Detecting pathological scalp recordings using features of the EEG background pattern

Rob van der Nagel, October 30, 2022

[1] Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, Enschede, The Netherlands
[2] Department of Clinical Neurophysiology, Medisch Spectrum Twente, Institute of Technical Medicine, Enschede, The Netherlands

Abstract: The purpose of this study was to improve feature-based pathology decoding models by utilizing features that describe properties related to the EEG background pattern. An extensive study was conducted to highlight and quantify the impact of key components, such as the differences between cropped-wise and image-based decoding strategies, single and multichannel inputs, computational and spatial features and binary classifiers. The experimental results were compared to a well-established baseline model adopted from Shirrmeister et al. (2017) that reported state-of-the-art performance while validated on previous datasets. Although not all models showed increased performance, the strategy to train a VGG using spectograms derived from raw single channel data located in the occipital region yielded the best decoding results, with specificities and sensitivities ranging between 82.5-85.0% and 73.75-76.25%, respectively. The cropped-wise models did not show significant improvements, although they performed in a narrow range close to the baseline performance. From the proposed background features, information that captured properties of the posterior dominant rhythm (PDR) was clearly one of the most discriminative features while considering different decoding strategies, inputs and classifiers. Individual experiments showed that including background features to a baseline set of computational features changed the decoding performance significantly considering different classifier-designs. Further, it was found that features computed from channels in the left temporal/frontal lobes, as well as those in the occipital region, were important. To ensure reproducability all the features and classifier-designs were uploaded to: https://github.com/RobvanderNagel94/Pathology-Decoding.

1 Introduction

Electroencephalography (EEG) is a non-invasive technique used for monitoring and recording brain electrical activity. EEGs provide essential information about brain functioning and assist clinicians in studying the effects of neurological conditions such as epilepsy, dementia, schizophrenia and depression (Horváth et al., 2016; Morita et al., 2011; Boutros et al., 2008; Thibodeau et al., 2006). Routine monitoring enables clinicians to study patients under varying conditions (e.g., exposure to certain stimuli, during task performance or deep sleep) to better understand the patients' brain activity related to neurological conditions. The relative low cost of performing EEGs for recording brain activity has made it a popular tool for routine care, making routine recordings a valuable source to detect neurological diseases.

EEG analysis in clinical neurology generally involves detecting transients and analyzing the background pattern (Van Putten, 2009; Schomer et al., 2012). Transients refer to relatively rare events, including physiological and pathological waveforms, while the background pattern refers to the mean statistical characteristics of the EEG. The decision to classify a recording as normal or pathological is made based on a range of specific descriptions of the background pattern and the presence or absence of transients given the patient's state of consciousness (awake, asleep, drowsy, comatose) and state (eyes open, eyes closed, hyperventilation, photic stimulation). It is important to note that recognising pathological waveforms requires years of training and diagnostic reliability is subjected to several limitations (i.e., individual training and experience, consistency of rating over time and subjective criteria for defining specific thresholds to describe the severity of an abnormality).

Visual EEG analysis remains the gold standard for interpretation. However, computerized classification algorithms have been proposed that can identify recordings as pathological or non-pathological based on labeled examples (Shirrmeister et al., 2017; Van Leeuwen et al., (2019); Gemein et al., 2020; Singh et al., 2021). Often referred to as pathology decoding, such computerized methods are promising and offer clinicians advantages over visual analysis. For example, an automated suggestion could significantly reduce workload, allowing the clinician to focus on pathological suggested recordings for further investigation. Furthermore, they allow clinicians to monitor a large pool of patients, which may lead to better patient care. It is important to understand that most recording sessions are routine checks intended for patients that already have been diagnosed, but even healthy individuals may experience EEG abnormalities. Another benefit is that computerized methods can adopt pattern-recognition technologies to detect patterns that humans cannot.

Although current methods perform satisfactory, pathology decoding research is still in its infancy and existing methods are far from replacing visual analysis. Additional research is needed aiming to increase decoding performance, as well as validating methods on other datasets. Prior research showed that the majority of work focused on validating "feature-based" and "end-to-end" models trained on publicly available scalp recordings (Obeid and Picone, 2016). Most of the papers proposed designs based on Convolutional Neural Networks, or ConvNets for short, that can be used for general classification tasks. The works of Shirrmeister et al. (2017) and van Leeuwen et al. (2019) have set strong baselines to decode pathology using Deep ConvNets in an end-to-end way. To the best of our knowledge, this specific ConvNet-design has been validated on separate datasets and produced superior decoding performance. On the other hand, Gemein et al. (2020) and Singh et al. (2021) showed comparable results while adopting different feature-based methods.

Even though a plethora of ConvNets have been proposed to decode pathology, little attention was given to the development of alternative approaches using features that capture specific properties of the EEG background activity, while changes in the background pattern have shown to be excellent indicators of general brain dysfunction (Jordan, 2004; Jin et al., 2006; van Putten, 2007; Knyazeva et al., 2008; Spronk et al., 2011). Lodder and van Putten (2012) proposed to quantify properties of the background pattern using five key features. In addition, van Putten (2007) earlier introduced to visualise the time-localised frequency information for a triplet of features to quantify the spatial distribution of EEG signals and their coherence, represented as three time-frequency images (Colorful Brain). The fact that labels are assigned to recordings based on, among other factors, specific changes in the background pattern, it might be possible to improve the decoding performance with features that quantify specific characteristics of the background pattern.

In this paper, an in-depth study was conducted on how the features proposed by Lodder and van Putten (2012) and van Putten (2007) can be applied within existing feature-based models to test the hypothesis that decoding performance can be improved by using features of the EEG background pattern. Our experimental design included experiments with different feature-based training strategies, inputs, features and classifiers to assess each component's contribution to the decoding performance. To validate the experimental results we used the well-established Deep ConvNet adapted from Shirrmeister et al. (2017) as a baseline comparison.

The rest of this paper is structured as follows. Subsection 1.1 contains the description of related work. Section 2 provides details on the data preparation and preprocessing (subsection 2.1), feature design (subsections 2.2-9) and experimental setup (subsection 2.3). The results are presented in section 3, including the evaluation of decoding models (subsection 3.1) and features (subsection 3.2). Section 4 discusses on the interpretations of the results. The conclusion, as well as suggestions for future work, is presented in section 5.

1.1 Related work

As mentioned, the studies that proposed computerized solutions for pathology decoding can broadly be divided into feature-based and end-to-end methods. Both methods are build from two parts: feature extraction and classification. End-to-end methods use raw or minimally preprocessed signal data as input, while feature-based methods use features to represent the raw signal data. It is important to note that different EEG representations have been used as input. Generic ConvNets require two-dimensional inputs, so studies proposed raw and feature inputs from *multi*-channel data, and raw and feature inputs from *single*-channel data (i.e., spectogram image taken from a specific channel, or any feature taken from a single channel that outputs a two-dimensional shape). For clarity, the following work briefly summarizes prior contributions with the specific decoding models depicted in Table 1.

Lopez et al. (2017) validated numerous models based on multichannel and single channel input data. The EEG signals were retrieved using a TCP montage from which they generated features based on mel-frequency cepstral coefficients (MFCCs). Several binary classifiers were used to decode the features, including k-nearest neighbors (KNN), random forest (RF) and a deep classifier (ConvNet-MLP). They used the first few minutes of each recording and explored with generating features from different segments of signal data (commonly referred to as crops or epochs). The dimensionality of the cropped-features was reduced using principle component analysis (PCA). A hidden Markov model with Gaussian mixture emissions (GMM-HMM) was proposed to model the cropped-features in a probabilistic framework.

Shirrmeister et al. (2017) proposed Shallow and Deep ConvNets. The general idea is that ConvNets apply learned filter operations to two-dimensional input data to extract high-level feature maps. By design, ConvNets allow for a joint optimization of the feature extraction and classification in one architecture. The Softmax function was used as activation function in the output layer of the network to classify a binary outcome. Reduced versions of the signals were used to train the ConvNets, using only the first 1, 2, 4, 8 or 16 minutes following the beginning of a recording. They proposed spectogram features taken from multichannel inputs within five frequency bands: delta (0–4 Hz), theta (4–8 Hz), alpha (8–14 Hz), low beta (14–20 Hz), high beta (20–30 Hz) and low gamma (30–50 Hz).

Table 1 Proposed work for pathology decoding. The papers proposed computerized solutions for a binary classification problem. Spec, Sens and Acc are in %, *= Softmax classifier, **= validated on TUH Abnormal Corpus (Obeid and Picone, 2016).

Paper	Method	Model	Spec	Sens	Acc
Van Leeuwen et al. (2019)	end-to-end *(multi)*	ConvNet-*	90	74.8	-
		ConvNet+Age-*	90	74.3	-
		ConvNet+Sleep-*	90	76.3	-
		ConvNet+LSTM-*	90	71.1	-
Amin et al.** (2019)	end-to-end *(multi)*	AlexNet-SVM	94.7	78.6	-
		VGG16-SVM	94	77.8	-
Yıldırım et al.** (2018)	end-to-end *(single)*	1D-ConvNet-*	-	-	79.3
Roy et al.** (2018)	end-to-end *(single)*	1D-ConvNet-*	-	-	76.9
		1D-ConvNet-RNN-*	-	-	82.2
		MLP	-	-	54.1
		LR	-	-	49.1
Bajpai et al.** (2021)	feature-based *(single)*	SeizureNet-SVM	100	90.5	96.7
		InceptionNet-SVM	100	74.6	88.4
		DenseNet-SVM	100	87.3	94.2
Singh et al.** (2021)	feature-based *(single)*	VGG19-RF	86.7	79.5	82.2
		VGG19-SVM	82.9	77.2	79.3
		VGG19-LR	84.4	78.4	80.8
Roy et al.** (2018)	feature-based *(single)*	ConvNet-*	-	-	70.4
		TCNN-RNN-*	-	-	71.5
Lopez et al.** (2017)	feature-based *(single)*	ConvNet-MLP-*	-	-	58
Gemein et al.** (2020)	feature-based *(multi)*	Deep ConvNet-*	91.9	75.9	84.6
		Shallow ConvNet-*	87.9	79.7	84.1
		TCN-*	91.6	79.7	86.2
		EEGNet-*	92.9	72.1	83.4
		RF	88.3	79	84.1
		SVM	92.7	66.7	80.8
		RG	92.7	77.8	85.9
		ASC	88.1	80.6	84.7
Shirrmeister et al.** (2017)	feature-based *(multi)*	Shallow ConvNet-*	81.9	75.4	78.8
		Deep ConvNet-*	94.1	75.1	85.4
Lopez et al.** (2017)	feature-based *(multi)*	RF	-	-	83
		KNN	-	-	68.8
		GMM-HMM	-	-	74.4
		PCA-HMM	-	-	78.8

Roy et al. (2018) and Yildirim et al. (2018) both validated one-dimensional ConvNets in an end-to-end way. By design, this architecture requires one-dimensional inputs. Roy et al. (2018) experimented with raw single channel data from re-referenced channel pairs and reported the most optimal combination. Yildirim et al. (2018) specifically used the T5–O1 channel pair from a re-referenced TCP montage. In addition, Roy et al. (2018) opted for a two-dimensional ConvNet trained on images produced by Gramian Angular Fields (GAF). The computation outputs a temporal correlation computed using polar coordinates, with each element a trigonometric sum between different time intervals.

Van Leeuwen et al. (2019) applied a modified ConvNet adapted from the work of Shirrmeister et al. (2017) used in an end-to-end way. The models are essentially similar but differ in minor degree.

