experiment. The F1 scores of the naive Bayes classifier and the total processing time using both techniques were evaluated. The trade-off between classifier performance and processing time was evaluated for multiple testing setups, and only one dimensionality reduction method was selected for the following experiments. Each experiment applying UMAP and PCA techniques was repeated five times and ten times, respectively. The mean F1 score and duration from all trials averaged across the category pairs are presented in Table 4 and Table 5.

Model	N comp. PCA	Dur. mean $\pm$ std [s]	F1 mean $\pm$ std
AE-FNUSA	17	298.5 $\pm$ 15.6	0.952 $\pm$ 0.001
AE-Mayo	16	171.5 $\pm$ 2.6	0.940 $\pm$ 0.003

Table 4: Index of separability among different iEEG signal categories determined by mean F1 score from ten trials using PCA method for dimensionality reduction. The mean duration of computational processes is shown in the third column.

Table 4 shows the results of the AE-FNUSA and AE-Mayo models employing the PCA method for dimensionality reduction. The criterium of 85% of explained variance determined the optimal number of PCA components to be 17 for the AE-FNUSA model and 16 for the AE-Mayo model. The processing time varied based on the different number of samples in each dataset. Both models achieved a separability index of more than 90%.

<b>AE-FNUSA</b>		
N comp. UMAP	Dur. mean $\pm$ std [s]	F1 mean $\pm$ std
100	4824.7 $\pm$ 32.7	0.862 $\pm$ 0.019
50	2853.5 $\pm$ 76.5	0.832 $\pm$ 0.069
20	1285.9 $\pm$ 26.6	0.755 $\pm$ 0.124
10	843.2 $\pm$ 2.8	0.884 $\pm$ 0.013
5	651.8 $\pm$ 5.1	0.884 $\pm$ 0.011
2	513.4 $\pm$ 1.1	0.881 $\pm$ 0.011
<b>AE-Mayo</b>		
N comp. UMAP	Dur. mean $\pm$ std [s]	F1 mean $\pm$ std
100	3486.4 $\pm$ 14.3	0.899 $\pm$ 0.006
50	1838.3 $\pm$ 17.9	0.855 $\pm$ 0.046
20	960.4 $\pm$ 76.9	0.775 $\pm$ 0.033
10	615.0 $\pm$ 20.0	0.861 $\pm$ 0.163
5	462.8 $\pm$ 20.1	0.914 $\pm$ 0.001
2	360.5 $\pm$ 13.1	0.910 $\pm$ 0.002

Table 5: Index of separability among different iEEG signal categories determined by mean F1 score from five trials implementing UMAP method to reduce dimensionality with different number of embedded components.

Table 5 presents the results of both AE-FNUSA and AE-Mayo models from five trials using the UMAP technique. The features extracted from the autoencoder were transformed into the various numbers of embedded components, which are listed in the left column. The middle column indicates the mean processing time of a single trial corresponding to the number of embedded components. The mean index of separability averaged across all category pairs can be seen in the third column.

Overall, applying the PCA method for dimensionality reduction resulted in a higher mean F1 score and significantly shorter processing time compared to the UMAP technique. Therefore, only the PCA method was used in the subsequent experiments.

The ability of AE features to differentiate iEEG segments into separate clusters was visually inspected by plotting two UMAP-transformed embedded components. The colours of the individual data points corresponded to the original labels. An example of the visualisation of the embedded components from features extracted by the AE-FNUSA model is shown in Figure 14.

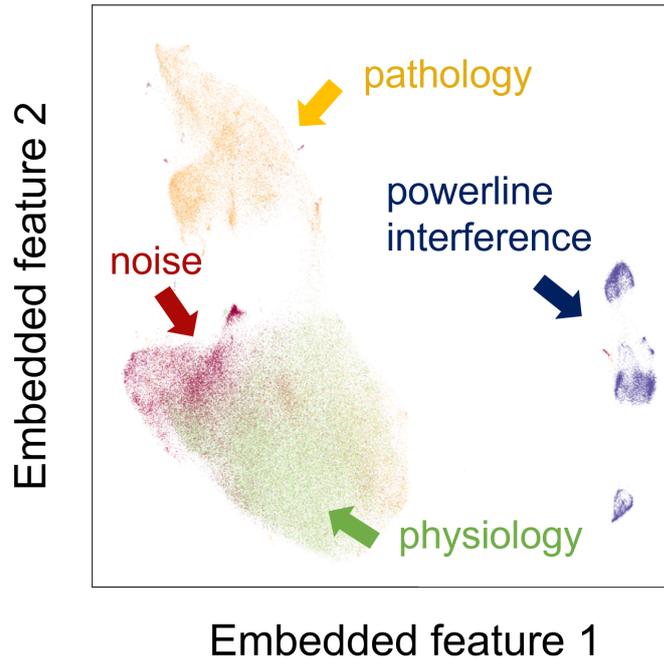


Figure 14: Visualisation of AE embedded features extracted by AE-FNUSA model reduced by UMAP into two dimensions. The embedded features representing physiological (green), pathological (yellow), noise (red), and powerline interference (violet) iEEG segments are coloured according to their original labels.

## 3.2 Experiment 2

**Specific question 2: “Are self-learned embedded features extracted by AE trained using 1-second iEEG segments capable of distinguishing between SOZ and NSOZ segments?”**

Experiment 2 aimed to explore whether the AE embedded features can represent characteristics of iEEG segments arising from SOZ and if they can differentiate them from NSOZ segments. Only pathological and physiological segments were used to ensure consistency with the real scenario, excluding noise. The analysis followed the framework presented in Section 2.7. The PCA method was chosen to reduce dimensionality with the number of components corresponding to 85% of the explained variance. A scree plot exhibiting the cumulative explained variance ratio respective to the number of PCA components is shown in Figure 15.

The experiment was run ten times for both AE-FNUSA and AE-Mayo models. The F1 results of the Naive Bayes classifier obtained as the average of all experiments are shown in Table 6.

Additionally, the interdependence between SOZ and pathology labels was evaluated using F1 and Cohen’s kappa scores. For the Mayo dataset, the achieved F1 and Cohen’s kappa score for the SOZ-pathology pairs was 0.45 and 0.26, respectively. For the FNUSA dataset,

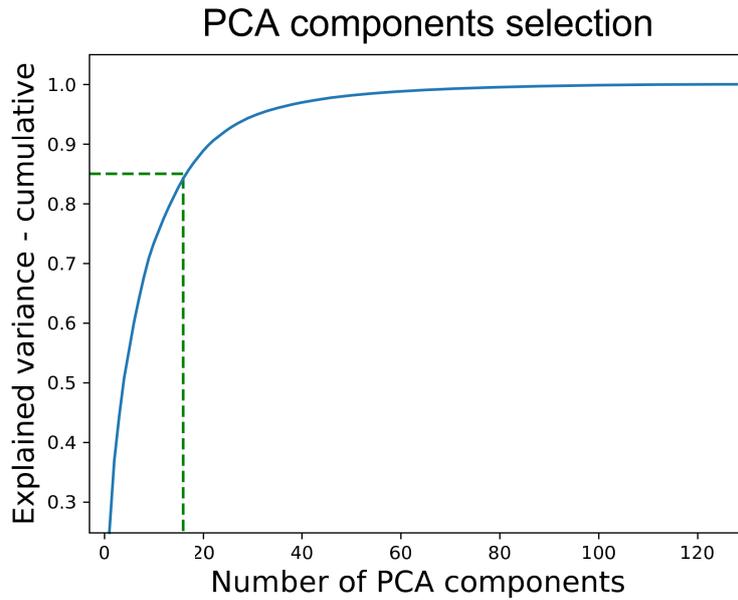


Figure 15: The scree plot visualising the contribution of individual PCA components to the cumulative explained variance using the FNUSA dataset. The selected number of components for the following analysis equalled 17, which corresponded to 85% of the cumulative explained variance ratio.

the F1 and Cohen’s kappa scores were 0.68 and 0.48, respectively. These results indicate that pathological segments did not directly relate to SOZ labeling. Thus, the demonstrated pathological patterns do not necessarily denote seizure-initiating brain tissue.

Model	N comp. PCA	Dur. mean $\pm$ std [s]	F1 mean $\pm$ std
AE-FNUSA	17	608.7 $\pm$ 3.6	0.710 $\pm$ 0.004
AE-Mayo	16	141.6 $\pm$ 1.3	0.720 $\pm$ 0.007

Table 6: Index of separability between SOZ and NSOZ segments of physiological and pathological iEEG signals (noise not included). The mean F1 score was acquired from ten trials using PCA method for dimensionality reduction.

Table 6 summarises the averaged results of the individual trials of Experiment 2. Each row corresponds to the model being evaluated. The second column contains the number of embedded components used for the classifier. The third column displays the mean processing time of each trial. The mean index of separability between SOZ and NSOZ segments is shown in the last column. It can be seen that both models reached similar F1 scores; 0.71 for the AE-FNUSA model and 0.72 for the AE-Mayo model.

The distribution approximated by the AE embedded features visualised by the first two PCA components, and UMAP embedded components can be seen in Figure 16.

