Detecting pathological scalp recordings using features of the EEG background pattern: A review

Rob van der Nagel, (October 30, 2022)
Faculty of Electrical Engineering, Mathematics and Computer Science
Department of Clinical Neurophysiology, Medisch Spectrum Twente, and Institute of Technical Medicine
Email: r.vandernagel@student.utwente.nl

***Abstract*—This paper reviewed the scientific literature on automated approaches to detect pathological electroencephalograms. We explored a novel idea regarding the use of features that describe specific properties of the EEG background pattern to enhance existing models. 81 papers were selected that solved a binary classification problem centered on EEG signals for specific and non-specific neurological outcomes. From the 81 papers, only two papers described aspects of the background activity. 9 papers that proposed specific methods for pathology decoding were extensively analysed and categorised into feature-based and end-to-end methods. Expert features such as the quantitative features proposed by Lodder and van Putten (2012) can be used directly in feature-based methods by adding them to an a priori selected set of handcrafted features. Furthermore, inspired by the visual output of the Colorful Brain proposed by van Putten (2007), we propose to use a triplet of features that quantifies the spatial distribution of the EEG signals and their coherence in existing feature-based models based on pre-trained pattern recognition models.**

***Index Terms*—EEG background pattern, pattern recognition, binary classification, deep learning, convolutional neural networks, feature importance visualization.**

I. INTRODUCTION

Electroencephalography (EEG) is a non-invasive and relatively affordable technique used to record the brain's electrical activity. Scientists and clinicians have long studied the electrical activity of the brain, and the understandings of the underlying mechanisms of EEG changes associated with cognition, behavioral states, motor function and neurological diseases have led to a wide range of practical applications (Brazier, 1961; Craik et al., 2019). Routine EEGs assist clinicians in diagnosing neurological diseases such as epilepsy, dementia and schizophrenia. Routine EEGs are essential in long-term studies of patients with already-diagnosed diseases, as the ability to analyse EEG recordings and study their physiological and pathological correlates provides useful insights into abnormal brain functioning.

Even though the gold standard for EEG interpretation is visual, the interest in automated analysis has become more prominent in clinical research. Significant progress has been made in automated seizure and spike detection using Convolutional Neural Networks (or ConvNets for short) (Acharaya et al., 2018; Cloostermans et al., 2018; Lourenco et al., 2020; Westover, 2020). Likewise, ConvNets have been proposed to detect abnormal electroencephalograms (Gemein et al., 2020; Shirrmeister et al. 2017). In fact, research on pathology decoding is still in its infancy as most studies primarily focused on validating ConvNets trained on publicly available routine scalp recordings (Obeid and Picone, 2013). Despite the wide variety of proposed ConvNets for pathology decoding, most studies use approaches without any focus for specific "expert" features. There is little attention in pathology decoding research to the background activity, but abnormalities in background characteristics can be excellent indicators of 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 the properties of the background activity using five key features. Putten (2007) earlier introduced to visualise the time-localised frequency information for a triplet of features that quantifies 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 enhance decoding performance of existing approaches with features that quantify specific characteristics of the background pattern.

The purpose of this review paper is to examine the scientific literature related to automated approaches for pathology decoding to see whether such features could be included in existing approaches. We propose to present a general overview of the field, review decoding strategies and categorize modelling approaches. With this result of this review paper, we propose to design reliable and reproducible experiments to evaluate whether the use of background features adds significant value to pathology decoding approaches.

The rest of this paper is structured as follows: Section II describes the planning, conducting and documenting phases of the research method, including search statistics and an overview of the reviewed work. Section III describes the results, including the categorisation of modeling approaches and features, limitations of the review and possible research improvements. Sections IV and V describe the discussion and conclusion, respectively.

II. Research Method

To structure the research, the well-established Systematic Literature Review (SLR) method was used (Kitchenham and Charters (2007)). Following these guidelines, our SLR was carried out in three stages: planning, conducting and documentation (see Figure 1).

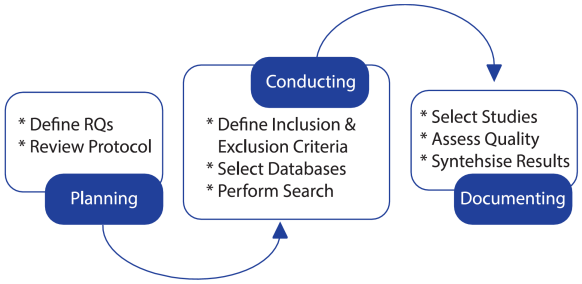

Figure 1. Process of the systematic literature review.

A. Planning

To achieve our research goal, we formulated the following research question:

RQ: *Which pathology decoding approaches do exist in the literature to include expert features in?*

We formulated queries composed of selected keywords and combinations with the AND/OR logical operators. The asterisk symbol "*" presents the wildcard operator. The query was focused on retrieving results based on the keywords present in the title, abstract and keywords list of the papers.

Query: (eeg pathology decoding) OR (eeg abnormal classification) AND (deep learning OR neural network* OR machine learning)

B. Conducting

Peer-reviewed papers published between January 2012 and January 2022, as well as electronic preprints, were chosen as the source material for this review paper. PubMed (https://www.ncbi.nlm.nih.gov/pubmed/), Scopus (https://www.scopus.com/), ScienceDirect (https://www.sciencedirect.com/) and arXiv (https://arxiv.org/) were searched to collect an initial list of papers containing specific search terms in the title or abstract.