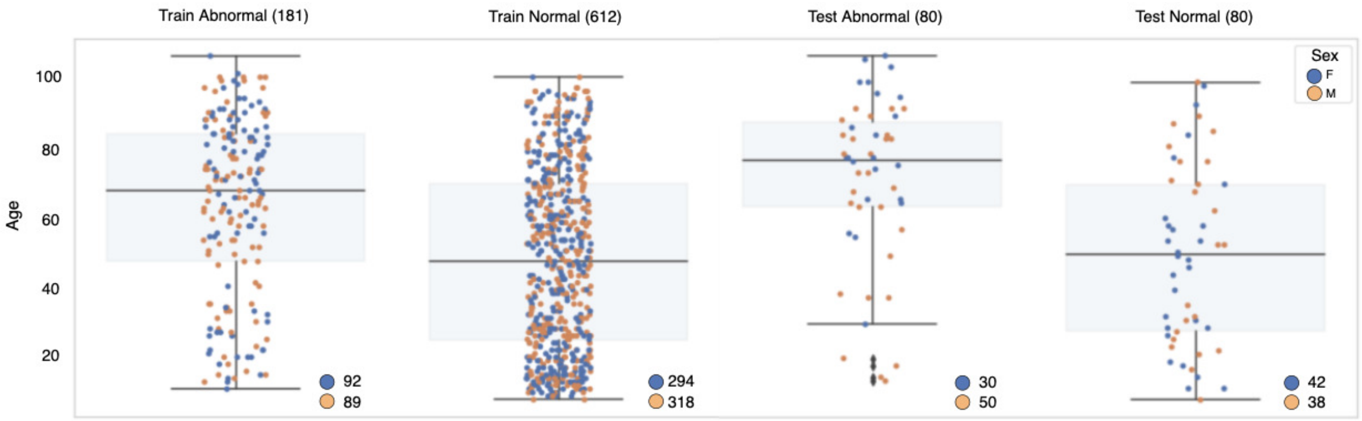

Fig. 1: Dataset of 953 labeled routine scalp recordings. The four boxplots show the patients' age distribution among the recordings. The test set was randomly chosen under the condition that an equal amount of normal and pathological recordings were selected (approx. 85% train and 15% test).

They were able to assign sleep stages to consecutive epochs of EEG and explored with including sleep stage probabilities and contextual information such as age and gender. Prior to the last classification layer they added an extra average pooling layer to take an average along the time point axis which enabled them to add new features to the network. Moreover, they used a technique to flip the channels of the left and right hemispheres with probability 0.5 to avoid overfitting to one hemisphere.

Gemein et al. (2020) introduced a variety of features categorized in time, frequency and synchrony features. They adopted the strategy to compute features from different crops within specific frequency bands, as earlier proposed by Shirrmeister et al. (2017). They experimented with binary classifiers, including support vector machine (SVM) and several automated Scikit-learn classifiers (ASC) to decode the cropped-features. They also introduced two deep classifiers earlier proposed to classify EEG-signals in motor-imagery for brain-computer interfaces. In addition, they proposed to compute a covariance matrix for each crop and opted to decode pathology using Riemannian Geometry (RG).

Amin et al. (2019) followed an end-to-end approach and proposed pre-trained ConvNets (VGG-16 and AlexNet) fine-tuned on raw multichannel data. As opposed to generic ConvNets, these ConvNets have fixed weights specifically designed to extract high-level feature maps from general images. A binary output was created by rescaling the output dimensions of the ConvNets, usually adding one or more dense layers. Instead of the Softmax function, they proposed a SVM-classifier at the end of the network.

Singh et al. (2021) proposed to use spectogram images derived from single channel data computed using Fourier (STFT), chromograms and MFCCs. They used a VGG-19 architecture to extract high-level feature maps from the spectograms and proposed to decode them using three classifier-designs, including Linear Regression (LR), RF and SVM. The F7+STFT combination yielded the best performance among all individual single channel tests.

Bajpai et al. (2021) computed spectogram images using STFT but focused on one specific channel. They used the first minute from each recording of the T5-O1 channel pair derived from a re-referenced TCP montage. Previously developed ConvNets for specific and nonspecific EEG-classification tasks were proposed, including DenseNet, InceptionNet and Seizurenet.

2 Materials and Methods

2.1 Data preparation and preprocessing

The study was performed with labeled routine scalp recordings collected by the department of Clinical Neurophysiology at Medisch Spectrum Twente. The EEGs were performed on patients diagnosed with epilepsy. Semi-random patient ID's were assigned to individual recordings to anonymize the data. Standard EEG caps were used with 19 Ag–AgCl electrodes placed according to the international 10–20 electrode system (see Fig. 2). Electrode impedances were kept below 5 kΩ to reduce polarization effects. For most recordings, a standard 20-minute protocol was followed. A common reference montage was used to analyse the EEG signals measured at a sampling rate of 250 Hz. Recordings were annotated for eyes open, eyes closed, hyperventilation and photic stimulation.

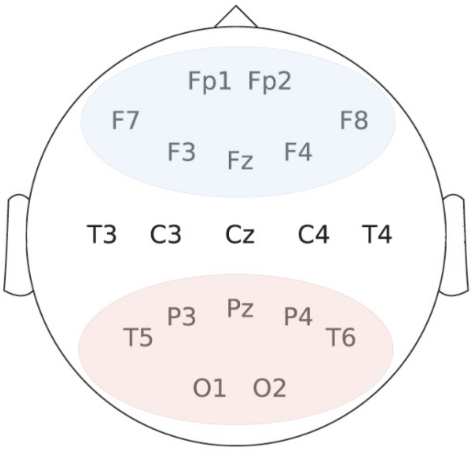

Fig. 2: Topographical map of the 19 electrodes placed following the international 10–20 system (Gemein et al., 2020). The red and blue overlaps present the posterior and anterior regions, respectively. Electrodes placed on the occipital, parietal, temporal and frontal lobes can be distinguished by letters in the image.

Recordings that were not compatible with the 10-20 system were not used and recordings with a sampling rate different than 250 Hz were resampled. Recordings containing less than 5 minutes of signal data were rejected and more than 30 minutes were capped. The first minute of each recording was excluded to accommodate for noise caused by positioning the cap's location on the scalp. The amplitude of the signals were capped at ± 800 μV to reject unphysiological high values. The distribution of values was examined per 10 seconds and segments with a deviation greater than five times the standard deviation were discarded. After preliminary preprocessing a total of 261 pathological and 692 normal recordings were used of male and female patients with ages ranging from 6 to 103 years (see Fig. 1).

2.2 Feature design

Based on the features proposed by Lodder and van Putten (2012) and Van Putten (2007) we specified nine feature definitions: five quantitative features (subsections 2.2.2-6) and four spatial features (subsections 2.2.7-9). Three features capture specific properties of the posterior dominant rhythm (subsections 2.2.2-3 and 2.2.7), five features quantify changes in the power-ratio between specific channel pairs (subsections 2.2.4-6 and 2.2.9) and a synchrony measure between neighboring channels (subsection 2.2.8). The following subsections briefly introduce the features. A more detailed description of the features definitions is found in Appendix 8.2.

2.2.1 Spectral estimation: Welch's averaged periodogram method (welch scipy v1.9.1) was used to estimate spectral power. To be more specific, let $V(t)$ be a matrix of a single recording with $c = \{1...N\}$ channels over time t. The signal data retrieved from channel c was divided into half-overlapping segments (256 data points per segment) with a window length (NFFT) of 512, each of which was detrended and windowed using a Hanning window. The resulting frequency resolution of the estimated spectrum is 1/NFFT Hz. Let the output of Welch's averaged periodogram method be denoted as:

$$P(c, j, f) : c \in [1...N], j \in [1...M], f \in [f_{min}, f_{max}]$$

Each entry in $P(c, j, f)$ contains a discrete Fourier coefficient for channel c, segment j and frequency f. Depending on the features described in the following subsections, either a common reference montage or small laplacian was used to retrieve the signal data.

2.2.2 Alpha rhythm frequency: Healthy adults commonly experience low amplitude and mixed frequency background rhythms (Lodder and van Putten, 2012). The patient's background rhythm is normally characterized by a dominant alpha rhythm in the posterior region (hence posterior dominant rhythm PDR) when the eyes are closed and when the patient is in a relaxed state of wakefulness (Niedermeyer, 1997). The frequency of the alpha rhythm increases with age, beginning at about 4 Hz at four months, 6 Hz at six months, 8 Hz at three years and stabilizing at about 10 Hz at ten years (Eeg-Olofsson, 1971; Lindsley, 1939; Niedermeyer, 2011). From this point, the peak frequency may decline slowly depending on patient-specific characteristics (Aurlien et al., 2004). Peak frequencies have less variability around adulthood although intra-individual variation may arise from fatigue in elderly patients. According to an early study of Brazier and Finesinger (1944) the mean peak frequency of 500 adult control subjects was 10.5 ± 0.9 Hz. However, Khan et al. (2018) reported intra-individual peak frequency variations of ± 0.67 Hz and ± 0.46 Hz for 120 patients with focal and generalized epilepsy, including 40 control subjects all aged from 30 to 40. A reference to discriminate between normal and pathological peak frequencies is based on an estimated mean trend of normal control subjects, earlier outlined by van der Stelt (2008), Segalowitz et al. (2010) and Lodder and van Putten (2011).

Table 2 The mean reference trend accepted for normal alpha rhythm peak frequencies at different ages.

Age (years)	Frequency (Hz)
0-1	5.3 ± 1.8
2-3	6.8 ± 1.8
4-5	7.9 ± 1.8
6-7	8.7 ± 1.8
8-15	9.5 ± 1.8
16-50	9.9 ± 1.8
>51	9.1 ± 1.8

The alpha rhythm frequency is defined as a dominant peak found in the power spectra estimated from channels in the posterior region. To be more specific, the alpha rhythm generator is thought to be located in the occipital lobes. It is not uncommon for the dominant frequency to fluctuate within a recording and change in location, amplitude and width, or not to emerge at all (i.e., absent peaks or multiple significant peaks found). These fluctuations make it difficult to robustly estimate the underlying alpha rhythm frequency.

We proposed an iterative curve-fitting method to localized segments of EEG to approximate the two most dominant peak locations, including their amplitudes and widths. To account for patient-specific characteristics and noisy segments, we quantified the importance of these peaks with respect to other peaks found in the localized spectrum using a correlation computation. Based on these findings, clusters were formed and parameters based on the two largest clusters were used to estimate the alpha rhythm frequency:

$$Q_f(j) = \sum_{i}^{M} f_i w_i \tag{1}$$

with f_i and w_i presenting the estimated frequency and normalized correlation components of the M_j peaks found in a recording. A detailed description of the steps taken is found in Appendix 8.2.1. Table 2. provides the reference trend approximated from normal control subjects. If the estimated peak frequency differs ± 1.8 Hz from the reference trend, the alpha rhythm frequency is considered to deviate from the norm. The results of the estimations are depicted in Fig. 3.

2.2.3 Alpha rhythm reactivity: Reactivity is characterized by an attenuation of rhythmic activity that occurs after the brain receives an external stimulus after an idle state (Schomer and Lopes da Silva, 2010). Eye openings, auditory inputs or pain can all be stimulus types. Dementia is associated with lower reactivity (van der Hiele et al., 2007; Babiloni et al., 2010). The prognostic value of reactivity in comatose patients has been explored by Logi et al. (2011), Douglass et al. (2002) and Ramachandrannair et al. (2005). The levels of attenuation varies with age (Gaál et al., 2010). According to Gaál et al. (2010), lower reactivity may be caused by reduced neuronal connectivity and weaker levels of neurotransmissions in elderly. Reduced reactivity may indicate pathology, but physiological variants may arise in certain states like drowsiness.

Reactivity was quantified as the difference in alpha power measured from EEG segments recorded in the occipital lobe between an idle (relaxed, eyes closed) and non-idle (eyes open) state. Given the estimated peak frequency Q_f, the suppression in alpha power was measured using a narrow frequency band around this peak. If no peak was found, a frequency band of $f \in [2, 12]$ Hz was used instead. Let P_{EC} and P_{EO} be the mean occipital power measured in a narrow frequency band on the estimated peak frequency for segments annotated with the eyes closed and eyes open state. The reactivity in alpha power was obtained following:

$$Q_r = 1 - \frac{P_{EO}}{P_{EC}} \tag{2}$$

Reactivity is substantial for $Q_r \geq 0.5$, moderate for $0.1 < Q_r < 0.5$ and low or absent for $Q_r \leq 0.1$. A detailed description is found in Appendix 8.2.2.