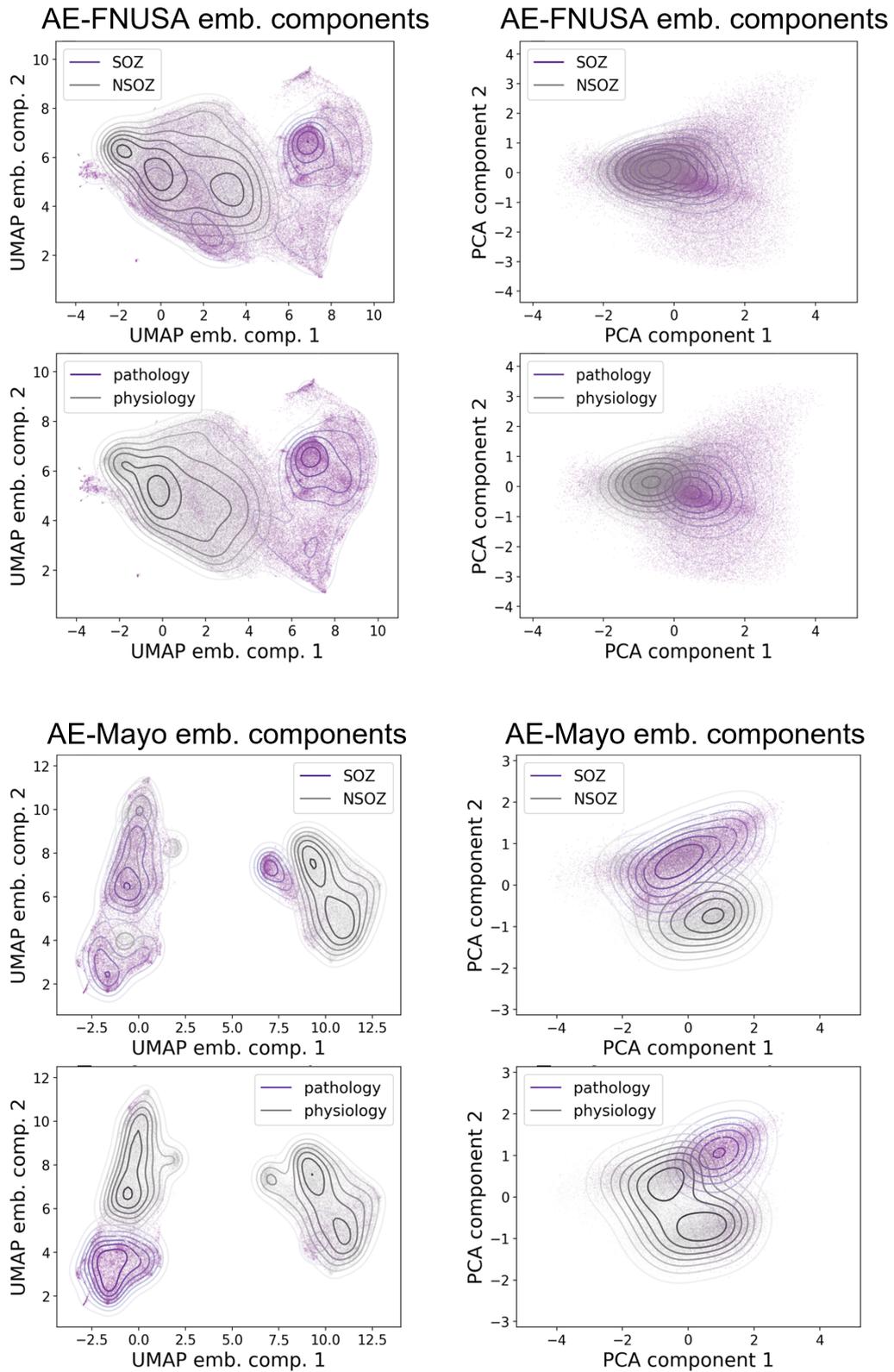


Figure 16: Visualisation of AE embedded components extracted by AE-FNUSA model (top four figures) and AE-Mayo model (bottom four figures). The distributions of SOZ and NSOZ segments (top row) and distributions of physiological and pathological segments (bottom row) are displayed as merged scatter and density plot. The AE embedded components are visualised using UMAP-transformed embedded components (left column) and using the first two PCA embedded components (right column).

### 3.3 Experiment 3

**Specific Question 3: “What is the performance of established biomarker-derived features for SOZ vs NSOZ iEEG discrimination in full-length clinical data (30-120 min) on a) a group level; b) a subject-level?”**

Experiment 3 analysed how conventional biomarkers (HFO, IS, REN) can differentiate between SOZ and NSOZ channels. Individual features were derived from the detection of HFO and IS and statistically tested. In the following section, summary statistics for all tested features described by the AUROC value and the p-value of the Mann-Whitney test are presented. The p-value is delineated in asterisk representation as follows: one asterisk corresponds to a p-value < 0.05; two asterisks: p < 0.01, three asterisks: p < 0.001, four asterisks: p < 0.0001. Non-asterisk value in a result table means the feature showed no statistical difference. The a) section presents the results of testing data from all the patients combined. The b) section shows the feature performance evaluated in each patient separately.

#### a) Group level evaluation

In this section, the potential of each feature to discriminate between SOZ and NSOZ channels is evaluated using AUROC and p-value. The features were tested for both institutional datasets together as well as separately.

The summary of SOZ vs NSOZ testing for HFO-derived features on a group level is presented in Tables 7 for ripples and Table 8 for fast ripples. Figure 17 visualises standardised boxplots of feature distributions for SOZ and NSOZ channels in both HFO bands.

#### • HFO - Ripples (80-200 Hz)

HFO - Ripples (80-200 Hz)				
Institution	Count/10min	Mean amplitude / 10 min	Event amplitude	Duration
Combined	0.64****	0.66****	0.64****	0.6****
FNUSA	0.67****	0.69****	0.69****	0.62****
Mayo	0.55	0.56	0.56	0.55

Table 7: The AUROC values and p-values (asterisk representation) of HFO features detected in ripple band for differentiation between SOZ and NSOZ channels. Statistical testing was performed for all patients together (Combined), for patients only from St. Anne’s Hospital (FNUSA) and for patients only from Mayo Clinic (Mayo).

• HFO - Fast ripples (200-500 Hz)

HFO - Fast ripples (200-500 Hz)				
Institution	Count/10min	Mean amplitude / 10 min	Event amplitude	Duration
Combined	0.58****	0.61****	0.58****	0.6****
FNUSA	0.57**	0.62****	0.59***	0.59****
Mayo	0.65**	0.61*	0.53	0.66**

Table 8: The AUROC values and p-values (asterisk representation) of HFO features detected in fast ripple band for differentiation between SOZ and NSOZ channels. Statistical testing was performed for all patients together (Combined), for patients only from St. Anne's Hospital (FNUSA) and for patients only from Mayo Clinic (Mayo).

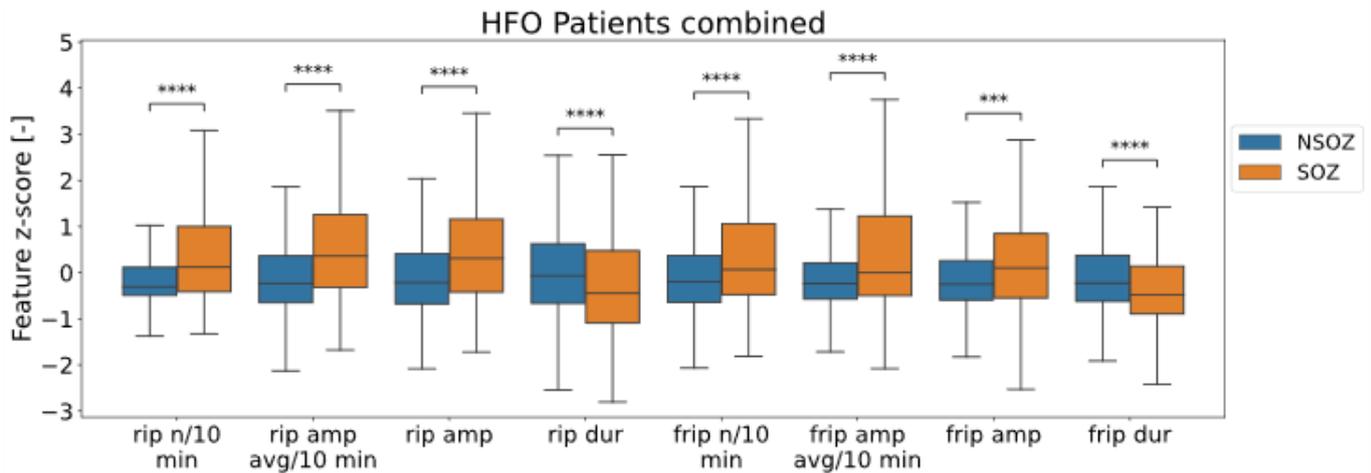


Figure 17: HFO features distributions and statistical differences between SOZ channels (orange) and NSOZ channels (blue) of all patients combined. The meaning of displayed features shortcuts is as follows; rip n/10 min: ripple count/10min, rip amp avg/10 min: ripple mean amplitude /10 min, rip amp: ripple event amplitude, rip dur: ripple duration, frip n/10 min: fast ripple count/10 min, frip amp avg/10min: fast ripple mean amplitude / 10 minutes, frip amp: fast ripple event amplitude, frip dur: fast ripple duration.

All individual features in the ripple and fast ripple band exceeded the AUROC value of 0.5. Table 7 shows that all features evaluated only for FNUSA patients reached higher values than Mayo patients and all patients combined. The highest AUROC score and statistical significance were observed for ripple event amplitude and mean ripple amplitude per 10/minutes. The results of the fast ripple band analysis are equivocal. The fast ripple amplitude performed better on the FNUSA data, whereas fast ripple rate and duration distinguished better on the Mayo dataset. The highest AUROC score was obtained for fast ripple duration in the Mayo dataset, which conversely reached the lowest score in the ripple band. Overall, AUROC values were

higher evaluating patients from a single institution compared to a combination of patients from both institutions.

Looking at Figure 17, it can be concluded that higher values and min-max range were observed in SOZ channels for rate, event amplitude, and mean amplitude features in both HFO bands. On the other hand, HFO duration in both bands was slightly longer in NSOZ channels, and the min-max range was similar to SOZ channels.

• **Interictal spikes**

Table 9 shows the group differentiation performance of the SOZ vs NSOZ channels of each feature related to interictal spikes. Standardised boxplots of the features' statistical characteristics for the SOZ and NSOZ channels are shown in Figure 18.

Interictal spikes							
Institution	Count / 10min	Mean amplitude / 10 min	Event amplitude	Left amplitude	Right amplitude	Left duration	Right duration
Combined	0.64****	0.62****	0.6****	0.6****	0.66****	0.55*	0.5
FNUSA	0.63****	0.6****	0.59***	0.58***	0.66****	0.52	0.52
Mayo	0.67***	0.68***	0.68***	0.71****	0.72****	0.68***	0.58

Table 9: The AUROC values and p-values (asterisk representation) of IS features for differentiation between SOZ and NSOZ channels. Statistical testing was performed for all patients together (Combined), for patients only from St. Anne's Hospital (FNUSA) and for patients only from Mayo Clinic (Mayo).