Table 1. Review Protocol of RQ

Inclusion (I)	1) Studies that contain keywords eeg and (abnormal*, patholog*, classif*, detect*, decod*, network*, *learning). 2) Studies that contain keywords (abnormal*, patholog*).
Exclusion (E)	1) Studies that contain keywords (ecg, *emg, *mri, pcg). 2) Studies that contain keywords (multi-class, multi-label, real-time). 3) Studies that mention (seizure, bci, motor*, behavior).

Unlike peer-reviewed publications, non-peer reviewed papers, such as those posted on arXiv, serve as important sources of state-of-the-art information because their publication cycle is usually shorter. As a result, unconventional research ideas are more likely to be shared in such repositories, which enhances the diversity of the reviewed work and reduces the possibility of bias introduced by peer-reviewing. Therefore, non-peer reviewed papers were also included in this review paper. For PubMed, Scopus and ScienceDirect

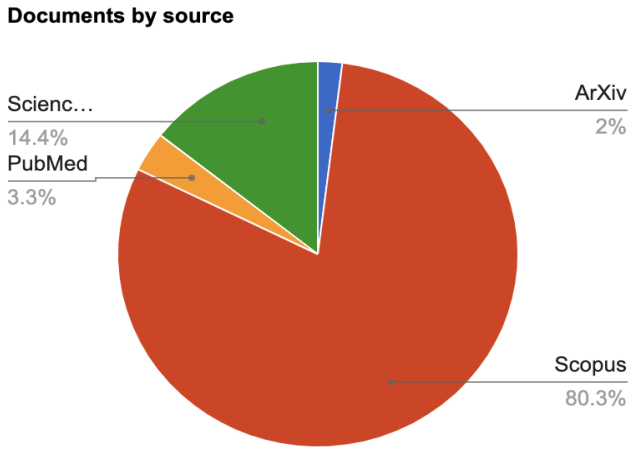

Figure 2. Phase 0: Retrieved papers (2484) after performing the initial search query.

the same exact **Query** as described in the Planning section was used. For arXiv, however, this query did not lead to any search results. Therefore, two searches were conducted: (eeg pathology decoding) and (eeg abnormal classification), whose results were combined. Also experiments with the query (eeg AND (deep learning OR machine learning OR convolutional neural network)) have been examined, but this lead to no papers or papers that were misaligned from the topic.

C. Documenting

Papers including their title and abstract were downloaded in textual form and analysed using Python 3.8. All search results

were transformed to a single dataframe, from which iterative queries were performed that match the criteria in Table 1. We performed our assessment in a systematic way so that we could evaluate the results for each query separately. After performing the initial search, a total of 2484 papers were collected. After performing queries $I1$ and $I2$, 503 papers were approved. Similarly, after performing queries $E1$, $E2$ and $E3$, 370 papers were approved.

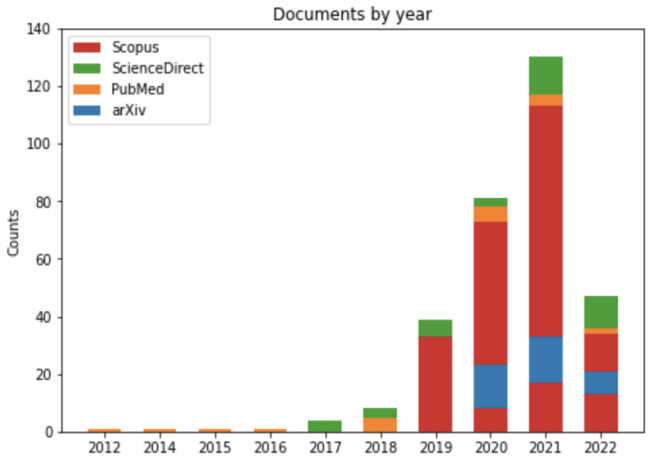

Figure 3. Phase 1: Approved papers (370) after performing the inclusion and exclusion criteria.

From this point, the eligibility of the selected papers was determined by reading their titles first. If the titles did not state clearly whether inclusion or exclusion criteria were met, the abstracts were read as well to review the overall narrative. We selected the papers in three search steps. In the first steps we filtered papers that did not describe any EEG classification/EEG pathology decoding task, and approved 195 papers. In the second step we read all abstracts, scanned their full text and included papers that focused on binary classification problems. This resulted in 81 approved papers. Still many papers did not specifically focus on a method for pathology decoding. As pathology decoding methods rely on solving a binary classification problem, many papers were approved that solved any type of binary classification problem (e.g., ictal/interictal, epileptic/non-epileptic, focal/non-focal, schizophrenia/non-schizophrenia, alcoholic/non-alcoholic, drowsy/non-drowsy, etc.)

To subdivide our search focus, we reviewed the 81 papers in the following way. We first analysed all modeling methodologies that aimed to classify any task related to neurological disorders (e.g. epilepsy, schizophrenia, depression, Parkinson's disease, Alzheimer's disease and general pathology). In the second part only those papers were selected that aimed to decode general pathology, and performed a more extensive analysis on 14 papers. Figure 4. illustrates the work reviewed. To guarantee search reproducibility, all scripts, search results and intermediate steps have been uploaded to Github: https://github.com/RobvanderNagel94/Pathology-Decoding.

III. Results

Pathology decoding approaches that aim to classify EEGs as abnormal vs. normal can broadly be categorized into *feature-based* and *end-to-end* methods. These methods are built from two separate parts: feature extraction and classification. In turn, for both parts different algorithms can be used. In the following sections we will detail on these specific parts for both feature-based and end-to-end methods.