2.2.4 Alpha power anterio–posterior gradient: Rhythmic activity in a normal awake brain is distributed with an anterior to posterior gradient along the scalp (Segalowitz et al., 2010). Schizophrenia and dementia have been linked to abnormalities in the gradient, while drowsiness and deep sleep may cause physiological patterns like slow-wave activity in the posterior region, together with enhanced alpha and theta activity anteriorly (Stevens and Kircher, 1998; Knyazeva et al., 2008). It is therefore important to evaluate the

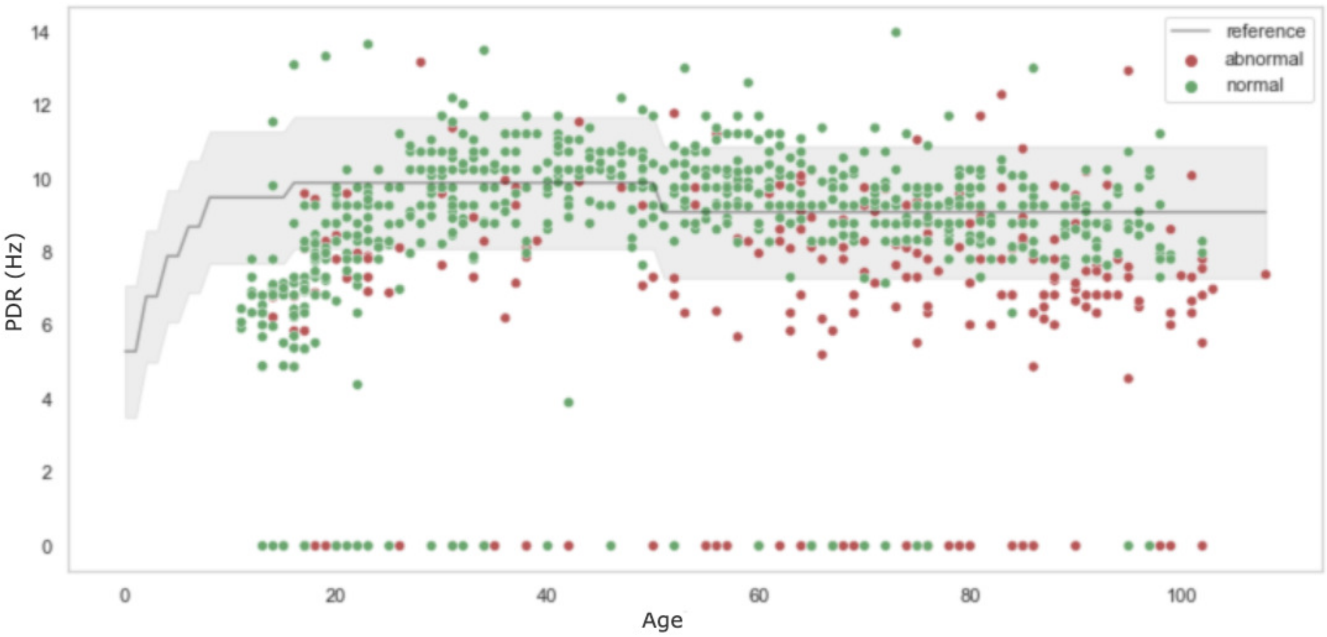

Fig. 3: Results of the alpha rhythm frequency estimation for 953 labeled scalp recordings. The red and green dots denote the pathological and normal labeled recordings. For 89% of the recordings one dominant peak was found, 6% had two dominant peaks and 5% showed no dominant peak (PDR = 0 Hz). The peak frequencies for normal patients follow the suggested mean reference trend, with lower frequencies for younger patients. For normal patients from adulthood to late adulthood, a positive skewed distribution is visible (i.e., more variation at higher frequencies), whereas pathological patients show a negative skewed distribution with more variation at lower frequencies. Note that a significant number of pathological patients showed peak frequencies defined well within the mean reference trend.

gradient within context of the situation to avoid misinterpretations caused by effects of external factors.

A normalized anterio-posterior power gradient was quantified by measuring power in the alpha band for specific channels located in the anterior and posterior regions. With these mean power estimates, the gradient was defined as:

$$Q_{ap} = \frac{P_{ant}}{P_{ant} + P_{pos}} \quad (3)$$

The gradient is considered within normal range for $Q_{ap} \leq 0.4$, moderately differentiated for $0.4 < Q_{ap} < 0.6$ and abnormal or deviant for $Q_{ap} \geq 0.6$. A detailed description is found in Appendix 8.2.3.

2.2.5 Interhemispheric asymmetries: The amplitude of the alpha rhythm frequency is nearly symmetrical for left and right hemispheres although patient variability exist (Segalowitz, 2009). Brains lesions have been linked to asymmetries between left and right hemispheres (Agius Anastasi et al., 2017; Jordan, 2004; van Putten, 2007). According to Maulsby et al. (1968), approximately 60% of healthy adults have asymmetrical amplitude differences, from which 17% have amplitude differences greater than 20% and 1.5% having amplitude differences greater than 50%. Based on these findings, interhemispheric amplitude differences greater than 50% are considered pathological. However, it is possible to observe asymmetries of >50%. Such findings are common in patients with asymptomatic EEG characteristics so asymmetry values should be interpreted with caution (Ebersole and Pedley, 2003).

Asymmetries were quantified by comparing the alpha power ratio between channel pairs in the corresponding hemispheres. Let LR_{avg} be the power ratio for a given asymmetry pair $C_{\{LR\}}$. A single asymmetry value was obtained following:

$$Q_s(C_{\{LR\}}) = \underset{f=\{2...12\}Hz}{mean} \left[LR_{avg}(C_{\{LR\}}, f) \right] \quad (4)$$

As a rule of thumb, pathological asymmetry values were defined by a difference in frequency exceeding 0.5 Hz. A detailed description is found in Appendix 8.2.4.

2.2.6 Diffused slow-wave activity: The presence of diffuse slow-wave activity is uncommon in awake and healthy individuals (Ebersole and Pedley, 2003). Normal variants exist, but the presence of slow-wave activity is usually indicative for pathology. Diffuse slowing is defined as increased power in theta and delta bands, while power is attenuated in the alpha and beta bands. Sedatives and anesthetics may cause physiological patterns of diffused slow-wave activity (Cloostermans et al., 2011; San-juan et al., 2010; Blume, 2006).

Diffuse slow-wave activity is present if the power measured in the eyes closed state has less than one third of power located above 8 Hz. Using this as guideline, a mean spectrum was calculated and the normalised power ratio between $P_{low} = \{2...8\}$ Hz and $P_{wide} = \{2...25\}$ Hz was obtained. The discrete power coefficients in the low and wide spectral bands were summed to find a ratio that quantifies the degree of slow-wave activity:

$$Q_d = \frac{P_{low}}{P_{wide}} \quad (5)$$

Pathological slow-wave activity is present for $Q_d \geq 0.6$ (i.e., less than 40% of power is above 8 Hz). For $Q_d < 0.6$ the recording is considered to have sufficiently fast rhythmic activity. A detailed description is found in Appendix 8.2.5.

2.2.7 Spatial distribution of power in the occipital lobe: Using a common reference montage, let P be the occipital power measured from channel $c \in O1$ within frequency range $f \in [2, 25]$ for segment j. A power spectrum was estimated using Welch's

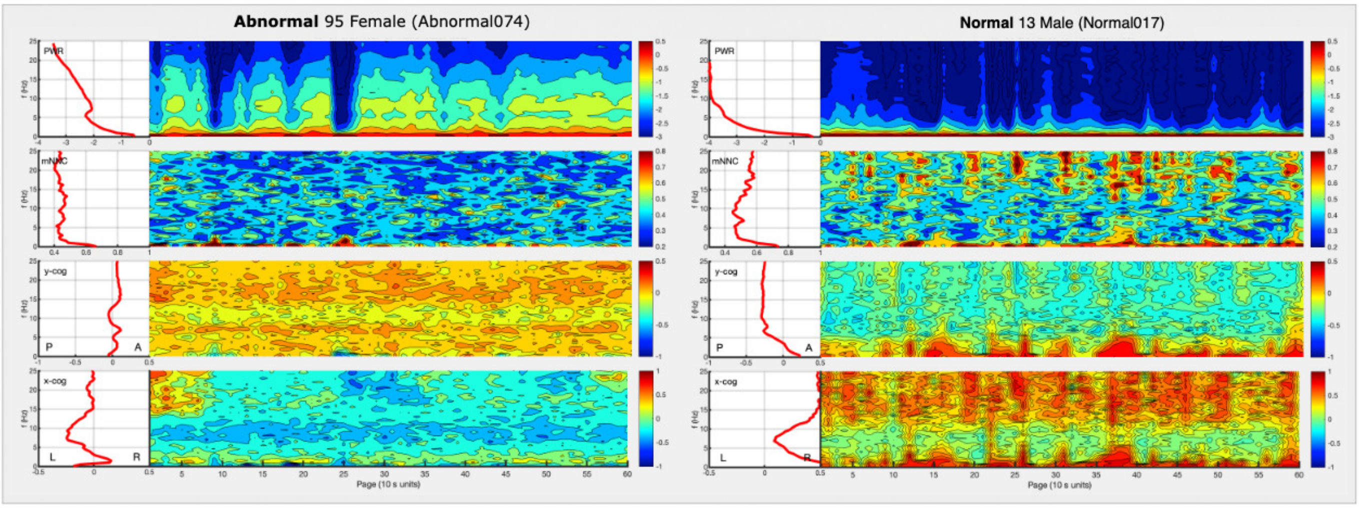

Fig. 4: Colorful Brain visualization of four spatial features. The left image shows a pathological recording from a 95 years old female. PWR shows a wide peak frequency at 8 Hz. mNNC shows a relatively weak signal coherence. COHx and COHy are distributed evenly along the scalp, with more power present in the left hemisphere at the peak frequency. The right image shows a normal recording from a 13 years old male. PWR shows a weak peak frequency at 7 Hz. mNNC shows a relatively weak signal coherence. COHx and COHy are distributed evenly, with more power present in the posterior and right regions.

method for each localized segment annotated with the eyes closed and eyes open state $j \in \{J_{EC}, J_{EO}\}$. Segments annotated during photic stimulation were not used. Let this be denoted as:

$$P(j) = \underset{f=\{2...25\}Hz}{mean} \left[\underset{c \in O1}{mean} \left[P(c,j,f) \right] \right] \tag{6}$$

Each entry in $P(j)$ contains a list of discrete Fourier coefficients for channel c and segment j. The individual periodograms were concatenated to output the spatial distribution of power in the occipital lobe (PWR).

2.2.8 *Spatial distribution of nearest neighbor coherence:*

A synchrony feature was quantified by estimating the coherence (coherence scipy v1.9.1) for each electrode position and its nearest neighbors. As an example, the coherence at channel position Cz was derived by the mean of four coherence values: coh(Cz,Fz), coh(Cz,C4), coh(Cz,C3) and coh(Cz,Pz). Likewise, the coherence at channel C3 was derived by the mean of four coherence values: coh(C3,Cz), coh(C3,F3), coh(C3,P3) and coh(C3,T3). This resulted in 19 mean coherence values from which only the maximum values were used (mNNC). Coherence values were calculated, after re-referencing to a small laplacian, for half-overlapping windows with a window length of 512, each of which was detrended and windowed using a Hanning window. The resulting feature is a time- and frequency-dependent function with coherence values ranging from 0 to 1 for different time intervals.

2.2.9 Spatial distribution of the power gradient: Although the nature of the brain is asymmetrical, the EEG is almost left-right symmetric in mean spectral power. However, there is a physiological asymmetry in the anterior to posterior direction depending on various factors. The distribution of power measured along the scalp can be expressed as functions of frequency, separating the 19 channels between two axis (left-right and up-down). Computing the gradients for different segments of EEG, a two-dimensional spatial distribution was obtained.

Using a laplacian montage, let $V(t)$ be a matrix of a single recording with $j = \{1...M\}$ channels over time t. We estimated spectral power following Welch's method and obtained discrete Fourier coefficients A_{ij} with $i = \{1...N\}$ coefficients. A_{ij} was weighted with it's Euclidean distance from reference location Cz(0,0) in the x-direction (dx_j) and y-direction (dy_j). The resulting weighted coefficients reflect the center-of-gravity of spectral power in the x- and y-directions (COHx and COHy) as a function of frequency(i).

$$C_{LR}(i) = \sum_{j=1}^{M} \frac{A_{ij} dx_j}{A_{ij}} \tag{7}$$

$$C_{AP}(i) = \sum_{j=1}^{M} \frac{A_{ij} dy_j}{A_{ij}} \tag{8}$$

2.3 *Experimental setup*

Fair comparisons to strong baselines are essential to compare experiments and to asses the robustness and reliability of the decoding results. However, the variation in inputs, features, training strategies and classifiers makes it difficult to assess which parts contribute significantly to the performance of a model, and which steps do not. Taking this into account, we proposed experiments based on two feature-based designs. The first design involved multichannel inputs to compute numerical features following a *cropped-wise* decoding strategy, and the second design involved single channel inputs to compute spatial features following an *image-based* decoding strategy.