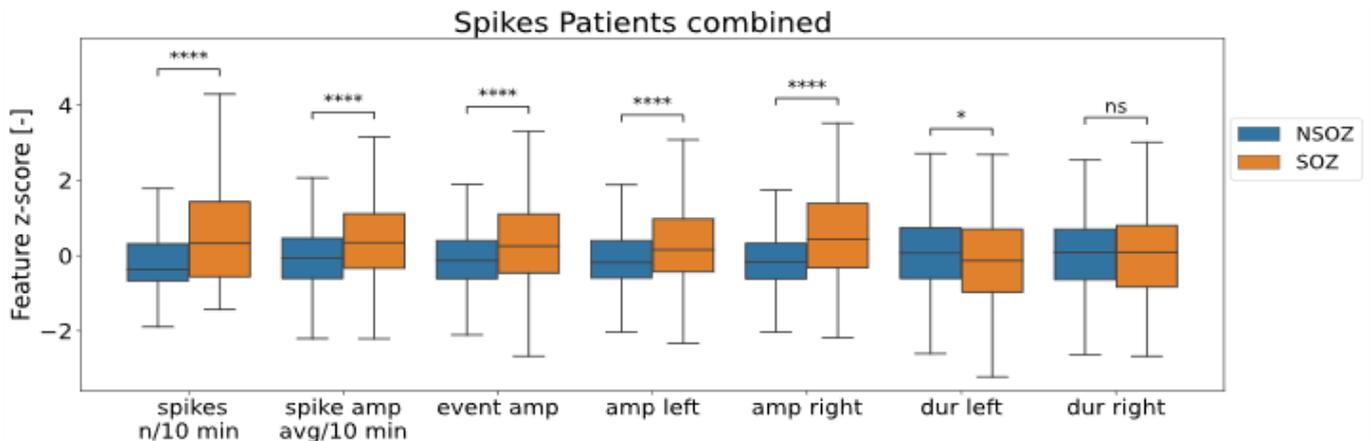
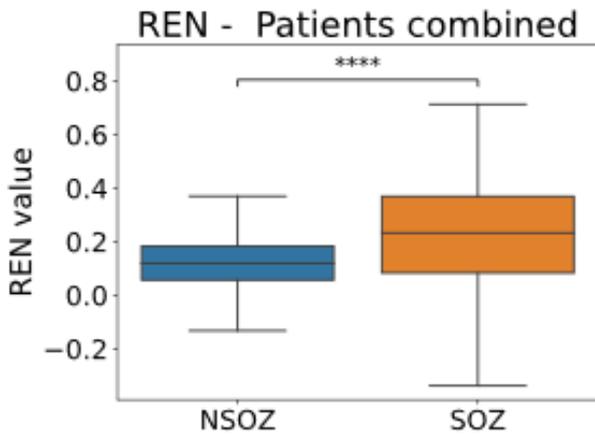


Figure 18: IS features distributions and statistical differences between SOZ channels (orange) and NSOZ channels (blue) of all patients combined. The meaning of displayed features short-cuts is as follows; spikes n/10 min: spike count /10 min, spike amp avg/10 min: spike mean amplitude/10 min, event amp: event amplitude, amp left: left amplitude, amp right: right amplitude, dur left: left duration, dur right: right duration.

In view of institutional comparison, all IS features achieved higher AUROC values on the Mayo dataset than on the FNUSA dataset and a combination of both. The right amplitude function achieved the highest AUROC values at both institutions. The right duration was the lowest performing feature in both datasets. Similar to the HFO features, the rate and amplitude features corresponding to the SOZ channels exhibited larger values and a broader min-max range than the NSOZ channels. Contrarily, duration features were nearly equivalent in both channel types.

• **Relative entropy**

The REN values of SOZ and NSOZ channels are depicted in Figure 19. The SOZ channels manifested higher values and more extensive min-max range, whereas NSOZ REN values appeared lower and more compact. The ability to distinguish between the two types of channels is shown in Table 10. It can be seen that the REN feature performed better on the FNUSA dataset compared to the Mayo and combined datasets.



Institution	Relative entropy
Combined	0.67****
FNUSA	0.75****
Mayo	0.57

Table 10: The AUROC values and statistical significance of REN for discrimination between SOZ and NSOZ channels.

Figure 19: REN statistical distribution of NSOZ channels (left) and SOZ channels (right).

b) **Subject level evaluation**

This section focuses on evaluating the features of each patient separately to assess the potential influence of individuals on the overall statistics. In the same way as in the previous section, individual feature performance was evaluated using AUROC and p-value. Due to the extensive content, detailed tables with results are available in the Appendix (Tables 18, 19,20, 21). The following part summarises the proportion of patients exceeding AUROC of 0.7 and 0.8 for each feature in all datasets (FNUSA – 18 patients, Mayo – 15 patients). Results for the HFO features are shown in Table 11 (ripples) and Table 12 (fast ripples).

• **HFO - Ripples (80-200 Hz)**

The largest proportion of patients achieving AUROC of 0.7 or higher can be observed for the ripple mean amplitude per 10 minutes (76% of all patients, 83% -

FNUSA, 66% - Mayo). Considering an AUROC threshold of 0.8, the best-performing feature was the ripple rate per 10 minutes (55% of all patients, 56% - FNUSA, 53% - Mayo). The fewest patients exceeding both AUROC thresholds was observed for ripple duration across all datasets.

HFO - Ripples (80-200 Hz)				
Institution	Count / 10min	Mean amplitude / 10 min	Event amplitude	Duration
Combined	(23; 18) / 33	(25; 17) / 33	(24; 13) / 33	(15; 11) / 33
FNUSA	(12; 10) / 18	(15; 8) / 18	(13; 7) / 18	(8; 5) / 18
Mayo	(11; 8) / 15	(10; 9) / 15	(11; 6) / 15	(7; 6) / 15

Table 11: Evaluation of the HFO features in the ripple band for SOZ-NSOZ discrimination task for each patient individually. The table represents the number of patients exceeding the AUROC value of 0.7 (first number in the bracket) and 0.8 (second number in the bracket); a) from all patients (combined), b) from St. Anne's Hospital dataset (FNUSA), and c) from Mayo Clinic dataset (Mayo).

• **HFO - Fast ripples (200-500 Hz)**

HFO - Fast ripples (200-500 Hz)				
Institution	Count / 10min	Mean amplitude / 10 min	Event amplitude	Duration
Combined	(21; 12) / 33	(22; 12) / 33	(20; 10) / 33	(24; 11) / 33
FNUSA	(12; 4) / 18	(12; 4) / 18	(11; 4) / 18	(12; 3) / 18
Mayo	(9; 8) / 15	(10; 8) / 15	(9; 6) / 15	(12; 8) / 15

Table 12: Evaluation of the HFO features in the fast ripple band for SOZ-NSOZ discrimination task for each patient individually. The table represents the number of patients exceeding the AUROC value of 0.7 (first number in the bracket) and 0.8 (second number in the bracket); a) from all patients (combined), b) from St. Anne's Hospital dataset (FNUSA), and c) from Mayo Clinic dataset (Mayo).

The fast ripple duration reached AUROC value above 0.7 in the highest number of patients (72% of all patients, 66% - FNUSA, 80% - Mayo). However, raising the threshold to 0.8 significantly decreased the proportion of patients exceeding this threshold to 33% of all patients (17% - FNUSA, 53% - Mayo). A similar trend can be seen for the fast ripple mean amplitude per 10 minutes. AUROC of 0.7 was exceeded by 67% of patients in all datasets, while the proportion decreased to 36% of all patients at the 0.8 threshold (22% - FNUSA, 53% - Mayo). In both cases, a larger decline was observed in the FNUSA dataset.

In general, except for the HFO duration, the ripple features overreached the set threshold in more patients than the fast ripple features.

- **Interictal spikes**

Table 13 summarises the results of individual IS-related features for each patient. The best discriminating feature derived from IS detections was the right amplitude overreaching AUROC of 0.7 in 61% of all patients (50% FNUSA, 73% - Mayo). Likewise, the spike rate per 10 minutes achieved over 0.7 in 61% of all patients (72% - FNUSA, 47% Mayo).

Interictal spikes							
Institution	Count / 10min	Mean amplitude / 10 min	Event amplitude	Left amplitude	Right amplitude	Left duration	Right duration
Combined	(20; 16) / 33	(16; 14) / 33	(16; 12) / 33	(16; 12) / 33	(20; 17) / 33	(15; 12) / 33	(19; 13) / 33
FNUSA	(13; 9) / 18	(6; 4) / 18	(6; 3) / 18	(8; 6) / 18	(9; 7) / 18	(7; 5) / 18	(9; 5) / 18
Mayo	(7; 7) / 15	(10; 10) / 15	(10; 9) / 15	(8; 6) / 15	(11; 10) / 15	(8; 7) / 15	(10; 8) / 15

Table 13: Evaluation of the IS features for SOZ-NSOZ differentiation for each patient separately. The table represents the number of patients exceeding the AUROC value of 0.7 (first number in the bracket) and 0.8 (second number in the bracket); a) from all patients (combined), b) from St. Anne's Hospital dataset (FNUSA), and c) from Mayo Clinic dataset (Mayo).

Increasing the AUROC up to 0.8 did not result in a significant decline of patients exceeding the threshold in any of the two features: right amplitude (52% of all patients, 39% - FNUSA, 67% Mayo), spike rate per 10 minutes (48% of all patients, 50% - FNUSA, 47% Mayo). The worst performance with a 0.7 threshold can be seen for the left duration feature reaching 45% proportion of all patients (39% - FNUSA, 53% - Mayo). The lowest performing features with AUROC set to 0.8 included event amplitude, left amplitude and left duration, all with a proportion of 36% of all patients.

- **Relative entropy**

A summary of the numbers of patients exceeding 0.7 and 0.8 AUROC values for relative entropy is shown in Table 14. The overall number of patients achieving above the 0.7 threshold was 73% (83% - FNUSA, 60% Mayo). For AUROC threshold of 0.8 the proportions reached 61% for all patients (72% - FNUSA, 47% Mayo).

Institution	Relative entropy
Combined	(24; 20) / 33
FNUSA	(15; 13) / 18
Mayo	(9; 7) / 15

Table 14: Summary of REN evaluation for SOZ-NSOZ discrimination for each patient separately. The table represents the number of patients exceeding the AUROC value of 0.7 (first number in the bracket) and 0.8 (second number in the bracket); a) from all patients (combined), b) from St. Anne’s Hospital dataset (FNUSA), and c) from Mayo Clinic dataset (Mayo).

Considering all presented biomarkers, the best overall performance for the 0.7 threshold was achieved for the ripple mean amplitude per 10 minutes. For the threshold of 0.8, the relative entropy showed better performance compared to the rest of the biomarkers.

### 3.4 Experiment 4

**Specific Question 4: “Do self-learned embedded features extracted by AE outperform conventional biomarker-derived features in discriminating SOZ and NSOZ channels?”**

The aim of Experiment 4 was to assess whether AE features can better differentiate between SOZ and NSOZ channels than biomarkers. Comparison were made using classification performances (F1, AUROC, AP) between naive Bayes classifiers trained on a specific set of features (Biomarker-trained, AE-trained, AE+Biomarker-trained). Classification was performed separately on the FNUSA and Mayo datasets. A detailed overview of the performances of each test patient can be seen in Appendix (Tables 22, 23, 24). For the Biomarker-trained classifier, the lowest F1 score was 0.11, and the highest F1 score was 0.84. The minimum and maximum for AUROC were 0.35 and 0.85. The minimal AP was 0.44, and the maximal AP equalled 0.82. Similar fluctuations were also observed for the AE-trained classifier. The min-max range for F1 score was 0-0.86, for AUROC 0.33-0.83 and for AP 0.44- 0.78.

