Machine Learning for detection of Interictal Epileptiform Discharges

Catarina da Silva Lourenço*a,*, Marleen C. Tjepkema-Cloostermansa,b, Michel J. A. M. van Putteena,b

*aDepartment of Clinical Neurophysiology, Institute for Technical Medicine, University of Twente, Technical Medical Centre, Enschede, The Netherlands
bNeurocentrum, Medisch Spectrum Twente MST, Enschede, The Netherlands

Abstract

The electroencephalogram (EEG) is a fundamental tool in the diagnosis and classification of epilepsy. In particular, Interictal Epileptiform Discharges (IEDs) reflect an increased likelihood of seizures and are routinely assessed by visual analysis of the EEG. Visual assessment is, however, time consuming and prone to subjectivity, leading to a high misdiagnosis rate and motivating the development of automated approaches. Research towards automating IED detection started 45 years ago. Approaches range from mimetic methods to deep learning techniques. We review different approaches to IED detection, discussing their performance and limitations. Traditional machine learning and deep learning methods have yielded the best results so far and their application in the field is still growing. Standardization of datasets and outcome measures is necessary to compare models more objectively and decide which should be implemented in a clinical setting.

Keywords: Electroencephalogram, Interictal Epileptiform Discharges, Automated Detection, Machine Learning, Deep Learning, Convolutional Neural Networks.

* Corresponding author.
Email addresses: c.dasilvalourenco@utwente.nl (Catarina da Silva Lourenço), M.Tjepkema-Cloostermans@mst.nl (Marleen C. Tjepkema-Cloostermans), mj.a.m.vanputten@utwente.nl (Michel J. A. M. van Putten)
Highlights

- Methods for automated interictal epileptiform discharge (IED) detection have evolved in the past 45 years.
- The distinct outcome measures and datasets complicate the comparison of approaches.
- Deep learning seems to surpass other approaches in automated IED detection.
1. Introduction

Electroencephalograms (EEGs) are widely used in the diagnosis and classification of epilepsy, a brain disease that entails a predisposition for epileptic seizures (Fisher et al., 2014; Lopes da Silva et al., 2003; Tatum et al., 2018). However, most seizures occur sporadically and at unknown times, so the availability of ictal EEGs is scarce. Alternatively, interictal EEGs are used in diagnosis. Half of the routine EEG recordings from epilepsy patients include interictal epileptiform discharges (IEDs), with this number rising to 80% in sleep recordings (de Curtis et al., 2012; Tatum et al., 2018).

IEDs are epileptiform transients that indicate an increased likelihood of seizures (de Curtis et al., 2012), appearing mainly as spikes and sharp waves. Several examples of IEDs are shown in Figure 1.

Figure 1

The gold standard for IED detection is visual analysis by experts (Lodder et al., 2014). This approach has several drawbacks, including a long learning curve and extensive analysis time, especially for long recordings. Further, human error, subjectivity, intra and inter-observer variability result in misdiagnosis rates up to 30%, leading to lack of treatment or prescription of medication with potentially harmful side effects (Lodder et al., 2014; Nowack, 1997).

This motivates the development of computer assisted IED detection with algorithms that match or outperform experts, reducing the time and resources spent on visual analysis, as well as the misdiagnosis rate. Automating IED detection is not trivial due to the oversimplified textbook definitions of these transients, similarities between IEDs and other transients and variations in discharge morphology between patients (Wilson and Emerson, 2002).

Still, extensive research has been carried out aiming to automate the detection of IEDs. We review the evolution in the type of algorithms in the literature, their performance and
limitations. We used the search terms 'automated interictal epileptiform discharge detection' and 'automated interictal spike detection' in Scopus. Papers written in English, based on human scalp EEG and describing an algorithm aimed to detect IEDs, dating back to the beginning of this field (1976) and up to March 2020, were selected.

2. EEG Data

Several EEG datasets including samples with IEDs have been used for the development and evaluation of automated IED detection. Some datasets are unequivocally too small to encompass the diversity of IED shapes necessary for a robust, generalizable algorithm. Examples of this are approaches based on data from one patient (Latka et al., 2003; Tzallas et al., 2006b).