The features and classifier-designs were implemented in Python 3.8.0. The output of the Colorful Brain was retrieved from Matlab. The experiments were performed on Windows OS with an AMD Ryzen 7 3700X 8-Core Processor 3.60 GHz and NVIDIA RTX 3080 Ti GPU. The Deep ConvNet was retrieved from Braindecode (Springenberg, 2017). The VGG-16 was retrieved from Tensorflow (v2.10.0). For the Deep ConvNet we used the same hyperparameters as proposed by Shirrmeister et al. (2017) and the initial weights of the VGG were used from image-net. The binary classifiers were retrieved from Scikit-learn (v0.21). The ConvNets were tuned using optimization framework Optuna (v3.0.2). Individual classifier-designs were optimized using a grid-search approach.

Table 3 All implemented features used for the cropped-wise decoding models sorted by feature domain. Feature domains include CWT/DWT, FT, Time, Background, Patient information and Connectivity.

CWT/DWT	**FT**
Bounded variation	Maximum
Maximum	Minimum
Minimum	Mean
Mean	Power
Power	Power ratio
Power ratio	Spectral entropy
Spectral entropy	Value range
Variance	Variance
Time	
Maximum	Hjorth activity
Minimum	Hjorth mobility
Mean	Hjorth complexity
Median	
Skewness	Zero crossing
Kurtosis	Zero crossing derivative
Line length	Higuchi fractal dimension
Flat spots	Petrosian fractal dimension
Lumpiness	
Energy	SVD entropy
Nonlinear energy	SVD Fisher information
Background	**Patient information**
Alpha rhythm peak frequency	Age
Alpha rhythm reactivity	Gender
Alpha power anterio–posterior gradient	
Interhemispheric asymmetries	**Connectivity**
Diffused slow-wave activity	Phase locking value

2.3.1 *Cropped-wise decoding models:* A baseline set of features was constructed using the time, frequency and connectivity features outlined by Gemein et al. (2020). The dimensionality of the cropped-features was reduced following a median aggregation over segments. A power spectrum was estimated for each channel within the following frequency bands: 0–2 Hz, 2–4 Hz, 4–8 Hz, 8–13 Hz, 13–18 Hz, 18–24 Hz, 24–30 Hz and 30–50 Hz. Gemein et al. (2020) reported superior performance using a Blackman-Harris window with 50% band overlap. In line with their suggestion we used the same window function with a 50% band overlap to weigh all segments (10-s epochs). The CWT and DWT computations (pywavelets v1.4.1) were constructed using a 'Morlet' and 'db4' wavelet, respectively. For each power estimate produced by the CWT, DWT and FT computations, several statistical properties were computed. Furthermore, statistical and time-complexity features were computed for each crop, and the patients' age and gender were added as binary variables. Phase locking values were computed for unique channel combinations within each crop.

An iterative feature selection procedure (selectkbest scikit-learn v1.1.2) was used to find the most optimal set of feature values. We tested both f-test and mutual information statistics, commonly used for numerical input data when the target variable is categorical. Although the decoding performance was not significantly changed, we used the f-test as it produced slightly better performance. A total of 8943 feature values were computed and 5236 values were chosen as baseline set.

We proposed to add the five background features, detailed in subsections 2.2-6, to the baseline set after feature selection. We used five classifier-designs earlier suggested in the literature and a majority voting classifier based on the three best performing classifiers. In summary, per recording we used 5236 feature values and added 12 feature values derived from 5 background features (i.e., 4 specific values and 8 asymmetry values). Thus, per recording 5248 feature values were used (see Fig. 6).

A brief summary of the features used is presented in Table 3. A definition of the decoding strategy is found in Appendix 8.1.1 and the definition of time-complexity features in Appendix 8.3.

2.3.2 *Image-based decoding models:* Singh et al. (2021) and Bajpai et al. (2021) showed promising results utilizing pre-trained ConvNets to extract features from spectogram images derived from raw single channel inputs. In the literature reviewed, two papers used spectrograms derived from a re-referenced T5-O1 channel pair (Yildirim et al., 2018; Bajpai et al., 2021). In line with this we proposed spectograms derived from the O1 channel (subsection 2.7) based on the assumption that this channel captures specific information of the posterior dominant rhythm. In addition, we proposed to use three spatial features as input described in subsections 2.8-9. The four spatial features are visualized in Fig. 4.

A VGG-16 was used to extract feature maps from the spectogram images. To save computational time, we fine-tuned each VGG using a generic design including the Softmax function. Adam stochastic optimizer function was used with a categorical cross entropy to estimate the loss. For each specific model the learned weights were used accordingly, replacing the initial Softmax function with a binary classifier. We used three classifier-designs earlier suggested by Singh et al. (2021).

A definition of the decoding strategy is found in Appendix 8.1.2. The description of the specific hyperparameter settings including train and test losses obtained for each model is found in Appendix 8.4.

2.3.3 *Model evaluation:* All models were trained using 8-fold cross-validation to divide the train set (see Fig. 1) into multiple subsets of train and validation based on the criteria of 85% train and 15% validation. Performance was validated by predicting to the test set and systematically comparing the predictions to the ground truth. Each prediction produced a probability between 0 and 1. The probability was mapped to a binary value and subsequently categorized in one of four categories. Using the counts of the categories, the decoding performance was quantified in specificity and sensitivity ratios:

$$\text{Specificity}(\%) = \frac{TN}{TN + FP} \cdot 100\%$$

$$\text{Sensitivity}(\%) = \frac{TP}{TP + FN} \cdot 100\%$$

with TP, FP, TN and FN the four categories, denoted by the amount of true positives, false positives, true negatives and false negatives, respectively. The relationship between sensitivity and specificity must be understood as a dynamic one. The ability to correctly identify patients with pathology usually comes at the expense of reduced specificity (meaning more false-positives), while high specificity (i.e., the ability to effectively rule out people without pathology) usually results in lower sensitivity (meaning more false-negatives).

McNemar's test was used to compare prediction errors between models (mcnemar statsmodels v0.13.2). The hypothesis was formulated as follows:

$H0$: prediction errors are significantly different.
$H1$: prediction errors are not significantly changed.

To test the hypothesis, a 2x2 contingency table was found. In case that both classifiers were correct in their predictions one value was counted for contingency value yes/yes. If one prediction was correct and the other incorrect, one value was counted for yes/no or no/yes, depending on the specific order. If both predictions were incorrect, one value was counted for no/no. H_0 was rejected for $p < 0.05$, meaning that the prediction errors of the classifiers differed significantly from each other. A continuity correction was applied.

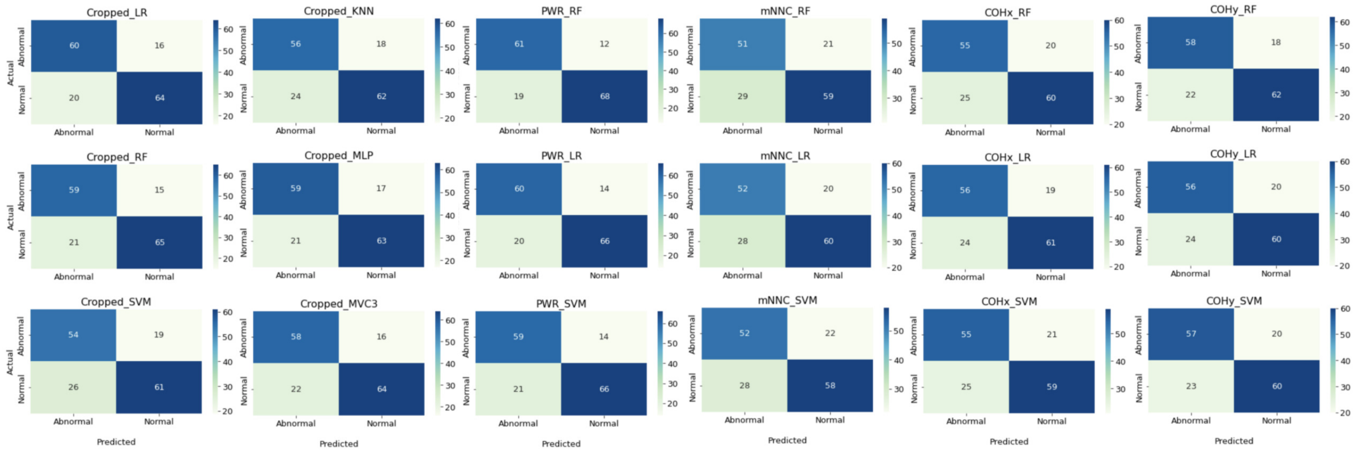

Fig. 5: Evaluation of the cropped-wise and image-based decoding models. Each confusion matrix presents the amount of true positives (upper-left), false positives (upper-right), false negatives (lower-left) and true negatives (lower-right) predicted by it's corresponding model. Note that all models predicted more false negatives (pathological examples predicted as normal) compared to false positives (normal examples predicted as pathological).

3 Results

3.1 *Evaluation of decoding models*

We present the decoding results of the cropped-wise and image-based models in Table 4 and their corresponding confusion matrices in Fig. 5. For the cropped-wise models, the random forest classifier achieved the highest performance with a specificity of 81.25% and sensitivity of 73.75%. Both linear regression and the voting classifier showed comparable results although their p-values were less significant. Overall, the support vector machine classifier resulted in the poorest performance. This result was in line with the results of Gemein et al. (2020) that showed that all feature-based models were congruent for bootstrapped accuracies, except for support vector machine. As can be seen in Table 4, the cropped-wise models did not produce predictions that were significantly different from the baseline. Even though individual classifiers showed variations in performance, none of them produced predictions significantly different.

In our evaluation we also assessed whether predictions would differ significantly with or without specific background features. We constructed an experiment using the baseline set of features, as described in subsection 2.3.1, and included one of the five background features separately. For the five one-to-one comparisons the p-values did not show any significant differences for different classifiers. Interestingly, we also attempted to include all five background features to the baseline set and also experimented with different classifiers. As a result, the highest p-value obtained showed a p-value of 0.032, which indicated that including all five features did significantly influence the decoding performance for all classifiers.

Besides computing the background features as numerical values, we also attempted to convert them into binary values based on the specific thresholds described in subsections 2.2.2-6. We noticed a decrease in performance for all classifiers. Individual inspection of the estimates showed that the majority of feature values were classified in a group (i.e., pathological or normal) that did not matched the predefined labels. Especially many patients with normal alpha rhythms were classified as pathological in this case.

Among the image-based decoding models, the PWR feature produced the best results with specificities and sensitivities ranging between 82.5-85.0% and 73.75-76.25% for different classifiers. For this spatial feature in particular, the p-values indicated that the prediction errors were significantly different compared to the baseline, regardless of the classifier used. The VGG's trained on the COHx, COHy and mNNC features produced worse results. As can be seen in Table 4, the decoding performance was clearly affected by the spatial feature used to train the VGG.

Moreover, all decoding models were found to have a higher ratio of false negatives than false positives (see Fig. 5), meaning that the models were more likely to classify pathological examples (pathological predicted as normal) than non-pathological examples (normal predicted as pathological). These results are in line with those of Lopez et al. (2017), Schirrmeister et al. (2017), Gemein et al. (2020) and Van Leeuwen et al. (2019).

Table 4 Decoding results for the cropped-wise and image-based decoding models. The p-values were computed comparing individual performances to the baseline performance (Deep ConvNet) using McNemar's test. A p-value of <0.05 indicates that the prediction errors are significantly different.