A. Feature-based and end-to-end methods

Feature-based methods use a priori selected features to represent the raw EEG data. Features describing time, frequency and connectivity have all been used to characterize EEG signals (Gemein et al., 2020). Papers that adopted a feature-based method commonly used Fourier or wavelet analysis to find spectral power estimates. This was true for studies that focused on decoding outcomes for specific diseases such as schizophrenia and depression, but also for non-specific cases like alcohol and drowsiness detection (Anuragi, 2019; Albaqami et al., 2020; Yildiz et al., 2021). Also features that describe channel connectivity have been widely adopted in many specific and non-specific decoding tasks (Mumta et al., 2018; Deng et al., 2022; Movahed et al., 2022). Overall, the papers reviewed that adopted a feature-based method primarily relied on one or multiple features as input data before the classification part.

Feature-based methods use binary classifiers to discriminate between feature-representations. To be more specific, let M be a matrix of features with $M \in \mathbb{R}^{I \times F}$ with corresponding labels $L \in \{i_1 ... i_I\}$ for $i \in \{1, 0\}$. In the latter, I presents the number of recordings and F the number of feature values. A classifier was used to discriminate between feature representations M and corresponding labels L. Binary classifiers such as random forest (RF), support vector machine (SVM), logistic regression (LR), k-nearest neighbors (KNN) and the various automated scikit-learn classifiers (ASC) have been used to decode pathology (Lopez et al., 2017; Roy et al., 2018; Gemein et al., 2020). Binary classifiers have also been used in end-to-end methods, typically added as an embedded layer in the architecture (Amin et al., 2019; Van Leeuwen et al., 2019; Singh et al., 2021; Bajpai et al., 2021).

As opposed to feature-based methods, end-to-end methods use raw or minimally preprocessed data as input. The general idea is to use ConvNets that systematically applies learned filters to raw input data to create high-level feature maps that summarize the characteristics of the input. ConvNets allow for a joint optimization of the feature extraction and classification in one structure. The Softmax function was used as activation in the output layer of the network to predict a multinomial probability. With the flexibility of the network design, the final output can also be used for multi-label classification. One major advantage of end-to-end methods

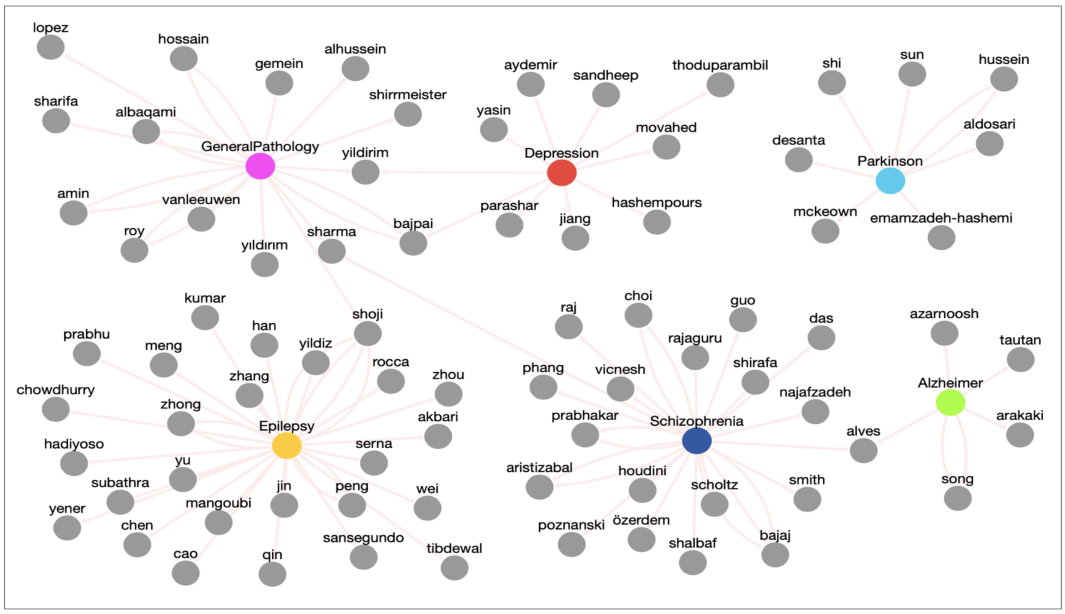

Figure 4. Phase 2: Approved papers (81) for the final review. The focus of each individual paper was assessed and categorised to one of six specific domains: epilepsy, schizophrenia, general pathology, depression, Parkinson's disease and Alzheimer's disease.

is that no additional preprocessing steps are required which makes them more practical for non-experts.

In general, feature-based and end-to-end methods differ in the feature extraction part. However, there are also different strategies for generating features and decoding EEGs. To clarify, we further separate feature-based methods in having *cropped-wise* or *image-based* decoding strategies.

B. Cropped-wise decoding

The works of Shirrmeister et al. (2017) and Gemein et al. (2020) have made tremendous progress in pathology decoding research, evaluating different feature-based and end-to-end methods on the TUH Abnormal Corpus (Obeid and Picone, 2013). They share the common factor of using multichannel inputs to represent the EEG data. In addition, they both use a cropped-wise strategy to either generate features from raw EEG data, or to feed batches of raw data to train an ConvNet (commonly referred to as cropped-wise decoding). For clarification, both methods follow a cropped-wise decoding strategy.