Table 15 shows the mean classification scores of leave-one-patient-out cross-validation. For the FNUSA dataset, the highest mean F1 score was reached for the AE-trained classifier. The highest mean AUROC and AP scores were achieved with the Biomarker-trained classifier. Looking at the Mayo dataset, the highest mean F1 and AUROC scores were reached using the merged AE+Biomarker-trained classifier. The mean AP was the same for the AE-trained and AE+Biomarker-trained classifier.

The leave-one-out performances of AE-trained, Biomarker-trained, and merged classifiers were tested for statistical differences using the Wilcoxon paired test with a significance level of  $\alpha = 0.05$ . No statistical difference was observed between any pair of classifiers in F1 score, AUROC or AP metric.

Instit.	Biomarker-trained classifier			AE-trained classifier			AE+Biomarker-trained classifier		
	F1 mean	AUROC mean	AP mean	F1 mean	AUROC mean	AP mean	F1 mean	AUROC mean	AP mean
FNUSA	0.60	0.70	0.67	0.66	0.68	0.63	0.54	0.67	0.64
Mayo	0.54	0.54	0.53	0.44	0.56	0.55	0.55	0.57	0.55

Table 15: The mean F1, AUROC, and AP scores of leave-one-patient-out cross-validation. The first three columns show the results of the classifier trained only on Biomarkers. The middle three columns demonstrate the results of the classifier trained only with AE features. The last three columns present results of the merged classifier incorporating both AE and Biomarkers features. The first row represents the mean scores on the FNUSA dataset. The second row represents mean scores on the Mayo dataset.

The maximal performances of the three classifiers (AE-trained, Biomarker-trained, AE+Biomarker-trained) for each patient were aggregated to determine the SOZ vs NSOZ discrimination potential in the case of selecting an optimal feature set and classifier. The average F1, AUROC, and AP scores for the best-performing classifiers achieved an increased classification score of 0.66, 0.64, and 0.60 for the Mayo dataset and 0.74, 0.75, and 0.71 for the FNUSA dataset. Further details are provided in the Appendix (Table 25).

## 4 Discussion and Conclusions

Accurate localisation of brain seizure onset zone is a critical task for successful epilepsy surgery. However, the vast amount of data collected within pre-operative monitoring, which is manually checked by physicians, is time-consuming and may lead to a human error. Implementation of automated algorithms could aid the data review process. The main objective of this thesis was to explore the feasibility of the in-house unsupervised Conv-GRU autoencoder to distinguish between SOZ and NSOZ channels in pre-surgical iEEG recordings. Four experiments were conducted to answer the research questions stated in section 1.7. The findings achieved during the experiments, the contributions and limitations of this thesis are discussed in this section.

**Specific Question 1: “Are self-learned embedded features extracted by AE trained using 1-second iEEG segments capable of distinguishing between physiological, pathological, noise and powerline interference segments?”**

The Experiment 1 examined two separate AE models for two datasets (one for each) and their capability to differentiate one-second iEEG segments of various categories. The embedded features produced by the autoencoders were transformed into lower dimensions using either UMAP or PCA technique. The performance of the embedded features was evaluated using the F1 score of the Naive Bayes classifier.

Classification using the UMAP method with different numbers of embedded components showed that a higher number of embedded components did not lead to better classification performance (Table 5). A similar trend was observed in both models. The highest F1 score was achieved when reducing AE features to five embedded components (AE-FNUSA - 0.884, AE-Mayo - 0.914). The lowest F1 score was seen when the AE features were transformed to 20 embedded components. Better performance was achieved when using a smaller number of embedded components (5 or 2) than when using a higher number of components. These results lead to the conclusion that a lower-dimensional feature space, which is computationally less demanding, can be used for classification without significant performance loss.

Figure 14 shows, the powerline interference was the most distinguishable category. That corresponds to the uniform appearance of spectrograms derived from the iEEG segments containing principally a single line emphasizing the dominant frequency of the interference. It is also in agreement with the observed F1 score, which was higher when powerline interference was present in the classification pair. The remaining categories exhibited certain overlaps; however, the overall separability appeared visually satisfactory.

The mean achieved F1 scores was 0.952 for the AE-FNUSA model and 0.940 for the AE-Mayo model (Table 4) This confirms that in-house Conv-GRU autoencoder architecture was efficiently designed to separate different types of iEEG signals.

**Specific Question 2: “Are self-learned embedded features extracted by AE trained using 1-second iEEG segments capable of distinguishing between SOZ and NSOZ segments?”**

The AE models were investigated in terms of SOZ vs NSOZ differentiation task on one-second iEEG segments. As aforementioned, only the PCA method was used for dimensionality reduction.

The visualisation of the first two PCA components and UMAP components (only for visualisation purposes) (Fig. 16) shows lower separability between segments than in Experiment 1 (Fig. 14). However, the two-dimensional visualisations do not account for all PCA components used to assess separability and are therefore only for approximate illustrative purposes. It can be seen that some embedded components representing SOZ segments were separable from embedded components belonging to the NSOZ. Comparison of the top and bottom rows in the figure shows that some pathological segments corresponded to SOZ segments. Based on the visual assessment, the Cohen’s kappa score was calculated between the SOZ and pathological labels to determine the label agreement. The resulting score showed differences between the labels, confirming the SOZ labels were not directly assigned to all pathological segments.

Both models reached similar F1 scores; AE-FNUSA achieved 0.710 and AE-Mayo achieved 0.720. These results suggest that unsupervised Conv-GRU AE has the potential to capture SOZ-related characteristics from the iEEG signal. Although the obtained results reached a lower F1 score than in Experiment 1, it can be assumed that a second-long time frame is not sufficiently representative for a complete characterisation of the SOZ.

**Specific question 3: “What is the performance of established biomarker-derived features for SOZ vs NSOZ iEEG discrimination in full-length clinical data (30-120 min) on a) a group level; b) a subject-level?”**

Statistical analysis of the full-length clinical recordings was performed within Experiment 3. The capability to differentiate between SOZ and NSOZ was evaluated using the AUROC score for each biomarker (HFO, IS, REN). The evaluation was performed on the subject and group level.

**a) Group level evaluation**

All HFO-derived features in the combined dataset (FNUSA and Mayo) showed a statistical difference between SOZ and NSOZ channels (Tables 7, 8). This confirms that HFO is a biomarker that can help in discriminating SOZ. The features representing the SOZ channels reached higher values compared to NSOZ channels (except for the duration feature) (see Fig 17). Nevertheless, the corresponding AUROC values ranging between 0.58 and

0.66 indicate that distinction between SOZ and NSOZ channels based on HFO feature values alone is insufficient.

Moreover, variable performance was observed between the FNUSA and Mayo datasets. For example, all ripple-derived features achieved higher AUROC in the FNUSA dataset compared to the Mayo dataset. None of the features in the Mayo dataset showed statistical significance (Table 7). However, the relatively low number of samples (channels) available for statistical evaluation in the Mayo dataset could be a biasing factor (see Appendix, Table 17).

The performance of the IS-derived features showed some similarities to the HFO performance. All features, except for the right duration, were statistically different in SOZ and NSOZ channels. The higher feature values were observed in the SOZ channels (Table 9). With respect to statistically significant features, AUROC ranged from 0.55 to 0.66. Thus, IS features alone cannot clearly separate SOZ from NSOZ channels despite the statistical difference. All IS features reached a higher AUROC in the Mayo dataset when comparing performance between institutions.

A possible reason why HFO features performed better in the FNUSA dataset and IS features in the Mayo dataset can be a different patient aetiology. Another reason might be a variability in the medical procedures preceding the iEEG acquisition between the institutes. Different electrode dimensions and brain onset locations may also be factors influencing this difference. Since different acquisition systems were used for recording in each institution, the collected iEEG signals might vary in some attributes. Therefore, a particular detector could work better on one type of signals than on another and affect the features derived from the detections. Different recording times and behavioural states can also play a role in event detections. A further reason could simply be that patients in one institution had more events present than patients from the other institution.

Relative entropy was the only feature estimated directly from the iEEG signal. Observed REN values were higher in the SOZ channels and statistically different from NSOZ channels (see Fig. 19 and Table 10). The achieved AUROC was 0.67, 0.75 and 0.57 for the combined, FNUSA and Mayo datasets, respectively. Looking at the performance within the FNUSA dataset, REN seems to be the most distinguishing biomarker feature. However, the same cannot be concluded about the Mayo dataset with a relatively low AUROC and no statistical difference between SOZ and NSOZ.

An advantage of REN is its direct calculation from the iEEG signal, so the outcome is independent of the event detections. A drawback of this feature might be its bivariate nature. The values estimated between two channels could be affected by the size and the distance of the recording electrodes. Furthermore, the values for each channel were calculated as an average of the adjacent electrodes and thus did not correspond exactly to the individual channels.

Even though differences between SOZ and NSOZ channels were observed, the discrim-

ination capability of the individual biomarkers did not appear to be sufficient for clinical use. However, it can be hypothesized that optimal utilisation of multiple biomarkers could result in better performance than using each biomarker-derived feature individually.

The different performance of the features on both datasets leads to assumptions that the feature performance is rather patient-specific. Therefore, the biomarkers were also evaluated in individual patients.

#### b) **Subject level evaluation**

The AUROC score and the statistical difference for SOZ vs NSOZ discrimination were evaluated for each patient separately based on assumptions about inter-patient variability in feature performance (see Appendix Tables 18, 19, 20, 21).

In this section, the subject-specific performance of individual features derived from HFO and IS were averaged to a single AUROC value, further referred to as mean HFO and mean IS.

The individual-based approach confirmed differences in feature performance between all patients and all biomarkers. For example, patient 6 underperformed in all biomarkers (mean HFO = 0.65, mean IS = 0.59, REN = 0.47). In contrast, patient 453 showed excellent performance for all biomarkers (mean HFO = 0.96, mean IS = 0.99, REN = 1). Mixed performance of individual biomarkers was observed in most of the patients, for instance in patient 40 (mean HFO = 0.64, mean IS = 0.58, REN = 0.9).

Overall, better performance of individual biomarkers was observed at the subject level compared to the group level assessment. Therefore, a more patient-specific approach based on biomarker-derived features might be beneficial to distinguish between SOZ and NSOZ channels.