Input	Extraction	Classifier	Sens (%)	Spec (%)	p-value
Raw	Deep ConvNet	Softmax	73.75	80.00	-
Cropped		LR	75.00	80.00	0.32
Cropped		SVM	67.50	76.25	0.90
Cropped		RF	73.75	81.25	0.16
Cropped		KNN	70.00	77.50	0.56
Cropped		MLP	73.75	78.75	0.71
Cropped		MVC3	72.50	80.00	0.57
PWR	**VGG-16**	**RF**	**76.25**	**85.00**	**0.018**
PWR	**VGG-16**	**LR**	**75.00**	**82.50**	**0.035**
PWR	**VGG-16**	**SVM**	**73.75**	**82.50**	**0.047**
mNNC	VGG-16	RF	63.75	73.75	0.73
mNNC	VGG-16	LR	65.00	75.00	0.62
mNNC	VGG-16	SVM	65.00	72.00	0.97
COHx	VGG-16	RF	68.75	75.00	0.87
COHx	VGG-16	LR	70.00	76.25	0.59
COHx	VGG-16	SVM	68.75	73.75	0.89
COHy	VGG-16	RF	72.50	77.50	0.56
COHy	VGG-16	LR	70.00	75.00	0.67
COHy	VGG-16	SVM	71.25	75.00	0.72

3.2 *Evaluation of features*

Fig. 6 shows the total amount of feature values used, with a high F-statistic indicating more "important" features (i.e., the relation between the feature values and labels describes a significant amount of variance). Considering the huge amount of produced feature values, visualizing them separately would be impractical. Therefore,

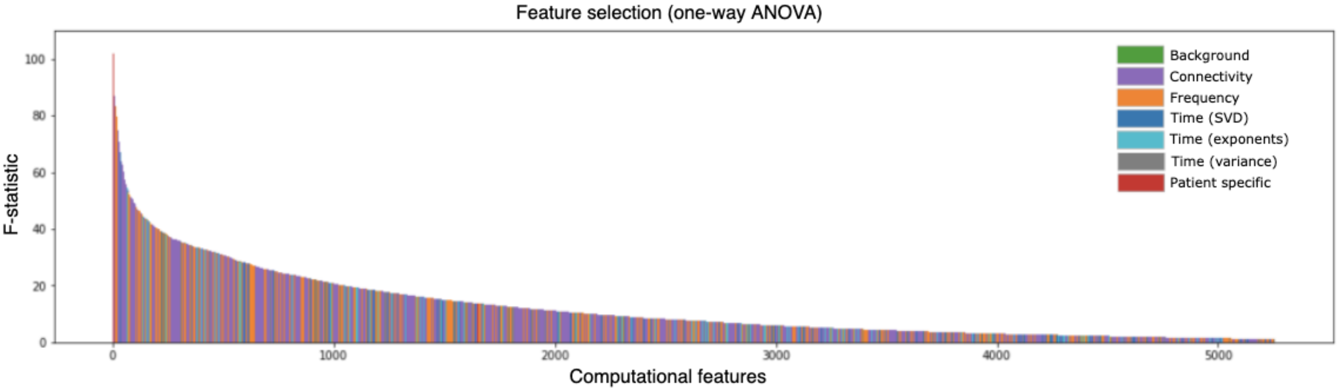

Fig. 6: Result of the feature selection procedure. From the initial 8943 computed feature values, only 5236 values were selected in the baseline set. Each feature value was classified into one of seven classes based on their general feature domain. After including the background features a total of 5248 feature values were used per recording.

each feature value was grouped in one of seven classes based on their general feature domain. From Fig. 6 it becomes clear that connectivity and frequency features were dominantly important. The computation of these features produced many feature values, although not all values were as important. For instance, the frequency estimates for combinations with channels C3, F7 and O1 produced much higher values for the F-statistic in different frequency bands, while other combinations did not have logical or recurrent patterns. This was also true for phase locking values with higher F-statistic values for channel combinations C3, F7, O1 and O2. The patient's age was the second most important factor. While a substantial amount of feature values were used, two background features appeared in the top 250, with the most important one the anterio-posterior gradient followed by the alpha rhythm frequency. Other background features were less important.

Moreover, we combined the spectogram images generated from the spatial features with pixel-space visualizations using gradient-weighted class activation mapping (Grad-CAM). Feature maps were extracted at the last layer of the network before classification. Subsequently, the pixel intensities were combined with the original spectrogram image to produce a single feature visualization. The PWR feature visualizations showed that the gradients highlighted regions within the alpha band. In fact, for all individual PWR feature visualizations a similar region was noticeable, with a few outliers showing regions at higher or lower frequencies. A common pattern was that if no well-defined PDR was visible, the uncertainty in the prediction was higher, and vice versa. Likewise, the COHx and COHy feature visualizations presented similar patterns with highlighted regions in the alpha band, mostly centered at the frequency of the PDR. In contrast, the visualizations for the mNNC feature highlighted regions scattered along the image. These visualizations were less obvious to understand, as no logic or recurrent patterns were noticeable.

4 Discussion

Although this paper supports the claim that decoding performance can be improved using features of the background pattern, the interpretation of the experimental results should be taken with care. Despite the fact that individual aspects of the entire decoding approach were tested (i.e., different training strategies, inputs, features and classifiers), the experimental design choices and the availability of labeled recordings could have greatly affected the results.

For instance, compared to the amount of recordings in TUH Abnormal Corpus (v2.0.0 contains approximately 1488 pathological and 1529 normal recordings), there were significantly fewer ones available in our dataset. Even so, the amount of pathological vs. normal was unevenly distributed, with only 181 pathological recordings used for training. We chose to divide the available recordings into 85% train and 15% test since fewer pathological recordings were available. To ensure an even comparison, we used the same distribution of pathological and normal recordings in the test set. Overall, lowering the amount of recordings in the test set would in most cases lead to a better performance as more recordings are available for training. However, in our experiments the test set may not be sufficiently representative to fully test the models' ability to generalize to unseen recordings. We observed that all models, even the well-established Deep ConvNet, produced lower performances when compared to previous experiments.

Furthermore, the preprocessing steps were kept relatively simple, mainly aimed at rejecting unphysiological values, though subtle artifacts may not have been detected. Computing features from noisy EEG segments generally produces poor estimates. This became apparent when evaluating the produced values of the background features with the help of the visual output of the Colorful Brain. Developing more sophisticated methods for artifact removal might have helped to gain better decoding performance. Although Gemein et al. (2020) proposed to reduce the dimensionality of the initial features using principle component analysis, they reported a decrease in decoding performance. However, considering the circumstances in which our experiments were conducted, such techniques might have been useful to detect and discard less obvious artifacts.

Moreover, the spectogram images produced by the Colorful Brain method were retrieved from Matlab. Specific color-maps can have a great impact on the output visualization of three-dimensional data which, in turn, might have affected the training procedure. As an example, the color-map used in Matlab is different compared to the color-maps in Python. Initially, this may not seem like a major factor, but the differences between the spectogram images were substantial. Besides different colors, there was a noticeable difference in the level of detail as well. This was verified by generating spectogram images using Python from the same channel as used in Matlab.

5 Conclusion

In this paper we showed that the features proposed by Lodder and van Putten (2012) and van Putten (2007) can be applied within existing feature-based models to improve overall decoding performance. We developed multiple feature-based models and assessed the contribution of individual components to the final decoding performance, including the difference between cropped-wise and image-based strategies, computational and spatial features and specific classifier-designs.

Although not all proposed models showed increased performance, the time-localised frequency information derived from a single channel located in the occipital region yielded the best decoding results for different classifiers, with specificities ranging between 82.5-85.0% and sensitivities between 73.75-76.25%. Compared to the Deep ConvNet, the combination of this spatial feature with a VGG approach to extract specific feature maps significantly changed the decoding performance. Despite that the same VGG approach was followed for the COHx, COHy and mNNC spatial features, these performances were not comparable to the baseline indicating that the PWR feature clearly had the most discriminatory power. In line with the results of Yildirim et al. (2018) and Bajpai et al. (2021), using features that captures spatial properties of the PDR is clearly effective to discriminate between pathological vs. non-pathological. Furthermore, individual experiments with configurations of computational and background features showed that including the five background features to the baseline set showed a significant increase in performance, with the anterio-posterior gradient and the alpha rhythm frequency being the most important features. Nonetheless, none of the cropped-wise models outperformed the Deep ConvNet.

The models validated on this dataset performed lower than previously achieved on the TUH Abnormal Corpus. Given this fact, they all performed in a narrow range as previously denoted by Shirrmeister et al. (2017), Van Leeuwen et al. (2019) and Gemein et al. (2020). Considering the contributions of individual components the features being used were much more important than the choice of binary classifiers. Also the specific configuration (i.e., multichannel inputs using a cropped-wise decoding strategy vs. single channel inputs using an image-based decoding strategy) was important. Although this paper didn't experimented with different lengths of inputs, finding the optimal ratio between noise and discriminatory information is important and remains an open challenge. The fact that our multichannel approach performed lower than the single channel approach could be linked to this ratio.

The general ability of ConvNets to learn complex, and not obvious patterns, makes them suitable for complex classification problems. However, this would not always result in a model that is able to capture expected patterns to make clear and understandable decoding decisions. Especially in the domain of clinical neurology where model explainability and interpretability is key to use them in practice. In comparison with earlier approaches validated on the TUH Abnormal Corpus, our feature-based models performed similarly to the end-to-end baseline, though more elaborate and comprehensive steps were required.

In closing, a lot of work has been done to improve current pathology decoding models, but there is a lot more room for improvement. For instance, from our cropped-wise experiments it became clear that the frequency and connectivity features computed from the C3, F7 and O1/O2 channels described more variance in relation with the labels. Although we used the O1 channel to compute spectograms, Singh et al. (2021) experimented with different single channel tests from which the F7+STFT combination yielded the best performance. Remarkably, the F7 channel tends to be discriminative for different datasets. This is a particularly useful insightful as the TUH Abnormal Corpus includes patients diagnosed with epilepsy, among other neurological diseases (validated using the available EEG-reports). This would suggest that this channel might be useful to detect diseases other than epilepsy. Doing additional test with single channels located in the left temporal/frontal lobes might be useful to make decisive conclusions whether this is true.

From the results in Table 4 it is noticeable that the COHx, COHy and mNNC features did not produce results in the range of other models. Using each feature separately as input might not have been sufficient to solely detect pathological recordings. Rather than experimenting with single spatial features, a combinatorial approach using multiple spatial features as input could be an alternative approach to resolve this issue. Although slight modifications to the network are required to accommodate for the different input shape, and the model will become more complex to tune, this could be a useful direction for utilizing multiple background features in one model.

6 Declaration of competing interest

The authors declare no competing financial interests that could have appeared to influence the work reported in this paper.

7 References

Abarbanel, H. D., Brown, R., amp; Kadtke, J. B. (1990). Prediction in chaotic nonlinear systems: Methods for time series with broadband Fourier Spectra. Physical Review A, 41(4), 1782–1807. https://doi.org/10.1103/physreva.41.1782

Acharya, U., Oh, S., Hagiwara, Y., Tan, J., Adeli, H. (2018). Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. Computers In Biology And Medicine, 100, 270-278. https://doi.org/10.1016/j.compbiomed.2017.09.017

Agius Anastasi, A., Falzon, O., Camilleri, K., Vella, M., Muscat, R. (2017). Brain Symmetry Index in Healthy and Stroke Patients for Assessment and Prognosis. Stroke Research And Treatment, 2017, 1-9. https://doi.org/10.1155/2017/8276136

Amin, S., Hossain, M. S., Muhammad, G., Alhussein, M., amp; Rahman, A. (2019). Cognitive Smart Healthcare for Pathology Detection and Monitoring., IEEE, 1–8. https://doi.org/10.1109/ACCESS.2019.2891390.

Anis, A., Lloyd, E. (1976). The Expected Value of the Adjusted Rescaled Hurst Range of Independent Normal Summands. Biometrika, 63(1), 111. https://doi.org/10.2307/2335090 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS.

Aurlien, H., Gjerde, I., Aarseth, J., Eldøen, G., Karlsen, B., Skeidsvoll, H., Gilhus, N. (2004). EEG background activity described by a large computerized database. Clinical Neurophysiology, 115(3), 665-673. https://doi.org/10.1016/j.clinph.2003.10.019

Babiloni, C., Lizio, R., Vecchio, F., Frisoni, G., Pievani, M., Geroldi, C. et al. (2011). Reactivity of Cortical Alpha Rhythms to Eye Opening in Mild Cognitive Impairment and Alzheimer's Disease: an EEG Study. Journal Of Alzheimer's Disease, 22(4), 1047-1064. https://doi.org/10.3233/jad-2010-100798

Bajpai, R., Yuvaraj, R., Prince, A. (2021). Automated EEG pathology detection based on different convolutional neural network models: Deep learning approach. Computers In Biology And Medicine, 133, 104434. https://doi.org/10.1016/j.compbiomed.2021.104434

Boutros, N. N., Arfken, C., Galderisi, S., Warrick, J., Pratt, G., Iacono, W. (2008). The status of spectral EEG abnormality as a diagnostic test for schizophrenia. Schizophrenia Research, 99(1-3), 225–237. https://doi.org/10.1016/j.schres.2007.11.020

Blume, W. (2006). Drug Effects on EEG. Journal Of Clinical Neurophysiology, 23(4), 306-311. https://doi.org/10.1097/01.wnp.0000229137.94384.fa

Brazier, M., Finesinger, J., Schwab, R. (1944). Characteristics of the normal electroencephalogram. II. The effect of varying blood sugar levels on the occipital cortical potentials in adults during quiet breathing 1. Journal Of Clinical Investigation, 23(3), 313-317. https://doi.org/10.1172/jci101496

Casdagli, M. (1989). Nonlinear prediction of Chaotic Time Series. Physica D: Nonlinear Phenomena, 35(3), 335–356. https://doi.org/10.1016/0167-2789(89)90074-2

Chollet, F. (2022). GitHub - keras-team/keras: Deep Learning for humans. GitHub. Retrieved 16 August 2022, from https://github.com/keras-team/keras.