Gemein et al. (2020) extracted features for each crop in multiple 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. For each analysed crop of raw data, a matrix of features was computed. The features were saved in feature matrix M_i with $i \in I$. The final dimension of the feature matrix was $M_i \in \mathbb{R}^{C_i \times F}$ where C_i represents the amount of analysed crops, I the total number of recordings and F the dimension of the feature vector. Every crop's feature vector was considered an independent example for time-resolved (i.e., non-aggregated) strategy was performed. Consequently, this increased the amount of training examples for the training phase but it also increased memory usage. Dimensionality reduction on the time-resolved crops was proposed using Principal Component Analysis (PCA). However, the dimensionality reduction led to a decrease in decoding accuracy. To reduce the dimensionality of the time-crop features matrices, a median aggregated decoding strategy was proposed to obtain a single feature vector of length F for each recording. The shape of the final aggregated feature matrix was $M_{agg} \in \mathbb{R}^{I \times F}$. By aggregating features into one feature vector, the dimensionality of the feature matrix was reduced, allowing for faster learning at the expense of discarding channel information. However, decoding accuracy was only modestly affected by the decision to aggregate the matrices. Several binary classifiers, ConvNets and ConvNet-archetypes earlier proposed in the literature were used.

Shirrmeister et al. (2017) introduced the well established Deep ConvNet (BD-deep4) architecture which employs an initial separated convolution followed by several blocks of convolutions and max-pooling operations with exponential linear units serving as activation functions. In addition, they used the temporal convolutional network (TCN) architecture that was originally proposed by Bai et al. (2018) as an alternative to recurrent neural networks (RNNs). On several datasets that commonly serve as benchmarks for RNNs, the TCN architecture has consistently proved to outperform RNNs. Further, another ConvNet architecture named Shallow ConvNet (or BD-Shallow) also previously introduced by Shirrmeister et al. (2017) was used. As with the BD-Deep4 network, there is an initial separated convolution. However, this is the only convolution in the entire architecture. In addition, the BD-EEGNet network originally introduced by Lawhern et al. (2016) was used. As opposed to the BD-Deep4 and BD-

TCN, the BD-EEGNet architecture has only a small number of parameters it uses.

Van Leeuwen et al. (2019) applied the ConvNet structure adapted from the work of Shirrmeister et al. (2017) with minor changes and used a time-resolved decoding strategy on a different dataset. They explored with including sleep stage probabilities (for REM, NREM1, NREM2, NREM3) and contextual information such as age and gender. Using a recently published algorithm, they were able to assign sleep stages to consecutive epochs of 30 seconds of EEG at a level equivalent to human experts. Prior to the last classification layer, they added an extra average pooling layer to take an average along the time point axis, which enables them to add new features to the network before the classification layer. Two ways were used to incorporate these features. Subject age was first normalized and then added to the feature vector after the average pooling layer. The sleep stage probabilities were averaged over one minute samples to match the input size of the network and then added to the same feature vector before the final classification layer. Furthermore, they used a data augmentation technique by flipping the channels of the left and right hemispheres with probability 0.5 to avoid over-fitting to one hemisphere.

C. Image-based decoding

On the other hand, works of Lopez et al. (2017), Roy et al. (2018), Singh et al. (2021) and Bajpai et al. (2021) focused on single-channels inputs, mainly experimenting with specific or optimised channel pairs. They all share the common factor of using images to represent the EEG data. Most papers reported the creation of spectograms by estimating the spectral characteristics for multiple overlapping EEG segments using a window function. A pre-trained ConvNet, such as VGG-16, VGG-19, AlexNet, etc., was used for this specific application. Such ConvNets typically perform well on general image-classification tasks.

Prior work from Lopez et al. (2017) introduced features based on mel-frequency cepstral coefficients (MFCCs), a feature often used to describe raw sound waves in speech recognition. Bajpai et al. (2021) generated spectograms using short-time fourier transformations (STFT). They used the first minute from each recording on the T5-01 channel pair from a re-referenced TCP montage. Further, they experimented with DenseNet, Inception-Resnet-v2 and Seizurenet and removed the last classification layer from the initial network and replacing it with a SVM. Also Amin et al. (2019) and Singh et al. (2021) followed a similar approach in removing the last classification layer in a pre-trained model. It must be noted that general image-recognition models such as VGG-16, VGG-19, DenseNet, Seizurenet, Inception-ResNet-v2 and AlexNet are models that cannot be used directly for pathology decoding. In order to create a binary classification problem, they removed the last classification layer and replaced it with one or two more dense layers, or a conventional classifier.

Singh et al. (2021) generated various spectograms based on STFT, MFCC and Chronograms from channel pairs located in the occipital region. They used the VGG-19 network and experimented with different classifiers combinations such as SVM, LR and RF. Yıldırım et al. (2020) and Roy et al. (2018) experimented with one- and two-dimensional ConvNet architectures, respectively. They introduced modified versions of ConvNets with different archetypes (ConvNet and TCNN-RNN) based on one- and two-dimensional input data. With all models described, an end-to-end method was followed. The works of Sharma et al. (2020) and Sharifa

Table S1
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

Razavian A (2018) used a similar approach with one- and two dimensional ConvNets.