#### **Specific Question 4: “Do self-learned embedded features extracted by AE outperform conventional biomarker-derived features in discriminating SOZ and NSOZ channels?”**

AE embedded features were compared against established biomarker-derived features on full-length clinical data. Leave-one-out cross-validation was used to compare the performance of Naive Bayes classifiers (F1, AUROC, AP) trained on a particular set of features. The resulting scores for each patient were tested using paired Wilcoxon rank-sum test to evaluate the statistical difference between the classifier performance.

The average score for all patients achieved better than chance classification performance (see Table 22). However, the overall ability to discriminate between SOZ and NSOZ segments was not ideal. To be noted, the classification was used for the purpose of comparing the sets of features. Therefore, no optimisation was carried out to improve the classification performance.

Large fluctuations were observed between classifier performances. In some cases, one classifier performed poorly, while the other classifier achieved good performance within the same

patient. For example, patient 60 reached an F1 score of 0.28 in the Biomarker-trained classification, whereas in the AE-trained classification, the achieved F1 score was 0.82. Merged classifier (AE+Biomarker-trained classifier) was introduced to investigate whether a combination of the two feature sets improves the SOZ vs NSOZ discriminability. However, simply combining of the features within the merged classifier did not lead to better SOZ vs NSOZ differentiation.

The Wilcoxon rank-sum test showed no statistical difference between the performances of the individual classifiers. Given the performance of biomarkers that are currently the baseline, no statistical difference between the classifiers indicates that AE embedded features might have equal potential for SOZ discrimination in clinical practice. Moreover, selecting the best performance among the three classifiers for each patient (see Table 25 in Appendix) resulted in average higher scores. Therefore, it can be assumed that further optimisation of a method combining unsupervised techniques and engineered biomarkers can enhance the current process of SOZ determination.

Several factors could affect the resulting scores in the analysis. The recorded iEEG signals are patient-specific (location of SOZ, electrode placement, type of epilepsy), which may cause diverse distribution of SOZ and NSOZ data in the feature space for each patient. Therefore, merging data from multiple patients may lead to overlapping features representing the SOZ and NSOZ segments. In addition, the relatively low number of patients only allowed for a small training sample size, which could cause classifier underfitting. The difference between the number of implanted electrodes in the SOZ and NSOZ channels could also play a role in the classification performance.

### **Limitations**

The relatively low number of subject samples is a general limitation of studies dealing with iEEG analysis due to the invasiveness, cost and time consumption of the data acquisition method. Another problematic issue is obtaining the gold-standard labels for the SOZ and NSOZ channels. As already mentioned, the iEEG channel designation may be biased by the subjective decision of a reviewer. Furthermore, it is impossible to cover the entire target brain tissue with the electrodes. Therefore the labelled location of the SOZ electrodes may not completely correspond to the anatomical region of the SOZ.

The disadvantage of using embedded features extracted by an unsupervised autoencoder can be seen in the lack of physical interpretation. The techniques applied for dimensionality reduction might reduce the information carried in the originally extracted features. Although engineered biomarkers (HFO, IS) are interpretable in terms of electrophysiology, their successful utilisation depends on the designed detectors and the presence of the events during recording.

### **Contributions & Outline**

Majority of the published studies related to deep learning methods in epilepsy used scalp EEG datasets which do not carry information about deeper brain structures. Most authors who did analyse intracranial data focused on iEEG processing or seizure classification but did not re-

search the differentiation of the SOZ. This thesis contributes to existing knowledge by investigating the feasibility of an unsupervised deep autoencoder-based method to distinguish between iEEG signals recorded from the SOZ and NSOZ brain regions.

The advantage of unsupervised methods is extracting the information from the data itself. Therefore, applying such methods might gain additional aspects of SOZ characteristics that are not covered by engineered biomarkers. Even though the classification performance of AE embedded features did not reach excellent results, the estimated potential to enhance the clinical process of SOZ assessment was comparable to the currently established methods. Further research is needed to develop an optimal strategy for selecting the most discriminating features characterising SOZ in individual patients.

More specifically, the following points will be pursued in the follow-up research. The algorithm optimally combining AE embedded features and biomarkers for separability of SOZ vs NSOZ will be investigated in the follow-up work. Further endeavours will be made to deploy the algorithm within a prospective study. Moreover, the algorithm will be examined on long-term iEEG data to explore long-term trends and their relation to the performance of the unsupervised algorithm.

This thesis validated that an unsupervised Conv-GRU autoencoder can be applied as a tool for sorting iEEG segments into different categories. Furthermore, this thesis unravels the feasibility of using the unsupervised Conv-GRU autoencoder for SOZ vs NSOZ iEEG discrimination and shows that deep autoencoder achieves similar performance as established biomarkers. The results presented in this thesis suggest that the combination of biomarker-derived features with AE-based features may further increase the differentiation capability for determining SOZ. Moreover, the findings of this thesis indicate that a combination of AE-based features and engineered biomarkers could lead to the development of systems for unsupervised iEEG analysis. Additional studies will be needed to explore the possibilities of using unsupervised deep-learning-based methods together with engineered biomarkers to reinforce the clinical support during the SOZ evaluation process.

# Bibliography

- [1] F. Mivalt, V. Kremen, V. Sladky, I. Balzekas, P. Nejedly, N. M. Gregg, B. N. Lundstrom, K. Lepkova, T. Pridalova, B. H. Brinkmann, P. Jurak, J. J. V. Gompel, K. Miller, T. Denison, E. K. S. Louis, and G. A. Worrell, "Electrical brain stimulation and continuous behavioral state tracking in ambulatory humans," *Journal of Neural Engineering*, vol. 19, 2 2022.
- [2] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. The MIT Press, 2016, vol. 800 p.
- [3] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep learning techniques for medical image segmentation: Achievements and challenges," *Journal of Digital Imaging*, vol. 32, pp. 582–596, 8 2019.
- [4] G. Eraslan, Žiga Avsec, J. Gagneur, and F. J. Theis, "Deep learning: new computational modelling techniques for genomics," *Nature Reviews Genetics*, vol. 20, pp. 389–403, 7 2019.
- [5] P. Nejedly, V. Kremen, V. Sladky, M. Nasser, H. Guragain, P. Klimes, J. Cimbalnik, Y. Varatharajah, B. H. Brinkmann, and G. A. Worrell, "Deep-learning for seizure forecasting in canines with epilepsy," *Journal of Neural Engineering*, vol. 16, 2019.
- [6] "Epilepsy," 6 2019. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/epilepsy>
- [7] *Atlas Epilepsy Care in the World 2005*. World Health Organization, 2005, vol. 91 p.
- [8] 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, 4 2005.
- [9] R. S. Fisher, "The new classification of seizures by the international league against epilepsy 2017," *Current Neurology and Neuroscience Reports*, vol. 17, p. 48, 6 2017.
- [10] J. J. Falco-Walter, I. E. Scheffer, and R. S. Fisher, "The new definition and classification of seizures and epilepsy," *Epilepsy Research*, vol. 139, pp. 73–79, 1 2018.
- [11] D. Schmidt, "Drug treatment of epilepsy: Options and limitations," *Epilepsy Behavior*, vol. 15, pp. 56–65, 5 2009.

- [12] L. Dalic and M. Cook, "Managing drug-resistant epilepsy: challenges and solutions," *Neuropsychiatric Disease and Treatment*, vol. Volume 12, pp. 2605–2616, 10 2016.
- [13] J. W. Miller and S. Hakimian, "Surgical treatment of epilepsy," *CONTINUUM: Lifelong Learning in Neurology*, vol. 19, pp. 730–742, 6 2013.
- [14] P. Ryvlin and S. Rheims, "Epilepsy surgery: eligibility criteria and presurgical evaluation," *Dialogues in Clinical Neuroscience*, vol. 10, pp. 91–103, 3 2008.
- [15] F. Rosenow, "Presurgical evaluation of epilepsy," *Brain*, vol. 124, pp. 1683–1700, 9 2001.
- [16] E. S. Louis, L. Frey, J. Britton, J. Hopp, P. Korb, M. Koubeissi, W. Lievens, and E. Pestana-Knight, *Electroencephalography (EEG): An Introductory Text and Atlas of Normal and Abnormal Findings in Adults, Children, and Infants*. American Epilepsy Society, 2016. [Online]. Available: <https://www.ncbi.nlm.nih.gov/books/NBK390354/>
- [17] S. Kovac, V. N. Vakharia, C. Scott, and B. Diehl, "Invasive epilepsy surgery evaluation," *Seizure*, vol. 44, pp. 125–136, 1 2017.
- [18] F. Cardinale, G. Casaceli, F. Raneri, J. Miller, and G. L. Russo, "Implantation of stereoelectroencephalography electrodes," *Journal of Clinical Neurophysiology*, vol. 33, pp. 490–502, 12 2016.
- [19] G. Li, S. Jiang, S. E. Paraskevopoulou, M. Wang, Y. Xu, Z. Wu, L. Chen, D. Zhang, and G. Schalk, "Optimal referencing for stereo-electroencephalographic (seeg) recordings," *NeuroImage*, vol. 183, pp. 327–335, 12 2018.
- [20] A. Shah and S. Mittal, "Invasive electroencephalography monitoring: Indications and presurgical planning," *Annals of Indian Academy of Neurology*, vol. 17, p. 89, 2014.
- [21] Renishaw, "Dixi medical seeg electrode banner." [Online]. Available: <https://resources.renishaw.com/gen/details/dixi-medical-seeg-electrode-banner--111671>
- [22] C. S. Nayak and A. C. Anilkumar, *EEG Normal Waveforms*. StatPearls Publishing, Treasure Island (FL), 2021. [Online]. Available: <http://europepmc.org/books/NBK539805>
- [23] G. Worrell and J. Gotman, "News amp; views in ... biomarkers in medicine," *Biomarkers in Medicine*, vol. 5, pp. 83–86, 2 2011.
- [24] S. Noachtar and J. Rémi, "The role of eeg in epilepsy: A critical review," *Epilepsy Behavior*, vol. 15, pp. 22–33, 5 2009.
- [25] M. de Curtis and G. Avanzini, "Interictal spikes in focal epileptogenesis," *Progress in Neurobiology*, vol. 63, pp. 541–567, 4 2001.
- [26] M. Pail, J. Cimbálník, R. Roman, P. Daniel, D. J. Shaw, J. Chrastina, and M. Brázdil, "High frequency oscillations in epileptic and non-epileptic human hippocampus during a cognitive task," *Scientific Reports*, vol. 10, 12 2020.