Cloostermans, M., de Vos, C., van Putten, M. (2011). A novel approach for computer assisted EEG monitoring in the adult ICU. Clinical Neurophysiology, 122(10), 2100-2109. https://doi.org/10.1016/j.clinph.2011.02.035

Douglass, L., Wu, J., Rosman, N., Stafstrom, C. (2002). Burst Suppression Electroencephalogram Pattern in the Newborn: Predicting the Outcome. Journal Of Child Neurology, 17(6), 403-408. https://doi.org/10.1177/088307380201700601

Ebersole JS, Pedley TA. (2003). Current practice of clinical electroencephalography. LWW medical book collection. Lippincott Williams Wilkins.

Gaál, Z., Boha, R., Stam, C., Molnár, M. (2010). Age-dependent features of EEG-reactivity—Spectral, complexity, and network characteristics. Neuroscience Letters, 479(1), 79-84. https://doi.org/10.1016/j.neulet.2010.05.037

Gemein, L., Schirrmeister, R., Chrabąszcz, P., Wilson, D., Boedecker, J., Schulze-Bonhage, A. et al. (2020). Machine-learning-based diagnostics of EEG pathology. Neuroimage, 220, 117021. https://doi.org/10.1016/j.neuroimage.2020.117021

Geng, D., Alkhachroum, A., Melo Bicchi, M., Jagid, J., Cajigas, I., Chen, Z. (2021). Deep learning for robust detection of interictal epileptiform discharges. Journal Of Neural Engineering, 18(5), 056015. https://doi.org/10.1088/1741-2552/abf28e IEEE Signal Processing in Medicine and Biology Symposium., 12.

Horváth, A. (2016). The Value of Long-Term EEG in the Diagnosis of Epilepsy in Alzheimer's Disease. Journal of Neurology Stroke, 4(2). https://doi.org/10.15406/jnsk.2016.04.00125

Jin, Y., Potkin, S., Kemp, A., Huerta, S., Alva, G., Thai, T. et al. (2005). Therapeutic Effects of Individualized Alpha Frequency Transcranial Magnetic Stimulation (TMS) on the Negative Symptoms of Schizophrenia. Schizophrenia Bulletin, 32(3), 556-561. https://doi.org/10.1093/schbul/sbj020

Juhàsz, C., Zrirmai, I. (1997). Changes of spectral EEG variables in patients with hemispheric stroke. Electroencephalography and Clinical Neurophysiology, 102(1), P6. https://doi.org/10.1016/s0013-4694(97)86228-1

Khan, A., Paulus, W., Stephani, C. (2018). Short-term intraindividual variability of the posterior dominant alpha frequency in the electroencephalogram. Clinical Neurophysiology, 129(1), 208-209. https://doi.org/10.1016/j.clinph.2017.11.002

Knyazeva, M., Jalili, M., Meuli, R., Hasler, M., De Feo, O., Do, K. (2008). 65 – Hypofrontality and EEG alpha rhythm in schizophrenia. Schizophrenia Research, 98, 60. https://doi.org/10.1016/j.schres.2007.12.132

Lachaux, J., Rodriguez, E., Martinerie, J., Varela, F. (1999). Measuring phase synchrony in brain signals. Human Brain Mapping, 8(4), 194-208. https://doi.org/10.1002/(sici)1097-0193(1999)8:4<194::aid-hbm4>3.0.co;2-c

Levin KH, Luders.H (2000). Comprehensive clinical neurophysiology. W.B. Saunders.

Lindsley, D. (1939). A Longitudinal Study of the Occipital Alpha Rhythm in Normal Children: Frequency and Amplitude Standards. The Pedagogical Seminary And Journal Of Genetic Psychology, 55(1), 197-213. https://doi.org/10.1080/08856559.1939.10533190

Lodder, S., van Putten, M. (2013). Quantification of the adult EEG background pattern. Clinical Neurophysiology, 124(2), 228-237. https://doi.org/10.1016/j.clinph.2012.07.007

Logi, F., Pasqualetti, P., Tomaiuolo, F. (2011). Predict recovery of consciousness in post-acute severe brain injury: The role of EEG reactivity. Brain Injury, 25(10), 972-979. https://doi.org/10.3109/02699052.2011.589795

Lopez, S., Suarez, G., Jungreis, D. , Obeid, I., and Picone, J. (2017) Automated identification of abnormal adult eegs.

Mandelbrot, B. (1967). How Long Is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension. Science, 156(3775), 636-638. https://doi.org/10.1126/science.156.3775.636

Mandelbrot, B., Wallis, J. (1969). Robustness of the rescaled range R/S in the measurement of noncyclic long run statistical dependence. Water Resources Research, 5(5), 967-988. https://doi.org/10.1029/wr005i005p00967

Maulsby RL., Kellaway P., and Graham M. (1968) The normative electroencephalographic data reference library. Final report, contract NAS-9-1200 Washington DC National aeronautics and space administration.

Moghadam, S., Pinchefsky, E., Tse, I., Marchi, V., Kohonen, J., Kauppila, M. et al. (2021). Building an Open Source Classifier for the Neonatal EEG Background: A Systematic Feature-Based Approach From Expert Scoring to Clinical Visualization. Frontiers In Human Neuroscience, 15. https://doi.org/10.3389/fnhum.2021.675154

Morita, A., Kamei, S., amp; Mizutani, T. (2011). Relationship between slowing of the EEG and cognitive impairment in parkinson disease. Journal of Clinical Neurophysiology, 1. https://doi.org/10.1097/wnp.0b013e3182273211

Mormann, F., Lehnertz, K., David, P., E. Elger, C. (2000). Mean phase coherence as a measure for phase synchronization and its application to the EEG of epilepsy patients. Physica D: Nonlinear Phenomena, 144(3-4), 358-369. https://doi.org/10.1016/s0167-2789(00)00087-7

Murphy, J. P. (1957). The Role of the EEG in the Differential Diagnosis of Brain Tumor. Southern Medical Journal, 50(8), 1013–1017. https://doi.org/10.1097/00007611-195708000-00008

Niedermeyer, E. (1997). Alpha rhythms as physiological and abnormal phenomena. International Journal Of Psychophysiology, 26(1-3), 31-49. https://doi.org/10.1016/s0167-8760(97)00754-x

Niedermeyer, E., Schomer, D., Lopes da Silva, F. (2011). Niedermeyer's electroencephalography. Wolters Kluwer Health/Lippincott Williams Wilkins.

Obeid, I. and Picone, J., 2016. The Temple University Hospital EEG Data Corpus. Frontiers in Neuroscience, 10.

Pedregosa, F., Varoquaux, G., Gramfort, A., Thirion, M., Grisel, O., Blondel, M. et al. (2022). GitHub - scikit-learn/scikit-learn: scikit-learn: machine learning in Python. GitHub. Retrieved 16 August 2022, from https://github.com/scikit-learn/scikit-learn.

Peh, W., Thomas, J., Bagheri, E., Chaudhari, R., Karia, S., Rathakrishnan, R. et al. (2021). Multi-Center Validation Study of Automated Classification of Pathological Slowing in Adult Scalp Electroencephalograms Via Frequency Features. International Journal Of Neural Systems, 31(06), 2150016. https://doi.org/10.1142/s0129065721500167

Peng, C., Buldyrev, S., Havlin, S., Simons, M., Stanley, H., Goldberger, A. (1994). Mosaic organization of DNA nucleotides. Physical Review E, 49(2), 1685-1689. https://doi.org/10.1103/physreve.49.1685

Petersén, I., Eeg-Olofsson, O. (1971). The Development of the Electroencephalogram in Normal Children from the Age of

1 Through 15 Years – Non-paroxysmal activity. Neuropediatrics, 2(03), 247-304. https://doi.org/10.1055/s-0028-1091786

RamachandranNair, R., Sharma, R., Weiss, S., Cortez, M. (2005). Reactive EEG Patterns in Pediatric Coma. Pediatric Neurology, 33(5), 345-349. https://doi.org/10.1016/j.pediatrneurol.2005.05.007

Rizal, A., Nugroho, H., Hidayat, R. (2018). Fractal Dimension for Lung Sound Classification in Multiscale Scheme. Journal Of Computer Science, 14(8), 1081-1096. https://doi.org/10.3844/jcssp.2018.1081.1096

Rommel, C. (2022). GitHub - braindecode/braindecode: Deep learning software to decode EEG, ECG or MEG signals. GitHub. Retrieved 16 August 2022, from https://github.com/braindecode/braindecode.

Rosenstein, M. T., Collins, J. J., amp; De Luca, C. J. (1993). A practical method for calculating largest Lyapunov exponents from small data sets. Physica D: Nonlinear Phenomena, 65(1-2), 117–134. https://doi.org/10.1016/0167-2789(93)90009-p

Roy, S., Kiral-Kornek, I. and Harrer, S. (2018). Deep learning enabled automatic abnormal eeg identification. Proceedings of the

Sananman, M. L. (1983). EEG Vs. Computerized Tomography of Brain in Neurological Diagnosis. Clinical Electroencephalography, 14(3), 116–129. https://doi.org/10.1177/155005948301400305

San-juan, D., Chiappa, K., Cole, A. (2010). Propofol and the electroencephalogram. Clinical Neurophysiology, 121(7), 998-1006. https://doi.org/10.1016/j.clinph.2009.12.016

Sanches, P., Tabaeizadeh, M., Moura, L., Rosenthal, E., Caboclo, L., Hsu, J. et al. (2022). Anti-seizure medication treatment and outcomes in acute ischemic stroke patients undergoing continuous EEG monitoring. Neurological Sciences. https://doi.org/10.1007/s10072-022-06183-9

Schirrmeister, R., Springenberg, J., Fiederer, L., Glasstetter, M., Eggensperger, K., Tangermann, M. et al. (2017). Deep learning with convolutional neural networks for EEG decoding and visualization. Human Brain Mapping, 38(11), 5391-5420. https://doi.org/10.1002/hbm.23730

Segalowitz, S., Santesso, D., Jetha, M. (2010). Electrophysiological changes during adolescence: A review. Brain And Cognition, 72(1), 86-100. https://doi.org/10.1016/j.bandc.2009.10.003

Shannon, C. (1948). A Mathematical Theory of Communication. Bell System Technical Journal, 27(4), 623-656. https://doi.org/10.1002/j.1538-7305.1948.tb00917.x

Singh, R., Ahmed, T., Kumar Singh, A., Chanak, P., Singh, S. (2021). SeizSClas: An Efficient and Secure Internet-of-Things-Based EEG Classifier. IEEE Internet Of Things Journal, 8(8), 6214-6221. https://doi.org/10.1109/jiot.2020.3030821

Springenberg, M. Schirrmeister, R. (2017). Braindecode. https://github.com/braindecode/braindecode/

Spronk, D., Arns, M., Barnett, K., Cooper, N., Gordon, E. (2011). An investigation of EEG, genetic and cognitive markers of treatment response to antidepressant medication in patients with major depressive disorder: A pilot study. Journal Of Affective Disorders, 128(1-2), 41-48. https://doi.org/10.1016/j.jad.2010.06.021

Stevens, A., Kircher, T. (1998). Cognitive decline unlike normal aging is associated with alterations of EEG temporo-spatial characteristics. European Archives Of Psychiatry And Clinical Neuroscience, 248(5), 259-266. https://doi.org/10.1007/s004060050047

Talagala, T., Li, F., Kang, Y. (2022). FFORMPP: Feature-based forecast model performance prediction. International Journal Of Forecasting, 38(3), 920-943. https://doi.org/10.1016/j.ijforecast.2021.07.002