In summary, the decoding methods proposed for single- and multichannel inputs are clearly differentiated (see Table S1). Schirrmeister et al. (2017), van Leeuwen et al. (2019) and Gemein et al. (2020) adopted a cropped-wise decoding strategy using multichannel inputs to represent the EEG data. In contrast, Lopez et al. (2017), Amin et al. (2019), Bajpai et al. (2021), Yıldırım et al. (2020), Roy et al. (2018) and Singh et al. (2021) adopted either an image-based decoding strategy, using single-channel inputs to create images based either on optimised channel pairs or channel pairs positioned in the occipital region, or no specific strategy using raw single-channel data as input. These spectograms, in turn, were mostly used to train a pre-trained ConvNet that is able to learn specific feature representations from images. Transfer learning was used to adjust the classification layer of the network with an alternative classifier or additional dense layers.

D. Expert vs. handcrafted features

Many features have been proposed in the literature for feature-based pathology decoding methods. However, in this paper we made a distinction between expert and handcrafted features. For example, features that describe time, frequency and synchrony are used to describe general characteristics of raw EEG data and therefore widely used for many EEG classification tasks. Conversely, expert features use the same techniques to estimate spectral power but use domain knowledge to derive information in the form of quantitative features linked to specific brain functioning. Within the literature of pathology decoding, no specific focus has been dedicated to features that describe the background pattern. Although Gemein et al. (2020) proposed a wide range of handcrafted features to describe raw EEG data in specific narrow frequency bands, no expert features have been used. van Leeuwen et al. (2019) experimented with contextual factors such as age and sleep stage and showed that adding specific features to the learning process of the network increased decoding accuracy (although not proven statistically significant).

For all 81 reviewed papers, only two papers dedicated specific focus on features of the background pattern. Moghadam and Vanhatalo (2021) focused on the development of key components of a neonatal EEG background, starting from the visual background scoring to classifier design, and finally visualising their performances. They tested three classifier designs based on 98 computational features. However, the specific features used were not described in the paper. Peh (2021) proposed three automated approaches: a threshold-based detection system (TDS), a shallow learning-based detection system (SLDS) and a deep learning-based detection system (DLDS) to detect slow-wave activity in EEG recordings. Again, channel-level annotations were required to train the detector which must be manually assigned.

E. Research limitations

As part of our review on pathology decoding approaches, only 14 papers were analyzed specifically dedicated to pathology decoding for neurological outcomes. With these papers, only 9 papers used distinct approaches, and their modeling methodologies were systematically evaluated. Table S1 shows that the decoding performances for the different methods result in a range between 78-86% with one paper that reported accuracies >95%. Due to the variety of steps used in the described methods for pathology decoding, performances cannot be compared reliably based on these performance measures. A direct comparison between specificity and sensitivity levels makes less sense if models use different data strategies (e.g., single- vs. multichannel input, input length of data, etc.). In addition, specific information regarding the inputs, preprocessing, features and training strategies were not always clearly described (or sometimes not described at all). This might lead to bias in the generalizing of decoding approaches as we might have missed crucial parts of processing pipelines. Subtle changes in model design choices (e.g., use of filter techniques to reduce the effect of artifacts, the amount of training samples used to train the network, use of specific architectures and archetypes) lead to different performances. Therefore such factors should be similar when comparing model performances.

F. Research improvements

As well as building robust models, model interpretability, and particularly explainability, are crucial to identify real powerful decoding approaches. End-to-end models have certain advantages as they allow raw EEG data as input making them applicable for non-experts. To the best of our knowledge, the models proposed by Schirrmeister et al. (2017) and van Leeuwen et al. (2019) are among the most advanced end-to-end methods to exist. On the other hand, Gemein et al. (2020) demonstrated that feature-based models using multichannel input data performed somewhat similarly to these end-to-end models. The feature-based method also have inherent advantages, such as feature and classifier interpretability and explainability. In addition, the ability to experiment with computational features and binary classifiers is a powerful instrument in building robust and reliable decoding models. Furthermore, feature-based methods also have disadvantages, such as the need for raw EEG data to be transformed into computational features before the classifier can be used, and the dependence on high quality data (i.e., the level of noise in recordings caused by artifacts) to achieve good decoding performance. As a result of the use of a variety of algorithms, converting raw EEG data into useful features and classifying recordings can be quite complex.

Feature-based models that use pre-trained ConvNets may benefit from a technique commonly known as gradient class activation mapping (Grad-CAM) to visualize feature

importance. By visualizing the input regions highlighted by the model, this technique makes ConvNets more transparent. This technique creates a coarse localization mapping of important regions in an image by using the gradient information flowing into the final convolutional layer. Grad-CAM requires no re-training and is applicable to any ConvNet architecture. More interestingly, Grad-CAM may be combined with existing pixel-space visualizations to create a high-resolution and class-discriminative visualization (Guided Grad-CAM). Although these techniques have been proposed in general deep learning research, none of the papers reviewed used any visual feature visualization technique. Thus, this technique might help to understand the decisions made by the ConvNet, which could therefore increase overall feature explainability.

IV. Discussion

Although machine learning approaches have proven useful for many EEG applications (i.e., motor-imagery classification for brain-computer interface, emotion classification, seizure and IED detection, etc.), they have certain limitations in practice. Especially in the domain of clinical neurophysiology, not all pathology decoding methods proposed in the literature can be used in real-life applications. A common misconception about pathology decoding models is that the performance (e.g., sensitivity, specificity, and less important, accuracy) is the most important indicator of a model's ability to generalize to unseen recordings. However, this depends on many more factors. For instance, the three models of Amin et al. (2019) reported state-of-the-art performance compared to all other papers. Though not incorrect, the paper reported the most optimal channel pairs. In practice, such models are not able to cope with the variety of patient-specific EEG characteristics.