- [27] J. Cimbalnik, B. Brinkmann, V. Kremen, P. Jurak, B. Berry, J. V. Gompel, M. Stead, and G. Worrell, "Physiological and pathological high frequency oscillations in focal epilepsy," *Annals of Clinical and Translational Neurology*, vol. 5, pp. 1062–1076, 9 2018.
- [28] J. Cimbalnik, P. Klimes, V. Sladky, P. Nejedly, P. Jurak, M. Pail, R. Roman, P. Daniel, H. Guragain, B. Brinkmann, M. Brazdil, and G. Worrell, "Multi-feature localization of epileptic foci from interictal, intracranial eeg," *Clinical Neurophysiology*, vol. 130, pp. 1945–1953, 10 2019.
- [29] P. Klimes, J. Cimbalnik, M. Brazdil, J. Hall, F. Dubeau, J. Gotman, and B. Frauscher, "Nrem sleep is the state of vigilance that best identifies the epileptogenic zone in the interictal electroencephalogram," *Epilepsia*, vol. 60, pp. 2404–2415, 12 2019.
- [30] J. Cimbalnik, M. Pail, P. Klimes, V. Travnicek, R. Roman, A. Vajcner, and M. Brazdil, "Cognitive processing impacts high frequency intracranial eeg activity of human hippocampus in patients with pharmacoresistant focal epilepsy," *Frontiers in Neurology*, vol. 11, 10 2020.
- [31] Y. Varatharajah, B. M. Berry, Z. T. Kalbarczyk, B. H. Brinkmann, G. A. Worrell, and R. K. Iyer, "Inter-ictal seizure onset zone localization using unsupervised clustering and bayesian filtering." *IEEE*, 5 2017, pp. 533–539.
- [32] K. V. Saboo, I. Balzekas, V. Kremen, Y. Varatharajah, M. Kucewicz, R. K. Iyer, and G. A. Worrell, "Leveraging electrophysiologic correlates of word encoding to map seizure onset zone in focal epilepsy: Task-dependent changes in epileptiform activity, spectral features, and functional connectivity," *Epilepsia*, vol. 62, pp. 2627–2639, 11 2021.
- [33] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, pp. 255–260, 7 2015.
- [34] C.-W. Tsai, C.-F. Lai, H.-C. Chao, and A. V. Vasilakos, "Big data analytics: a survey," *Journal of Big Data*, vol. 2, p. 21, 12 2015.
- [35] E. D. Nering, *Linear Algebra and Matrix Theory*, 2nd ed. Wiley, 1970.
- [36] I. T. Jolliffe and J. Cadima, "Principal component analysis: a review and recent developments," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, p. 20150202, 4 2016.
- [37] L. McInnes, J. Healy, and J. Melville, "Umap: Uniform manifold approximation and projection for dimension reduction," 2 2018.
- [38] I. Rish, "An empirical study of the naive bayes classifier," ser. JCAI 2001 workshop on empirical methods in artificial intelligence, vol. 3, 2001, pp. 41–46.
- [39] S. Węglarczyk, "Kernel density estimation and its application," *ITM Web of Conferences*, vol. 23, p. 00037, 11 2018.

- [40] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 5 2015.
- [41] J. Ahmad, H. Farman, and Z. Jan, "Deep learning methods and applications," pp. 31–42, 2019.
- [42] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," vol. 2018-January. Institute of Electrical and Electronics Engineers Inc., 3 2018, pp. 1–6.
- [43] A. Ajit, K. Acharya, and A. Samanta, "A review of convolutional neural networks." IEEE, 2 2020, pp. 1–5.
- [44] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," 6 2014. [Online]. Available: <http://arxiv.org/abs/1406.1078>
- [45] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, pp. 1735–1780, 11 1997.
- [46] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 6 2017. [Online]. Available: <http://arxiv.org/abs/1706.03762>
- [47] J. Nemcek, T. Vicar, and R. Jakubicek, "Weakly supervised deep learning-based intracranial hemorrhage localization," 5 2021. [Online]. Available: <http://arxiv.org/abs/2105.00781>
- [48] Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," *Neurocomputing*, vol. 452, pp. 48–62, 9 2021.
- [49] S. Gu, B. Kelly, and D. Xiu, "Autoencoder asset pricing models," *Journal of Econometrics*, vol. 222, pp. 429–450, 5 2021.
- [50] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 12 2014. [Online]. Available: <http://arxiv.org/abs/1412.6980>
- [51] A. Harati, S. Lopez, I. Obeid, J. Picone, M. P. Jacobson, and S. Tobochnik, "The tuh eeg corpus: A big data resource for automated eeg interpretation." IEEE, 12 2014, pp. 1–5.
- [52] T. Wen and Z. Zhang, "Deep convolution neural network and autoencoders-based unsupervised feature learning of eeg signals," *IEEE Access*, vol. 6, pp. 25 399–25 410, 2018.
- [53] C. Meisel and K. A. Bailey, "Identifying signal-dependent information about the preictal state: A comparison across ecog, eeg and ekg using deep learning," *EBioMedicine*, vol. 45, pp. 422–431, 7 2019.

- [54] D. Lu and J. Triesch, "Residual deep convolutional neural network for eeg signal classification in epilepsy," 3 2019. [Online]. Available: <http://arxiv.org/abs/1903.08100>
- [55] H. Khan, L. Marcuse, M. Fields, K. Swann, and B. Yener, "Focal onset seizure prediction using convolutional networks," *IEEE Transactions on Biomedical Engineering*, vol. 65, pp. 2109–2118, 9 2018.
- [56] ostas . Tsiouris, V. C. Pezoulas, M. Zervakis, S. Konitsiotis, D. D. Koutsouris, and D. I. Fotiadis, "A long short-term memory deep learning network for the prediction of epileptic seizures using eeg signals," *Computers in Biology and Medicine*, vol. 99, pp. 24–37, 8 2018.
- [57] "Multicenter intracranial eeg dataset for classification of graphoelements and artifactual signals," *Scientific Data*, vol. 7, p. 179, 12 2020.
- [58] S. Butterworth *et al.*, "On the theory of filter amplifiers," *Wireless Engineer*, vol. 7, no. 6, pp. 536–541, 1930.
- [59] J. Cimbálník, A. Hewitt, G. Worrell, and M. Stead, "The cs algorithm: A novel method for high frequency oscillation detection in eeg," *Journal of Neuroscience Methods*, vol. 293, pp. 6–16, 1 2018.
- [60] D. T. Barkmeier, A. K. Shah, D. Flanagan, M. D. Atkinson, R. Agarwal, D. R. Fuerst, K. Jafari-Khouzani, and J. A. Loeb, "High inter-reviewer variability of spike detection on intracranial eeg addressed by an automated multi-channel algorithm," *Clinical Neurophysiology*, vol. 123, pp. 1088–1095, 6 2012.
- [61] P. Řehulka, J. Cimbálník, M. Pail, J. Chrastina, M. Hermanová, and M. Brázdil, "Hippocampal high frequency oscillations in unilateral and bilateral mesial temporal lobe epilepsy," *Clinical Neurophysiology*, vol. 130, pp. 1151–1159, 7 2019.
- [62] Y. Sasaki and R. Fellow, "The truth of the f-measure," 2007.
- [63] W.-M. . D. Combinatorics and A. D. Bucchianico, "test," 1996.
- [64] T. Fawcett, "An introduction to roc analysis," *Pattern Recognition Letters*, vol. 27, pp. 861–874, 6 2006.
- [65] E. Zhang and Y. Zhang, "Average precision," pp. 192–193, 2009.
- [66] F. Wilcoxon, "Individual comparisons by ranking methods," pp. 80–83, 1945. [Online]. Available: <http://www.jstor.org/about/terms.html>.

# List of Figures

- Figure 1 Electrodes used for iEEG recording. . . . . 3
- Figure 2 A schematic example of 2D convolution. . . . . 7
- Figure 3 Structure of a hidden GRU unit. . . . . 7
- Figure 4 Architecture of Conv-GRU autoencoder network for unsupervised iEEG feature extraction. . . . . 8
- Figure 5 Unsupervised intracranial EEG (iEEG) clustering paradigm for aiding clinical staff in intracranial EEG assessment prior to epilepsy surgery. . . . . 10
- Figure 6 Illustrative examples of signals from the used datasets. . . . . 13
- Figure 7 Example of an iEEG signal segment and its corresponding spectrograms. . . . . 14
- Figure 8 A schematic example of an autoencoder. . . . . 15
- Figure 9 Data workflow during feature extraction. . . . . 16
- Figure 10 The general framework for determining the index of separability in Experiment 1 and Experiment 2. . . . . 18
- Figure 11 Graphical illustration of Experiment 1 and Experiment 2. . . . . 20
- Figure 12 Schematic illustration of two perspectives in Experiment 3. . . . . 21
- Figure 13 Framework of Experiment 4 for evaluation of performance of AE-trained and Biomarker-trained models. . . . . 23
- Figure 14 Visualisation of AE embedded features. . . . . 26
- Figure 15 Selection of PCA components. . . . . 27
- Figure 16 Visualisation of AE embedded components. . . . . 28
- Figure 17 HFO features distributions and statistical differences between SOZ and NSOZ channels. . . . . 30
- Figure 18 IS features distributions and statistical differences between SOZ and NSOZ channels. . . . . 31
- Figure 19 REN statistical distribution of NSOZ and SOZ channels. . . . . 32

# List of Tables

Table 1	Training parameters for Conv-GRU autoencoder training and optimization. . . . .	15
Table 2	Total number of iEEG segments in each category in FNUSA and Mayo datasets.	19
Table 3	Number of SOZ and NSOZ samples in FNUSA and Mayo datasets. . . . .	19
Table 4	Index of separability among different iEEG signal categories - PCA . . . . .	24
Table 5	Index of separability among different iEEG signal categories - UMAP . . . . .	25
Table 6	Index of separability between SOZ and NSOZ segments . . . . .	27
Table 7	The AUROC values and p-values (asterisk representation) of HFO features detected in ripple band for differentiation between SOZ and NSOZ channels. . . . .	29
Table 8	The AUROC values and p-values (asterisk representation) of HFO features detected in fast ripple band for differentiation between SOZ and NSOZ channels. . . . .	30
Table 9	The AUROC values and p-values (asterisk representation) of IS features for differentiation between SOZ and NSOZ channels. . . . .	31
Table 10	The AUROC values and statistical significance of REN for discrimination between SOZ and NSOZ channels. . . . .	32
Table 11	Evaluation of the HFO features in the ripple band for SOZ-NSOZ discrimination task for each patient individually. . . . .	33
Table 12	Evaluation of the HFO features in the fast ripple band for SOZ-NSOZ discrimination task for each patient individually. . . . .	33
Table 13	Evaluation of the IS features for SOZ-NSOZ differentiation for each patient separately. . . . .	34
Table 14	Summary of REN evaluation for SOZ-NSOZ discrimination for each patient separately. . . . .	35
Table 15	Results of leave-one-patient-out cross-validation within Experiment 4. . . . .	36
Table 16	Overview of recorded iEEG channels collected from St Anne’s University Hospital.	53
Table 17	Overview of recorded iEEG channels collected from Mayo Clinic . . . . .	54
Table 18	Results of SOZ-NSOZ channel differentiation of HFO features detected in ripple band (80-200 Hz) for each patient separately. . . . .	55
Table 19	Results of SOZ-NSOZ channel differentiation of HFO features detected in fast ripple band (200-500 Hz) for each patient separately. . . . .	56
Table 20	Results of SOZ-NSOZ channel differentiation of IS features for each patient separately. . . . .	57
Table 21	Results of SOZ-NSOZ channel differentiation of relative entropy for each patient separately. . . . .	58
Table 22	Results of leave-one-patient-out cross-validation of the AE-trained classifier. . . . .	59