Thibodeau, R., Jorgensen, R. S., Kim, S. (2006). Depression, anxiety, and resting frontal EEG asymmetry: A meta-analytic review. Journal of Abnormal Psychology, 115(4), 715–729. https://doi.org/10.1037/0021-843x.115.4.715

Tjepkema-Cloostermans, M., de Carvalho, R., van Putten, M. (2018). Deep learning for detection of focal epileptiform discharges from scalp EEG recordings. Clinical Neurophysiology, 129(10), 2191-2196. https://doi.org/10.1016/j.clinph.2018.06.024

Yıldırım, Ö., Baloglu, U., Acharya, U. (2018). A deep convolutional neural network model for automated identification of abnormal EEG signals. Neural Computing And Applications, 32(20), 15857-15868. https://doi.org/10.1007/s00521-018-3889-z

van Leeuwen, K., Sun, H., Tabaeizadeh, M., Struck, A., van Putten, M., Westover, M. (2019). Detecting abnormal electroencephalograms using deep convolutional networks. Clinical Neurophysiology, 130(1), 77-84. https://doi.org/10.1016/j.clinph.2018.10.012

van Putten, M. (2008). The Colorful Brain: Visualization of EEG Background Patterns. Journal Of Clinical Neurophysiology, 25(2), 63-68. https://doi.org/10.1097/wnp.0b013e31816bdf85

van Putten, M., Kind, T., Visser, F., Lagerburg, V. (2005). Detecting temporal lobe seizures from scalp EEG recordings: A comparison of various features. Clinical Neurophysiology, 116(10), 2480-2489. https://doi.org/10.1016/j.clinph.2005.06.017

van Putten, M., de Carvalho, R., Tjepkema-Cloostermans, M. (2018). Deep learning for detection of epileptiform discharges from scalp EEG recordings. Clinical Neurophysiology, 129, e98-e99. https://doi.org/10.1016/j.clinph.2018.04.248

van der Hiele, K., Vein, A., Reijntjes, R., Westendorp, R., Bollen, E., van Buchem, M. et al. (2007). EEG correlates in the spectrum of cognitive decline. Clinical Neurophysiology, 118(9), 1931-1939. https://doi.org/10.1016/j.clinph.2007.05.070

van der Stelt, O. (2008). Development of human EEG posterior alpha rhythms. Clinical Neurophysiology, 119(8), 1701-1702. https://doi.org/10.1016/j.clinph.2008.04.001

8 Appendices

8.1 Appendix A: Definition of decoding strategies

8.1.1 Cropped-wise decoding strategy: Let V_i be a matrix containing all feature values for recording $i \in I$ with $I = 953$. The dimension of the matrix is denoted as $V_i \in \mathbb{R}^{C_i \times F}$. Note that C can be different for individual recordings. Each row in the matrix consist of a feature vector with an equal amount of computed feature values per analysed segment:

$$V_i = \begin{bmatrix} X_{11} & X_{12} & ... & X_{1M} \\ X_{21} & X_{22} & ... & X_{2M} \\ ... & ... & ... & ... \\ X_{N1} & X_{N2} & ... & X_{NM} \end{bmatrix}$$

with X_{NM} a numerical value for $C_i = \{1...N\}$ segments and $F = \{1...M\}$ feature values. A median aggregation function was used to retrieve one median feature value from all segments. Thus, a column-wise aggregation was performed to retrieve one feature vector per recording. This process was repeated for each recording such that all computational feature values were stored as one aggregated feature matrix $V_{agg} \in \mathbb{R}^{I \times F}$ with corresponding labels $L = \{l_1...l_I\}$ for $l \in [0, 1]$.

8.1.2 Image-based decoding strategy: Let $V_i\{T\}$ be a spectogram image for recording $i \in I$ with $I = 953$ computed from feature $T \in$ {PWR, mNNC, COHx, COHy}:

$$V_i\{T\} = \begin{bmatrix} X_{11} & X_{12} & ... & X_{1M} \\ X_{21} & X_{22} & ... & X_{2M} \\ ... & ... & ... & ... \\ X_{N1} & X_{N2} & ... & X_{NM} \end{bmatrix}$$

with $0 \leq X_{NM} \leq 255$ a pixel value for $\{1...N\}$ frequency values and $\{1...M\}$ time values. By design, the VGG requires a dimensional input of (224, 224, 3). The spectogram images were rescaled to dimensions $N = 244$ and $M = 244$. Thus, four separate datasets of spectogram images were created with corresponding labels $L = \{l_1...l_I\}$ for $l \in [0, 1]$.

8.2 Appendix B: Definition of background features

8.2.1 Alpha rhythm frequency: Using a common reference montage, a power spectrum was computed for channels $c \in \{O1, O2\}$ separately within frequency band $f \in [2, 18]$ using only segments annotated with the eyes closed state $j \in J_{EC}$. For clarity, the following procedures are described for one power spectrum $P(f)$. The estimated power spectrum was log transformed:

$$P_{log}(f) = log(P(f))$$

Dominant peak location: A curve P_{curve} was constructed using three components: two peak components P_{pk1} and P_{pk2} and background component P_{bg}. P_{curve} was fitted to P_{log} and parameters were optimized in a nonlinear least-squares way following the Levenberg–Marquardt method (least_squares scipy v1.9.1).

$$P_{log}(f) \approx P_{curve}(f) = P_{pk1}(f) + P_{pk2}(f) + P_{bg}(f)$$

$$P_{pk1}(f) = A_1 \cdot exp\left(\frac{(f - f_1)^2}{\Delta_1^2}\right)$$

$$P_{pk2}(f) = A_2 \cdot exp\left(\frac{(f - f_2)^2}{\Delta_2^2}\right)$$

$$P_{bg}(f) = B - C \cdot log(f)$$

Parameters A_1 and A_2 are the amplitudes, f_1 and f_2 the center frequencies and Δ_1 and Δ_2 the widths. C is a power-law approximation and B a normalization factor. Parameters were optimized in an iterative way to prevent localized optimizations of components from interfering with the approximation of P_{curve}. Initial values for B and C were set to zero and the parameters were approximated by fitting the background component to P_{log}, while minimizing the error:

$$\{B, C\} = \underset{f \in [2,18]}{argmin} |P_{log} - P_{bg}|$$

After obtaining parameters B and C, parameter C was kept fixed. Initial values for A_1 and f_1 were guessed using methods (find_peaks and peak_prominences scipy v1.9.1) to find the most dominant peak's location and it's amplitude in $(P_{log} - P_{bg})$, while Δ_1 was set to 1. One peak was fitted to P_{log}:

$$\{A_1, f_1, \Delta_1, B\} = \underset{f \in [2,18]}{argmin} |P_{log} - P_{pk1} - P_{bg}|$$

In a similar fashion, initial values for A_2 and f_2 were set based on the second most dominant peak found in $(P_{log} - P_{pk1} - P_{bg})$, while Δ_2 was set to 1. The second peak was fitted to P_{log}:

$$\{A_2, f_2, \Delta_2, B\} = \underset{f \in [2,18]}{argmin} |P_{log} - P_{pk1} - P_{pk2} - P_{bg}|$$

During the iterative process, two evaluations were performed to verify whether P_{curve} was accurately approximated. Firstly, a ratio was calculated between the spectral power of P_{pk1} and P_{log} after one peak had been fitted. It was assumed that the spectrum did not have a dominant frequency component if P_{pk1} contributed less than 50% of power in P_{log}. Secondly, if $f_1 \notin [2, 18]$, or Δ_1 exceeded a threshold TH, the epoch was rejected. In addition, the second peak was discarded by setting A_2 to zero if $\Delta_2 > TH$ or if $f_2 \notin [2, 18]$. The threshold TH was empirically chosen as 2 Hz.

Dominant peak correction: The detected frequency components were improved by an intermediate step after obtaining parameters for each peak in the epoch and approximating the curve. As a result of this iterative approach, the dominant frequency components were robustly located, but the amplitudes of these components were not precisely determined. Searching for the peak around the estimated frequencies in P_{log} was a relatively simple yet effective solution. Assuming center frequencies f_1 and f_2, peak estimates were shifted towards the positive gradients on P_{log} until a local maximum was achieved. In the event that two peaks were present, and both were updated to the same point, one peak was discarded.

Dominant peak estimation: Having updated the peak parameters of P_{curve}, a power ratio was found between the estimated peak components and other, less dominant, peaks found in the localized segments. To do this, the center frequencies f_1 and f_2 were allocated, and a frequency range was found around them based on their center widths Δ_1 and Δ_2. A new spectrum P_{res} was found excluding these frequency ranges:

$$P_{res}(f) = \begin{cases} P(f) & , f \notin R_{peaks} \\ exp(B - C \cdot log(f)) & , f \in R_{peaks} \end{cases}$$

with $R_{peaks} \subseteq [f_{min}, f_{max}]$ being the frequency range around the dominant frequency components. Given the original spectrum P and new spectrum P_{res}, a correlation coefficient was computed:

$$c = 1 - \text{corr}(P, P_{res})$$

If P_{res} was noisy or more peak components were present with relatively high peaks, the correlation parameter was low. Conversely, if peaks contributed to most of the power in P_{log}, a high value for c was obtained.

After obtaining the peak components and correlation coefficient for each localized segment, the parameters were concatenated and sorted on frequency. The components were grouped into clusters provided that their frequencies did not exceed 0.2 Hz, and clusters consisting of one peak or smaller in size than 75% of the largest cluster were discarded. Let all the clusters formed be denoted as $\{C_j\}_{j \in \{1,\ldots,N\}}$ with $C_j = \{A_i, f_i, c_i\}_{i \in \{1,\ldots,M_j\}}$ containing the amplitude, frequency and correlation components of the M_j peaks in the cluster.

Note that the procedure described in the latter was performed for both log spectra P_{O1} and P_{O2}. Parameters based on the two largest clusters formed in the set were used to estimate alpha rhythm frequency. In the event that only one cluster was available, we assumed that the recording contained one rhythm. Based on a given cluster, a weighted average of the frequency was calculated:

$$w_i = \frac{c_i}{\sum_k^M c_k}$$

$$Q_f(j) = \sum_i^M f_i w_i$$

In addition to calculating peak frequencies for a given recording, two further steps were performed. First, if two dominant frequencies were detected, but only one was in the alpha band, the outlying frequency was discarded as it was assumed that it did not belong to the alpha rhythm. Second, if two frequencies were found outside of the alpha range, the one with the largest amplitude was identified as the dominant frequency.

8.2.2 Alpha rhythm reactivity: Using a laplacian montage, let P_{EC} and P_{EO} be the mean occipital power measured in a 0.5 Hz frequency band on the estimated peak frequency for segments with the eyes closed and eyes open state:

$$P_{EC} = \underset{f=\{Q_f \pm .25\}Hz}{mean} \left[P(c,j,f)\right], c \in \{O1\}, j \in J_{EC}$$

$$P_{EO} = \underset{f=\{Q_f \pm .25\}Hz}{mean} \left[P(c,j,f)\right], c \in \{O1\}, j \in J_{EO}$$

Given the power estimates in the eyes closed and open states, the reactivity in alpha power was quantified as:

$$Q_r = 1 - \frac{P_{EO}}{P_{EC}}$$

8.2.3 Alpha power anterio–posterior gradient: Using a laplacian montage, the mean power in the alpha band was computed as:

$$P_{alpha}(c,j) = \underset{f=\{8\ldots12\}Hz}{mean} \left[P(c,j,f)\right]$$

Only segments in the eyes closed state were used and a mean spectrum from these segments was found:

$$P_{EC}(c) = \underset{j \in J_{EC}}{mean} \left[P_{alpha}(c,j)\right]$$

Using channels $\mathbf{c_{ant}} = \{Fp1, Fp2, F7, F8, F3, Fz, F4\}$ and $\mathbf{c_{pos}} = \{T5, T6, P3, P4, Pz, O1, O2\}$, the mean alpha power of the anterior and posterior regions were found following:

$$P_{ant} = \underset{c \in \mathbf{c_{ant}}}{mean} \left[P_{EC}(c)\right]$$

$$P_{pos} = \underset{c \in \mathbf{c_{pos}}}{mean} \left[P_{EC}(c)\right]$$

With these mean power estimates, the normalised anterio-posterior gradient was defined as:

$$Q_{ap} = \frac{P_{ant}}{P_{ant} + P_{pos}}$$

8.2.4 Interhemispheric asymmetries: Using a laplacian montage, let the channel pairs based be denoted as:

$\mathbf{c_{\{LR\}}}$ = $\{\{Fp1, Fp2\}, \{F7, F8\}, \{F3, F4\}, \{T3, T4\}, \{C3, C4\}, \{T5, T6\}, \{P3, P4\}, \{O1, O2\}\}$