Even so, the brute-force approach of trying all channel pairs to find an optimal one does not scale well on large datasets. Furthermore, the capability of ConvNets to generalize to unseen samples highly depends on a good convergence of the train and test losses. In practice, hyperparameter tuning on large and diverse datasets is time-consuming and does not always result in stable converging behavior of train and test losses (less reliable models). By increasing the dataset with labeled recordings over time (to make the model more reliable), such models become dependent on constant parameter tuning with increasing complexity. Such problems could be resolved with advanced infrastructures or cloud-based solutions to optimize computing time but many medical care institutions do not accommodate these technologies. It is therefore necessary for the physician or specialist to tune hyperparameters manually to train pathology decoding models.

Moreover, the majority of work reviewed validated models on labeled recordings from the TUH Abnormal Corpus. As a benchmark dataset, this dataset is particularly useful but only one paper validated their approach on another dataset (Van Leeuwen et al., 2019). They used an even larger and more heterogeneous dataset, comprising over 8000 recordings. Further, in order to make reliable conclusions about the decoding performance, a thorough analysis on data characteristics (e.g., distribution of patient age, gender, neurological conditions, etc.) should be conducted.

V. Conclusion

The scientific literature on pathology decoding was reviewed in this paper, as well as the possibility of including expert features in existing approaches. Two types of pathology decoding approaches exist based on feature-based and end-to-end methods. A wide variety of handcrafted features were proposed to describe general characteristics of the EEG signals. Also various binary classifier designs have been used to discriminate between feature representations. The extraction of features from raw EEG data involves many approaches. A wide variety of ConvNet architectures for specific (BD-Deep4, BD-TCN, BD-EEGNet, BD-Shallow) and non-specific (VGG-16, VGG-19, AlexNet, SeizureNet, InceptionNet, DenseNet) EEG classifications have been used. In addition, different classifier combinations including archetypes were also examined.

Expert features such as the quantitative features proposed by Lodder and van Putten (2012) can be used directly in feature-based methods by adding them to an a priori selected set of handcrafted features. Another option is to embed the features in the network structure as proposed by van Leeuwen et al. (2019). Feature-based methods allow for direct comparison between handcrafted and expert features, making evaluation of decoding performance intuitive. However, the expert features cannot be directly applied to feature-based methods adopting an image-based decoding strategy. Unlike cropped-wise decoding, image-based decoding uses images instead of raw signal data. By constructing spectrograms from single-channel data centered in the occipital region, the spectral characteristics of the PDR were highlighted. From this point of view, highlighting other spectral characteristics related to the background pattern could enhance decoding performance as these characteristics might capture other information about general brain dysfunction. The expert features can be computed for overlapping EEG segments to create a 3-dimensional image where the "amplitude" of the time-frequency information is visualized using a color map. Inspired by the visual output of the Colorful Brain proposed by van Putten (2007), we propose a triplet of features that quantifies the spatial distribution of the EEG signals and their coherence.

VI. Reference links

Agius Anastasi A (2017) Acharya et al. (2018); Agius Anastasi A (2017) Brazier (1961) Craik et al. (2019) Schomer DL (2012) Prabhu D. (2022) György Buzsáki (2012) Aich (2014) Spence and Schneider (2009) M. G. Knyazeva (2013) Yi Jin 1 (2006) Ishii et al. (2010) Spronk et al. (2011) Roberta Lizio and Babiloni (2011) Niedermeyer (1997) Lodder and van Putten (2012) Grant (2014) Houfek (1959) Tjepkema-Cloostermans et al. (2018) da Silva Lourenço et al. (2019) Schirrmeister et al. (2017) Gemein et al. (2020) Putten

(2004) van Putten and Tavy (2004) Agius Anastasi A (2017) Kitchenham and Charters (2007) Lopez et al. (2017) Bajpai et al. (2021) Singh et al. (2021) Amin et al. (2019) Roy et al. (2018) Krizhevsky et al. (2012) Shaojie Bai (2018) Lawhern et al. (2016) Yıldırım et al. (2020) van Leeuwen et al. (2019) Umar Asif and Harrer (2020) van Putten (2007) Hezam Albaqami (2020) Abdulnasir Yildiz (2021) Arti Anuragi (2021) Mumtaz W (2018) Song Z (2022) Reza Akbari Movahed (2021) Moghadam and Vanhatalo (2021) Peh (2021) Sharma et al. (2020)

References

Sherif Said Abdulnasir Yildiz, Hasan Zan. Classification and analysis of epileptic eeg recordings using convolutional neural network and class activation mapping. *Biomedical Signal Processing and Control*, 68, 2021. doi: https://doi.org/10.1016/j.bspc.2021.102720.

U.R. Acharya, S.L. Oh, Y. Hagiwara, J.H. Tan, and H. Adeli. Deep convolutional neural network for the automated detection and diagnosis of seizure using eeg signals. *Computers in Biology and Medicine*, 100:270–278, 2018. doi: 10.1016/j.compbiomed.2017.09.017. cited By 723.

Camilleri K Vella M Muscat R. Agius Anastasi A, Falzon O. Brain symmetry index in healthy and stroke patients for assessment and prognosis. *Stroke Res Treat*, 2017. doi: https://doi.org/10.1155/2017/8276136.

Tapas Kumar Aich. Absent posterior alpha rhythm: An indirect indicator of seizure disorder? *Indian J Psychiatry*, page 61–66, 03 2014. doi: 10.4103/0019-5545.124715.