Table 23	Results of leave-one-patient-out cross-validation of the Biomarker-trained classifier. . . . .	60
Table 24	Results of leave-one-patient-out cross-validation of the AE+Biomarker-trained classifier. . . . .	61
Table 25	Maximal performances from the AE-trained, Biomarkers-trained, AE+Biomarker-trained classifiers. . . . .	62

# List of Abbreviations

<b>AE</b>	Autoencoder
<b>AUROC</b>	Area Under Receiver Operating Characteristic Curve
<b>AP</b>	Average Precision
<b>CNN</b>	Convolutional Neural Networks
<b>DL</b>	Deep Learning
<b>DRE</b>	Drug-Resistant Epilepsy
<b>ECoG</b>	Electrocorticography
<b>EEG</b>	Electroencephalography
<b>EFR</b>	Embedded Feature Representation
<b>FNUSA</b>	St. Anne 's University Hospital Brno
<b>GRU</b>	Gated Recurrent Units
<b>HFO</b>	High Frequency Oscillations
<b>IS</b>	Interictal Spikes
<b>IZ</b>	Irritative Zone
<b>iEEG</b>	Intracranial Encephalography
<b>KDE</b>	Kernel density Estimation
<b>MAE</b>	Mean Absolute Error
<b>MRI</b>	Magnetic Resonance Imaging
<b>NSOZ</b>	Non-Seizure-Onset Zone
<b>PCA</b>	Principal Component Analysis
<b>REN</b>	Relative Entropy
<b>RNN</b>	Recurrent Neural Networks

**SEEG** Stereoelectroencephalography

**SOZ** Seizure Onset Zone

**UMAP** Uniform Manifold Approximation and Projection

## A Appendix

Detailed information about used datasets and results of individual patients' analysis from Experiment 3 and Experiment 4 are shown in this section.

**Dataset FNUSA**

Patient ID	Institution	Number of recorded iEEG channels	Number of SOZ channels	Number of NSOZ channels	Number of IZ channels	Sampling frequency	Signal duration [h]
3	FNUSA	74	4	66	4	5000	0,5
4	FNUSA	67	4	55	8	5000	0,5
6	FNUSA	111	37	40	34	5000	0,32
13	FNUSA	105	6	87	12	5000	0,64
15	FNUSA	110	8	102	0	5000	0,59
21	FNUSA	100	10	76	14	5000	0,52
31	FNUSA	130	14	56	60	5000	0,58
33	FNUSA	134	8	111	15	5000	0,53
40	FNUSA	95	12	83	0	5000	0,54
42	FNUSA	106	13	93	0	5000	0,52
43	FNUSA	100	7	94	0	5000	0,54
45	FNUSA	175	13	163	0	5000	0,67
47	FNUSA	156	6	150	0	5000	0,52
50	FNUSA	173	2	171	0	5000	0,56
60	FNUSA	153	2	151	0	5000	0,51
63	FNUSA	120	3	117	0	5000	0,67
71	FNUSA	169	9	160	0	5000	0,53
72	FNUSA	161	4	157	0	5000	0,52

Table 16: Overview of recorded iEEG channels collected from St Anne's University Hospital (Brno, Czech Republic) in patients diagnosed with drug-resistant epilepsy (DRE). All channels were manually reviewed and assessed as SOZ = seizure onset zone, NSOZ = non-seizure onset zone, IZ = irritative zone.

**Dataset Mayo**

Patient ID	Institution	Number of recorded iEEG channels	Number of SOZ channels	Number of NSOZ channels	Number of IZ channels	Sampling frequency	Signal duration [h]
36	Mayo	8	4	4	0	5000	2
69	Mayo	8	6	2	0	5000	2
201	Mayo	6	2	4	0	5000	2
344	Mayo	8	2	6	0	5000	1,25
387	Mayo	12	4	8	0	5000	2
406	Mayo	16	3	13	0	5000	2
419	Mayo	8	4	4	0	5000	2
421	Mayo	12	2	10	0	5000	2
448	Mayo	8	6	2	0	5000	2
449	Mayo	24	2	22	0	5000	2
451	Mayo	16	4	12	0	5000	2
453	Mayo	16	2	14	0	5000	2
552	Mayo	12	8	4	0	5000	2
569	Mayo	8	3	5	0	5000	2
573	Mayo	4	3	1	0	5000	2

Table 17: Overview of recorded iEEG channels collected from Mayo Clinic (Rochester, USA) in patients diagnosed with drug-resistant epilepsy (DRE). All channels were manually reviewed and assessed as SOZ = seizure onset zone, NSOZ = non-seizure onset zone, IZ = irritative zone.

HFO – Ripples (80-200 Hz)					
ID	Institution	Ripples count/10min	Ripples mean amplitude /10 min	Ripples amplitude	Ripples duration
3	fnusa	0.81*	0.92**	0.96**	0.7
4	fnusa	0.81*	0.73	0.7	0.77
6	fnusa	0.57	0.7**	0.68**	0.63
13	fnusa	0.73	0.77*	0.74*	0.69
15	fnusa	0.61	0.72*	0.73*	0.83**
21	fnusa	0.58	0.63	0.63	0.94****
31	fnusa	0.88****	0.54	0.57	0.51
33	fnusa	0.8**	0.74*	0.69	0.9***
40	fnusa	0.57	0.61	0.63	0.6
42	fnusa	0.95****	0.95****	0.95****	0.84****
43	fnusa	0.97****	0.85**	0.84**	0.58
45	fnusa	0.67*	0.79***	0.77**	0.5
47	fnusa	0.87**	0.74*	0.73	0.5
50	fnusa	0.75	0.91*	0.9	0.5
60	fnusa	1*	0.86	0.74	0.76
63	fnusa	0.66	0.91*	0.87*	0.59
71	fnusa	0.84***	0.87**	0.87***	0.92****
72	fnusa	0.82*	0.94**	0.91**	0.59
36	mayo	1*	0.94	0.94	0.75
69	mayo	0.6	0.8	0.8	0.8
201	mayo	0.75	0.88	1	0.62
344	mayo	0.5	0.5	0.5	0.6
387	mayo	0.93	0.93	0.79	1
406	mayo	0.69	0.61	0.58	0.47
419	mayo	0.88	0.62	0.75	0.69
421	mayo	0.83	1	1	0.58
448	mayo	0.75	0.75	0.75	0.83
449	mayo	0.78	0.97*	1*	0.62
451	mayo	0.64	0.48	0.73	1*
453	mayo	1*	1*	0.96	1*
552	mayo	1**	0.88	0.78	0.66
569	mayo	1*	0.6	0.53	0.53
573	mayo	1	1	0.67	1

Table 18: The AUROC and p-values (asterisk representation) for SOZ-NSOZ channel differentiation of HFO features detected in ripple band (80-200 Hz), evaluated within Experiment 3 for each patient separately.

HFO – Fast ripples (200-500 Hz)					
ID	Institution	Fast ripples count/10min	Fast ripples mean amplitude /10 min	Fast ripples amplitude	Fast ripples duration
3	fnusa	0.7	0.61	0.52	0.76
4	fnusa	1***	0.76	0.91**	0.74
6	fnusa	0.58	0.64*	0.71**	0.71**
13	fnusa	0.93***	0.87**	0.8*	0.77*
15	fnusa	0.74*	0.64	0.65	0.81**
21	fnusa	0.71*	0.79**	0.73*	0.79**
31	fnusa	0.33*	0.67*	0.69*	0.79***
33	fnusa	0.71	0.75*	0.68	0.68
40	fnusa	0.67	0.75**	0.74**	0.56
42	fnusa	0.55	0.93****	0.92****	0.84****
43	fnusa	0.72	0.57	0.6	0.64
45	fnusa	0.5	0.44	0.56	0.55
47	fnusa	0.71	0.74*	0.74*	0.83**
50	fnusa	0.63	0.72	0.75	0.71
60	fnusa	0.95*	0.87	0.65	0.75
63	fnusa	0.73	0.79	0.77	0.69
71	fnusa	0.99****	0.95****	0.95****	0.75*
72	fnusa	0.78	0.72	0.74	0.63
36	mayo	1*	0.81	0.75	1*
69	mayo	0.9	1	1	0.6
201	mayo	0.62	1	1	0.88
344	mayo	0.6	0.6	0.6	0.8
387	mayo	0.79	1	0.36	0.79
406	mayo	0.61	0.78	0.72	0.5
419	mayo	0.56	0.56	0.62	0.75
421	mayo	0.92	0.83	0.83	1
448	mayo	0.83	0.67	0.5	1
449	mayo	0.94	0.75	0.84	0.94
451	mayo	0.85	0.61	0.76	0.91*
453	mayo	0.85	1*	0.96	0.88
552	mayo	0.69	0.91*	0.53	0.72
569	mayo	0.6	0.53	0.6	0.73
573	mayo	1	1	1	0.67

Table 19: The AUROC and p-values (asterisk representation) for SOZ-NSOZ channel differentiation of HFO features detected in fast ripple band (200-500 Hz), evaluated within Experiment 3 for each patient separately.