A power ratio for each left-right channel pair $C_{\{LR\}} \in \mathbf{c_{\{LR\}}}$ was found for:

$$LR(C_{\{LR\}}, j, f) = \frac{P\left(C_{\{R\}}, j, f\right) - P\left(C_{\{L\}}, j, f\right)}{P\left(C_{\{R\}}, j, f\right) + P\left(C_{\{L\}}, j, f\right)}$$

By calculating the absolute mean of all segments $j = \{1 \ldots M\}$:

$$LR_{avg}(C_{\{LR\}}, f) = \left| \underset{j=\{1\ldots M\}}{mean} \left[LR(C_{\{LR\}}, j, f)\right] \right|$$

and averaging over $f = \{2 \ldots 12\}$ Hz, a single asymmetry value was obtained for each channel pair:

$$Q_s(C_{\{LR\}}) = \underset{f=\{2\ldots12\}Hz}{mean} \left[LR_{avg}(C_{\{LR\}}, f)\right]$$

8.2.5 *Diffuse slow-wave activity:*

Using a laplacian montage, a mean spectrum was first calculated and the normalised power ratio between $P_{low} = \{2...8\}$ Hz and $P_{wide} = \{2...25\}$ Hz was obtained for channels:

$\mathbf{c_{slow}}$ = $\{F7, F8, F3, F4, Fz, T3, T4, T5, T6, C3, C4, Cz, P3, P4, Pz, O1, O2\}$

Using only segments with the eyes closed state, the mean spectrum over all channels was found:

$$P_{EC}(f) = \underset{C \in \mathbf{c_{slow}}}{mean} \left[\underset{j \in J_{EC}}{mean} \left[P(c,j,f) \right] \right]$$

The discrete power coefficients in the low and wide spectral bands were summed following:

$$P_{low} = \sum_{f=2Hz}^{8Hz} P_{EC}(f)$$

$$P_{wide} = \sum_{f=2Hz}^{25Hz} P_{EC}(f)$$

From these estimates, a ratio was found that quantifies the degree of slow-wave activity:

$$Q_d = \frac{P_{low}}{P_{wide}}$$

8.3 *Appendix C: Definition of time-complexity features*

For the following features detailed, let $x(t)$ be a time series with $t > 0$ time points.

8.3.1 *Higuchi fractal dimension:*

Higuchi's method for estimating fractal dimension is iterative in nature. The reconstructed sequence of $x(t)$ was computed following:

$$x_m^k = \left\{ x(m), x(m+k), ..., x\left(m + \left[\frac{N-m}{k}\right] \cdot k\right) \right\}$$

with k the interval between two adjacent time series and $m = 1, 2, ..., k$ the initial value of the reconstructed sequence of $x(t)$. The box length was calculated for each produced sequence x_m^k:

$$L_m(k) = \frac{\sum_{i=1}^{(N-m)/k} |x(m+ik) - x(m+(i-1)k)| \cdot (N-1)}{[(N-m)/k] \cdot k}$$

The average box length of the sequence was computed as:

$$L(k) = \sum_{m=1}^{k} L_m(k)$$

From this, an useful relation is that the log of the average box length behaves proportional to fractal dimension D times the log of reciprocal k:

$$L(k) \propto \left(\frac{1}{k}\right)^D$$

The linear relation between both parts can be described as a function with slope D presenting Higuchi's fractal dimension. We estimated the slope parameter using a least-squares fit (lstsq numpy v1.22.1) by plotting $\log(L(k))$ as a function of $\log\left(\frac{1}{k}\right)$.

8.3.2 *Petrosian fractal dimension:*

Petrosian's method has similarities with Higuchi's method but differs in a rule to binarize $x(t)$ before estimating the fractal dimension. Suppose that each value was transformed to a binary value following:

$$z_i = \left\{ \begin{cases} 1 & x_i > \bar{x} \\ -1 & x_i \leq \bar{x} \end{cases} \quad i = 1, 2, ..., n \right\}$$

with $\bar{x}$ the mean of $x(t)$ and x_i a value in $x(t)$. Each binary value was stored in sequence z_i and the total number of adjacent value changes was computed:

$$N_\Delta = \sum_{i=1}^{N-2} \frac{z_{i+1} - z_i}{2}$$

The fractal dimension D was found using the relation:

$$D = \frac{log_{10}^N}{log_{10}^N + log_{10}\left(\frac{N}{N+0.4N_\Delta}\right)}$$

8.3.3 *Hurst exponent:*

Time series $x(t)$ of length n was divided into a number of shorter series of lengths $n = n, n/2, n/4, ...$, and the average rescaled range was calculated for each value of n. To do this, first the mean of $x(t)$ was computed, denoted as $\bar{x}$. Then a separate mean-adjusted series was created and subsequently used to calculate the cumulative deviate series $Z(t)$:

$$Z(t) = \sum_{i=1}^{t} (x_i - \bar{x})$$

The range $R(n)$ was found by subtracting the maximum and minimum value from $Z(t)$ and the standard deviation $S(n)$ was found following:

$$R(n) = \max(z_1, z_2, ..., z_n) - \min(z_1, z_2, ..., z_n)$$

$$S(n) = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

By taking the average of the rescaled range $R(n)/S(n)$ over all partial series of length n, a straight line was fitted describing the power law behavior:

$$\frac{R(n)}{S(n)} \propto C(n)^H$$

with C a constant and n the time span of the observation. Hurst's exponent was estimated by plotting $\log[R(n)/S(n)]$ as a function of $\log n$. The slope parameter was estimated using a least-squares fit (lstsq numpy v1.22.1).

8.3.4 Largest Lyapunov exponent: The first step involves constructing an attractor dynamics from $x(t)$. The reconstructed trajectory X can be expressed as a matrix with each row a phase-space vector. The delayed sequence of $x(t)$ was constructed using the following criteria:

$$M = \{X(i), X(i+J), ..., X(i+(m-1)\cdot J)\}$$

with J the lag or reconstruction delay, and m the embedding dimension. Liebert and Schuster (1989) described a method for choosing the lag based on the correlation sum. Nonetheless, it remains an open challenge to determine the right lag parameter. J was computed using the FFT and the final parameters used were $J = 4$ and $m = 10$.

Following the reconstruction of the series dynamics, the nearest neighbors of each point on the trajectory were identified. The closest neighbor $X_{\hat{j}}$ was selected by searching for a point that minimizes distance from a specific reference point X_j:

$$d_j(0) = \min_{X_{\hat{j}}} \|X_j - X_{\hat{j}}\|$$

with $d_j(0)$ the initial distance from the j^{th} point to its nearest neighbor. $\|...\|$ denotes the Euclidean norm. As an additional step we required the nearest neighbors to be separated by a longer duration than the time series mean duration:

$$|j - \hat{j}| > \text{mean period}$$

Each pair of neighbors was considered as nearby for initial conditions for a variety of trajectories. Using the mean rate of separation between the nearest neighbors, the largest Lyapunov exponent was estimated. As far as the method for calculating λ_1 is concerned, according to Rosensteins' method, the largest Lyapunov exponent is defined as (Rosenstein et al., 1993):

$$d(t) = Ce^{\lambda_1 t}$$

with $d(t)$ the average divergence at time t and C a constant that normalizes the initial separation. We assumed that the j^{th} pair of neighbors diverge approximately at the rate by the largest Lyapunov exponent:

$$d_j(t) \propto C_j e^{\lambda_1 (i\Delta t)}$$

The equation the latter presents a set of approximately parallel lines $j = 1, 2, ..., N$, each with a slope roughly proportional to λ_1. The slope parameter (i.e., the largest Lyapunov exponent) was estimated from the average of all parallel lines using a least-squares fit (lstsq numpy v1.22.1).

8.3.5 Fisher information and entropy: First a set of embedding sequences was created. The delayed sequence of $x(t)$ was constructed as:

$$M = \{X(i), X(i+J), ..., X(i+(m-1)\cdot J)\}$$

with $J = 4$ and $m = 10$. Given the embedding sequences in M, we constructed a new matrix from the reduced singular values of the SVD transformation (svd numpy v1.22.1). Let's denote the output of the SVD transformation as:

$$M = U\Sigma V^t$$

with M the matrix to be decomposed, U the left singular matrix, Σ the diagonal matrix containing singular eigenvalues and V the right singular matrix. The diagonal eigenvalues were extracted and used to compute the entropy H and Fisher information I, with x_i the ith value of the diagonal eigenvalues.

$$H = -\sum_{i=1}^{N} x_i \cdot log_{10}(x_i)$$

$$I = \sum_{i=2}^{N} \frac{(x_i - x_{i-1})^2}{x_i}$$

8.3.6 Phase locking value: Consider a pair of time series $x_i(t)$ for $i = 1, 2$. Let the output of the Hilbert transform be denoted as the analytical signal $z_i(t)$:

$$z_i(t) = A_i(t) \cdot e^{-j\phi_i(t)}$$

The instantaneous phase $\phi_i(t)$ is the angle between the real and the imaginary parts of the Hilbert analytical signal. After the analytical signals were obtained (hilbert scipy v1.9.1), the relative phase was computed:

$$\Delta\phi(t) = arg\left(\frac{z_1(t)z_2(t)}{||z_1(t)|| \cdot ||z_2(t)||}\right)$$

with the argument of a complex number z, denoted $arg(z)$, the angle between the positive real axis and the line joining the origin. ||...|| denotes the Euclidean norm. A phase close to 1 describes synchronized signals, with 0 reflecting no phase synchrony.

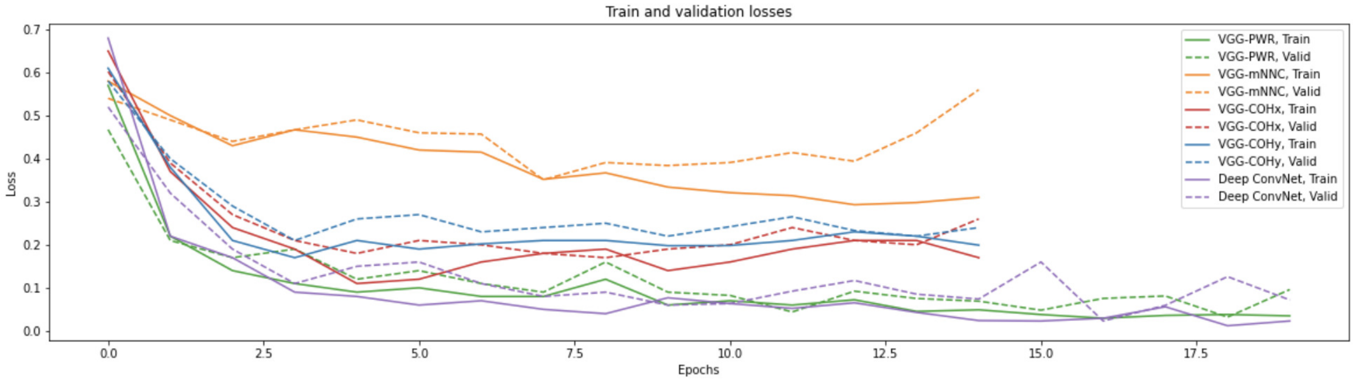

Fig. 7: Losses incurred during training and validation of the networks. Validation losses for the Deep ConvNet, VGG-PWR, VGG-COHx and VGG-COHy show convergence after approximately 10 epochs. The VGG-mNNC does not show proper convergence.

8.4 *Appendix D: Results of the deep classifiers*

Table 5 VGG-model hyperparameter optimization results.

Hyperparameters	VGG-PWR	VGG-mNNC	VGG-COHx	VGG-COHy
No. dense nodes	2000	2500	2000	2000
Batch size	5	5	5	5
Epochs	20	15	15	15
Learning rate	5e-5	10e-4	5e-4	5e-5
Dropout	0.09	0.12	0.05	0.08
Activation function	Softmax	Softmax	Softmax	Softmax
Optimizer	Adam	Adam	Adam	Adam
Loss	Categorical cross-entropy	Categorical cross-entropy	Categorical cross-entropy	Categorical cross-entropy