Syed Amin, M. Shamim Hossain, Ghulam Muhammad, Musaed Alhussein, and Abdur Rahman. Cognitive smart healthcare for pathology detection and monitoring. *IEEE Access*, PP:1–1, 01 2019. doi: 10.1109/ACCESS.2019.2891390.

Dilip Singh Sisodia Arti Anuragi. Alcohol use disorder detection using eeg signal features and flexible analytical wavelet transform. *Biomedical Signal Processing and Control*, 52:384–393, 2021. doi: https://doi.org/10.1016/j.bspc.2018.10.017.

R. Bajpai, R. Yuvaraj, and A.A. Prince. Automated eeg pathology detection based on different convolutional neural network models: Deep learning approach. *Computers in Biology and Medicine*, 133, 2021.

M. A. B Brazier. A history of the electrical activity of the brain: The first half-century. *Macmillan*, 1961.

A. Craik, Y. He, and J.L. Contreras-Vidal. Deep learning for electroencephalogram (eeg) classification tasks: A review. *Journal of Neural Engineering*, 16(3), 2019. doi: 10.1088/1741-2552/ab0ab5.

Catarina da Silva Lourenço, Marleen Tjepkema-Cloostermans, Teixeira Luís, and Michel van Putten. Deep learning for interictal epileptiform discharge detection from scalp eeg recordings. pages 1984–1997, 09 2019. doi: 10.1007/978-3-030-31635-8_237.

Lukas Gemein, Robin Schirrmeister, Patryk Chrabąszcz, Daniel Wilson, Joschka Boedecker, Andreas Schulze-Bonhage, Frank Hutter, and Tonio Ball. Machine-learning-based diagnostics of eeg pathology. *NeuroImage*, 220:117021, 06 2020. doi: 10.1016/j.neuroimage.2020.117021.

Abdel-Baki S.G. Weedon J. Arnedo V. Chari G. Koziorynska E. Lushbough C. Maus D. McSween T. Mortati K.A. et al. Grant, A.C. Eeg interpretation reliability and interpreter confidence: a large single-center study. *Epilepsy Behav.*, 32:102–107, 2014.

Costas A. Anastassiou Christof Koch György Buzsáki. The origin of extracellular fields and currents — eeg, ecog, lfp and spikes. *Nature Reviews Neuroscience*, 13:407–420, 05 2012.

Abdulhamit Subasi Amitava Datta Hezam Albaqami, Ghulam Mubashar Hassan. Automatic detection of abnormal eeg signals using wavelet feature extraction and gradient boosting decision tree. *Biomedical Signal Processing and Control*, 70, 2020.

Ellingson R.J. Houfek, E.E. On the reliability of clinical eeg interpretation. *J. Nerv. Ment. Dis.*, 128:425–437, 1959. doi: 10.1007/s00702-013-1070-5.

Ryouhei Ishii, Ryu Kurimoto, Koji Ikezawa, Yasunori Aoki, Michiyo Azechi, Hidetoshi Takahashi, Takayuki Nakahachi, Masao Iwase, Hiroaki Kazui, and Masatoshi Takeda. Frontal shift of posterior alpha activity is correlated with cognitive impairment in early alzheimer's disease: A magnetoencephalography–beamformer study. *Psychogeriatrics: the official journal of the Japanese Psychogeriatric Society*, 10:138–43, 09 2010. doi: 10.1111/j.1479-8301.2010.00326.x.

Kitchenham and Charters. Guidelines for performing systematic literature reviews in software engineering. *Keele University*, Version 2.3, 7 2007. URL https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.117.471&rep=rep1&type=pdf.

A. Krizhevsky, I. Sutskever, and G.E. Hinton. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 2:1097–1105, 2012.

Vernon Lawhern, Amelia Solon, Nicholas Waytowich, Stephen Gordon, Chou Hung, and Brent Lance. Eegnet: A compact convolutional network for eeg-based brain-computer interfaces. *Journal of Neural Engineering*, 15, 11 2016. doi: 10.1088/1741-2552/aace8c.

Shaun Lodder and Michel van Putten. Quantification of the adult eeg background pattern. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, 124, 08 2012. doi: 10.1016/j.clinph.2012.07.007.

Sebas Lopez, G Suarez, D Jungreis, I Obeid, and Joseph Picone. Automated identification of abnormal adult eegs. *IEEE Signal Processing in Medicine and Biology Symposium.*, 2015, 12 2017.

R. Meuli M. Hasler O. De Feo K. Q. Do M. G. Knyazeva, M. Jalili. Alpha rhythm and hypofrontality in schizophrenia. *Department of Neurology*, 08 2013.

Pinchefsky Tse I. Viviana Marchi Jukka Kohonen Minna Kauppila Manu Airaksinen Karoliina Tapani Päivi

Nevalainen Cecil Hahn Emily W. Y. Tam Nathan J. Stevenson Moghadam, S. and Sampsa Vanhatalo. Building an open source classifier for the neonatal eeg background: A systematic feature-based approach from expert scoring to clinical visualization. *Front Hum Neurosci*, 2021. doi: 10.3389/fnhum.2021.675154.

Kamel N Ali SSA Malik AS Mumtaz W, Saad MNBM. An eeg-based functional connectivity measure for automatic detection of alcohol use disorder. *Artif Intell Med.*, 2018. doi: 10.1016/j.artmed.2017.11.002.