Interictal spikes								
ID	Institution	Spikes count / 10 min	Spike mean amplitude / 10 min	Event amplitude	Left amplitude	Right amplitude	Left duration	Right duration
3	fnusa	0.73	0.54	0.54	0.74	0.55	0.63	0.71
4	fnusa	0.78	0.84*	0.84*	0.67	0.9**	0.52	0.5
6	fnusa	0.48	0.66**	0.62*	0.66**	0.53	0.65*	0.51
13	fnusa	0.9**	0.56	0.67	0.86**	0.64	0.88**	0.92***
15	fnusa	0.97****	0.74*	0.72*	0.88***	0.96****	0.95****	0.85***
21	fnusa	0.84***	0.92****	0.91****	0.86***	0.85***	0.65	0.54
31	fnusa	0.94****	0.59	0.49	0.63	0.88****	0.67*	0.63
33	fnusa	0.57	0.64	0.58	0.52	0.5	0.57	0.55
40	fnusa	0.59	0.6	0.6	0.7*	0.54	0.48	0.58
42	fnusa	0.81***	0.83***	0.83***	0.99****	0.72**	0.73**	0.99****
43	fnusa	0.9***	0.8**	0.76*	0.83**	0.86**	0.43	0.79*
45	fnusa	0.6	0.65	0.66	0.6	0.65	0.56	0.47
47	fnusa	0.97****	0.66	0.63	0.66	0.84**	0.87**	0.73
50	fnusa	0.78	0.73	0.75	0.65	0.58	0.67	0.47
60	fnusa	0.59	0.5	0.51	0.94*	0.51	0.89	0.82
63	fnusa	0.71	0.58	0.58	0.4	0.62	0.7	0.66
71	fnusa	0.96****	0.63	0.67	0.59	0.81**	0.64	0.73*
72	fnusa	0.93**	0.66	0.63	0.66	0.74	0.91**	0.93**
36	mayo	0.62	0.56	0.56	0.62	1*	1*	0.81
69	mayo	1	0.83	0.83	0.5	0.92	0.58	0.58
201	mayo	0.62	1	1	0.75	0.88	0.75	0.75
344	mayo	0.8	0.9	0.7	1	1	1	1
387	mayo	0.59	0.59	0.66	0.59	0.53	0.5	0.62
406	mayo	0.97*	1*	1*	1*	0.94*	0.86	0.86
419	mayo	0.69	0.62	0.69	0.69	0.81	0.88	0.69
421	mayo	1	0.92	1	1	0.67	1	1
448	mayo	0.96	1	1	0.5	1	0.58	0.5
449	mayo	0.57	0.91	0.89	0.52	0.45	0.64	0.84
451	mayo	0.58	0.52	0.48	0.7	0.7	0.88	0.55
453	mayo	0.96	1*	1*	1*	1*	0.96	1*
552	mayo	0.97*	0.97*	0.84	1**	0.91*	0.59	0.72
569	mayo	0.6	0.47	0.47	0.67	0.67	0.6	0.87
573	mayo	0.67	1	1	1	1	0.67	1

Table 20: The AUROC and p-values (asterisk representation) for SOZ-NSOZ channel differentiation of IS features evaluated within Experiment 3 for each patient separately.

Relative Entropy		
ID	Institution	REN
3	fnusa	0.88*
4	fnusa	0.93**
6	fnusa	0.47
13	fnusa	0.87**
15	fnusa	0.99****
21	fnusa	0.89****
31	fnusa	0.86****
33	fnusa	0.94****
40	fnusa	0.9****
42	fnusa	0.56
43	fnusa	0.76*
45	fnusa	0.79***
47	fnusa	0.99****
50	fnusa	0.84
60	fnusa	1*
63	fnusa	0.99**
71	fnusa	0.95****
72	fnusa	0.69
36	mayo	1*
69	mayo	0.67
201	mayo	0.75
344	mayo	0.6
387	mayo	0.75
406	mayo	0.82
419	mayo	0.81
421	mayo	0.67
448	mayo	0.33
449	mayo	0.82
451	mayo	0.52
453	mayo	1*
552	mayo	0.5
569	mayo	0.87
573	mayo	1

Table 21: The AUROC and p-values (asterisk representation) for SOZ-NSOZ channel differentiation of relative entropy evaluated within Experiment 3 for each patient separately.

<b>AE scores - FNUSA</b>			
Left-out patient ID	F1	AUROC	AP
3	0.63	0.61	0.56
4	0.72	0.62	0.57
6	0.58	0.52	0.51
13	0.47	0.57	0.54
15	0.68	0.7	0.65
21	0.83	0.8	0.72
31	0.37	0.6	0.58
33	0.39	0.55	0.53
40	0.54	0.64	0.6
42	0.75	0.73	0.66
43	0.75	0.79	0.78
45	0.68	0.66	0.6
47	0.78	0.71	0.63
50	0.65	0.73	0.71
60	0.82	0.82	0.75
63	0.66	0.66	0.61
71	0.71	0.69	0.62
72	0.82	0.78	0.69
<b>Mean</b>	<b>0.66</b>	<b>0.68</b>	<b>0.63</b>

<b>AE scores - Mayo</b>			
Left-out patient ID	F1	AUROC	AP
36	0.76	0.78	0.74
69	0.7	0.72	0.66
201	0.7	0.65	0.59
344	0	0.47	0.5
387	0.24	0.57	0.57
406	0.33	0.33	0.44
419	0.6	0.48	0.49
421	0.86	0.83	0.75
448	0	0.5	0.5
449	0.51	0.44	0.47
451	0	0.5	0.5
453	0.72	0.64	0.58
552	0.67	0.51	0.5
569	0	0.49	0.5
573	0.52	0.44	0.47
<b>Mean</b>	<b>0.44</b>	<b>0.56</b>	<b>0.55</b>

Table 22: The F1, AUROC, and AP performances from leave-one-patient-out cross-validation performed within Experiment 4 on the FNUSA dataset (left table) and on the Mayo dataset (right table) of the AE-trained classifier.

<b>Biomarkers scores - FNUSA</b>			
Left-out patient ID	F1	AUROC	AP
3	0.69	0.73	0.69
4	0.79	0.8	0.75
6	0.24	0.54	0.52
13	0.11	0.51	0.51
15	0.55	0.63	0.59
21	0.82	0.84	0.8
31	0.75	0.79	0.76
33	0.73	0.78	0.77
40	0.69	0.71	0.66
42	0.44	0.61	0.59
43	0.66	0.71	0.67
45	0.49	0.62	0.59
47	0.84	0.85	0.82
50	0.47	0.62	0.59
60	0.27	0.54	0.53
63	0.79	0.8	0.75
71	0.84	0.85	0.81
72	0.61	0.68	0.64
<b>Mean</b>	<b>0.60</b>	<b>0.70</b>	<b>0.67</b>

<b>Biomarkers scores - Mayo</b>			
Left-out patient ID	F1	AUROC	AP
36	0.63	0.58	0.54
69	0.63	0.51	0.5
201	0.43	0.39	0.45
344	0.8	0.74	0.66
387	0.33	0.53	0.52
406	0.69	0.56	0.53
419	0.64	0.64	0.59
421	0.39	0.35	0.44
448	0.58	0.55	0.53
449	0.07	0.43	0.49
451	0.62	0.46	0.48
453	0.64	0.62	0.57
552	0.74	0.75	0.69
569	0.67	0.5	0.5
573	0.23	0.48	0.49
<b>Mean</b>	<b>0.54</b>	<b>0.54</b>	<b>0.53</b>

Table 23: The F1, AUROC, and AP performances from leave-one-patient-out cross-validation performed within Experiment 4 on the FNUSA dataset (left table) and on the Mayo dataset (right table) of the Biomarker-trained classifier.

<b>AE + Biomarkers scores - FNUSA</b>			
Left-out patient ID	F1	AUROC	AP
3	0.72	0.74	0.69
4	0.78	0.76	0.69
6	0.2	0.49	0.49
13	0.1	0.49	0.5
15	0.39	0.56	0.53
21	0.85	0.86	0.83
31	0.78	0.79	0.74
33	0.53	0.66	0.64
40	0.67	0.72	0.68
42	0.18	0.44	0.48
43	0.62	0.69	0.65
45	0.63	0.68	0.64
47	0.71	0.75	0.71
50	0	0.49	0.5
60	0.28	0.56	0.55
63	0.73	0.76	0.71
71	0.73	0.77	0.74
72	0.74	0.77	0.73
<b>Mean</b>	<b>0.54</b>	<b>0.67</b>	<b>0.64</b>

<b>AE + Biomarkers scores - Mayo</b>			
Left-out patient ID	F1	AUROC	AP
36	0.6	0.58	0.55
69	0.64	0.55	0.53
201	0.54	0.41	0.46
344	0.79	0.79	0.72
387	0.38	0.6	0.58
406	0.65	0.63	0.58
419	0.63	0.62	0.58
421	0.27	0.5	0.5
448	0.33	0.42	0.47
449	0.26	0.49	0.5
451	0.58	0.43	0.47
453	0.63	0.62	0.57
552	0.77	0.76	0.69
569	0.65	0.49	0.5
573	0.47	0.59	0.56
<b>Mean</b>	<b>0.55</b>	<b>0.57</b>	<b>0.55</b>

Table 24: The F1, AUROC, and AP performances from leave-one-patient-out cross-validation performed within Experiment 4 on the FNUSA dataset (left table) and on the Mayo dataset (right table) of the AE+Biomarker-trained classifier.

<b>The maximal performance per patient FNUSA</b>			
Left-out patient ID	F1	AUROC	AP
3	0.72	0.74	0.69
4	0.79	0.8	0.75
6	0.58	0.54	0.52
13	0.47	0.57	0.54
15	0.68	0.7	0.65
21	0.85	0.86	0.83
31	0.78	0.79	0.76
33	0.73	0.78	0.77
40	0.69	0.72	0.68
42	0.75	0.73	0.66
43	0.75	0.79	0.78
45	0.68	0.68	0.64
47	0.84	0.85	0.82
50	0.65	0.73	0.71
60	0.82	0.82	0.75
63	0.79	0.8	0.75
71	0.84	0.85	0.81
72	0.82	0.78	0.73
<b>Mean</b>	<b>0.74</b>	<b>0.75</b>	<b>0.71</b>

<b>The maximal performance per patient Mayo</b>			
Left-out patient ID	F1	AUROC	AP
36	0.76	0.78	0.74
69	0.70	0.72	0.66
201	0.70	0.65	0.59
344	0.80	0.79	0.72
387	0.38	0.60	0.58
406	0.69	0.63	0.58
419	0.64	0.64	0.59
421	0.86	0.83	0.75
448	0.58	0.55	0.53
449	0.51	0.49	0.50
451	0.62	0.50	0.50
453	0.72	0.64	0.58
552	0.77	0.76	0.69
569	0.67	0.50	0.50
573	0.52	0.59	0.56
<b>Mean</b>	<b>0.66</b>	<b>0.64</b>	<b>0.60</b>

Table 25: The maximal F1, AUROC, and AP performances from the three classifiers (AE-trained, Biomarkers-trained, AE+Biomarker-trained) evaluated within Experiment 4 for each patient from the FNUSA dataset (left table) and from the Mayo dataset (right table).