Niedermeyer. Alpha rhythms as physiological and abnormal phenomena. *Department of Neurology, John Hopkins University School of Medicine and Hospital, Baltimore, MD 21287-7247, USA.*, 1997. doi: 10.1016/s0167-8760(97)00754-x.

Bagheri E Chaudhari R Karia S Rathakrishnan R Saini V Shah N Srivastava R Tan YL Dauwels J. Peh, Thomas J. Multi-center validation study of automated classification of pathological slowing in adult scalp electroencephalograms via frequency features. *Int J Neural Syst.*, 2021. doi: 10.1142/S0129065721500167.

Najib Murr Prabhu D., Emmady. Eeg triphasic waves. *StatPearls*, 01 2022.

M. Van Putten. The revised brain symmetry index. *Clin Neurophysiol*, 118(11):2362–7, 2004. doi: https://doi.org/10.1016/j.clinph.2007.07.019.

Shima Shahyad Gholam Hossein Meftahi. Reza Akbari Movahed, Gila Pirzad Jahromi. A major depressive disorder classification framework based on eeg signals using statistical, spectral, wavelet, functional connectivity, and nonlinear analysis. *Journal of Neuroscience Methods,*, 358, 2021. doi: https://doi.org/10.1016/j.jneumeth.2021.109209.

Giovanni B. Frisoni Raffaele Ferri Guido Rodriguez Roberta Lizio, Fabrizio Vecchio and Claudio Babiloni. Electroencephalographic rhythms in alzheimer's disease. *Int J Alzheimers Dis*, 05 2011. doi: 10.4061/2011/927573.

S. Roy, I. Kiral-Kornek, and S. Harrer. Deep learning enabled automatic abnormal eeg identification. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2018-July:2756–2759, 2018. cited By 23.

Robin Schirrmeister, L. Gemein, K. Eggensperger, Frank Hutter, and Tonio Ball. Deep learning with convolutional neural networks for decoding and visualization of eeg pathology. pages 1–7, 12 2017. doi: 10.1109/SPMB.2017.8257015.

Niedermeyer's Schomer DL, Da Silva FL. Electroencephalography: basic principles, clinical applications, and related fields. *Lippincott Williams Wilkins*, 2012.

Vladlen Koltun Shaojie Bai, J. Zico Kolter. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. v2, 2018.

Sullivan J Carlsson SCNN. Sharifa Razavian A, Azizpour H. Features off-the-shelf : an astounding baseline for recognition. in: Proc: Ieee conf on comp vis pattern recognit workshops (cvpr). 20:806–13, 2018.

Manish Sharma, Sohamkumar Patel, and U Rajendra Acharya. Automated detection of abnormal eeg signals using localized wavelet filter banks. *Pattern Recognition Letters*, 133, 03 2020. doi: 10.1016/j.patrec.2020.03.009.

Rishav Singh, Tanveer Ahmed, Amit Kumar Singh, Prasenjit Chanak, and Sanjay Kumar Singh. Seizsclas: An efficient and secure internet-of-things-based eeg classifier. *IEEE Internet of Things Journal*, 8(8):6214–6221, 2021. doi: 10.1109/JIOT.2020.3030821.

Wang J Yi G. Song Z, Deng B. An eeg-based systematic explainable detection framework for probing and localizing abnormal patterns in alzheimer's disease. *J Neural Eng.*, 2022. doi: 10.1088/1741-2552/ac697d.

Sarah Spence and Mark Schneider. The role of epilepsy and epileptiform eegs in autism spectrum disorders. *Pediatric research*, 65:599–606, 07 2009. doi: 10.1203/PDR.0b013e31819e7168.

Desirée Spronk, Martijn Arns, K.J. Barnett, Nicholas Cooper, and Evian Gordon. 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*, 128:41–8, 01 2011. doi: 10.1016/j.jad.2010.06.021.

M.C. Tjepkema-Cloostermans, R.C.V. de Carvalho, and M.J.A.M. van Putten. Deep learning for detection of focal epileptiform discharges from scalp eeg recordings. *Clinical Neurophysiology*, 129(10):2191–2196, 2018. doi: 10.1016/j.clinph.2018.06.024.

Jianbin Tang Umar Asif, Subhrajit Roy and Stefan Harrer. Seizurenet: Multi-spectral deep feature learning for seizure type classification. *IBM Research Australia*, 2020.

K.G. van Leeuwen, H. Sun, M. Tabaeizadeh, A.F. Struck, M.J.A.M. van Putten, and M.B. Westover. Detecting abnormal electroencephalograms using deep convolutional networks. *Clinical Neurophysiology*, 130(1):77–84, 2019.

Michel J.A.M. van Putten. The colorful brain: Compact visualisition of routine eeg recordings. *Department of Clinical Neurophysiology, Medisch Spectrum Twente, and Institute of Technical Medicine*, pages 1–2, 2007.

M.J.A.M. van Putten and D.L.J. Tavy. Continuous quantitative eeg monitoring in hemispheric stroke patients using the brain symmetry index. *Keele University*, 35, 2004.

Aaron S Kemp Steven T Huerta Gustavo Alva Trung Minh Thai Danilo Carreon William E Bunney Jr Yi Jin 1, Steven G Potkin. Therapeutic effects of individualized alpha frequency transcranial magnetic stimulation (alphatms) on the negative symptoms of schizophrenia. 07 2006. doi: 10.1093/schbul/sbj020.

Ö. Yıldırım, U.B. Baloglu, and U.R. Acharya. A deep convolutional neural network model for automated identification of abnormal eeg signals. *Neural Computing and Applications*, 32(20), 2020.