Although most authors used private datasets, impeding direct comparison of results across publications, some works were based on the same publicly available EEGs (or a subset thereof) (Abibullaev et al., 2010; Güler and Ubeyli, 2005, 2007; Güler et al., 2005; Guo et al., 2011; Iscan et al., 2011; Martinez-Vargas et al., 2011; Orhan et al., 2011; Ubeyli, 2008; Übeyli, 2009a, 2009b, 2010; Wang et al., 2011). The five-class dataset, made available by Andrzejak et al. (Andrzejak et al., 2001), consists of 100 single channel EEG segments without discernible artefacts. Classes A and B (extracranial) were recorded from 5 controls with eyes open and closed; classes C and D (intracranial) were recorded from 5 epilepsy patients in seizure free intervals in the epileptogenetic zone (D) and hippocampal formation of the opposite side of the brain (C); class E contained seizure activity from those patients. While this dataset is useful for comparison, it still entails limitations such as the small number of patients and the absence of extracranial recordings from epilepsy patients.

Another relevant characteristic concerning the datasets used for training and testing the algorithms is the expert opinion used for labeling the IEDs. Since identifying these discharges
is a subjective task prone to variability, it is crucial to have enough experts agreeing on the presence of an IED for it to be a part of the dataset. Inter-rater agreement (i.e. how often experts agree with one another) can be estimated by calculating the percentage of agreement or with more sophisticated metrics such as Cohen’s kappa (Cohen, 1960).

The number of experts annotating the EEGs in the reviewed studies ranges from one (Chaibi et al., 2015; Dingle et al., 1993; El-Gohary et al., 2008; Gabor and Seyal, 1992; Horak et al., 2015; Ko and Chung, 2000; Sankar and Natour, 1992; Webber et al., 1994) to eleven (Halford et al., 2013). While the opinion of one expert is highly relevant, a single subjective classification might not be enough for the correct identification of all the IEDs in a recording. Thus, most authors opt for including the opinion of two or three experts.

The way in which these opinions are taken into account is also pertinent. One way of doing this is to ‘accept’ the IED only if all experts agree on its presence (i.e. if consensus is reached) (Gotman and Wang, 1991, 1992; Inan and Kuntalp, 2007; Kurth et al., 2000; Kutlu et al., 2009; Pietilä et al., 1994). This eliminates doubtful patterns but can disregard some IEDs, classifying them as normal samples in the dataset. Instead of consensus, an agreement threshold can also be used – for instance, consider all IEDs identified by at least 7 of 8 experts (Guedes de Oliveira et al., 1983). Assigning weights according to the expertise of each clinician (Hostetler et al., 1992) or calculating the mean of the probability assigned to each sample by each expert (Wilson et al., 1999) are other ways or integrating multiple opinions.

3. Performance Metrics

To assess model performance, the number of true and false positives and negatives (i.e. correctly and incorrectly identified IEDs and non-IEDs) are essential characteristics. The number of true and false positives is particularly relevant since they show how many IEDs are
detected, as well as the number of false detections, which should be low in comparison to the correctly detected IEDs.

Several metrics can then be derived, including sensitivity (defined as the number of true positives over true positives and false negatives), specificity (true negatives over true negatives and false positives) and accuracy (true positives and true negatives over all samples).

An IED detector should have both a high sensitivity (to detect as many IEDs as possible) and high specificity (correctly identifying non-IEDs). Either will not suffice, as low sensitivity and high specificity yield a very low number of detections while a high sensitivity with low specificity would lead to a high number of false positives. An alternative metric to specificity is the false positive rate, either defined as 1-specificity or as the number of false positives per hour.

Accuracy is often reported, too, defined as sensitivity * P(a) + specificity * (1-P(a)) with P(a) the prior likelihood. Note that this metric alone does not provide enough information about the classifier performance if sensitivity and specificity are not reported as well, as the accuracy generally depends on the prior likelihood of an IED, P(a). The only exception occurs in the unlikely scenario that the sensitivity equals the specificity, as in that case accuracy = sensitivity = specificity. Further, in the case of unbalanced datasets (which is typically the case in IED detection, as non-IEDs are more abundant), high accuracy often corresponds to a high number of true negatives, providing little to no information on IED detection itself. Thus, while accuracy is reported in many studies (Benlamri et al., 1997; Güler and Übeyli, 2005; Güler et al., 2005; Nigam and Graupe, 2004; Srinivasan et al., 2005; Tzallas et al., 2006a; Übeyli, 2008; Übeyli, 2008, 2009a, 2009b; Wang et al., 2011) (often as the only metric) it is not suitable for performance assessment nor comparison of different methods as typically the prior likelihoods of IEDs are different for the EEG data sets used between various studies.
4. Developed Methods

Over the last 40 years, several approaches to automate IED detection have been developed. Table 1 provides a comprehensive overview of the automated IED detection methods developed between 1976 and 2020. Supplementary Table S1 provides detailed information on each study.

All approaches can be broadly divided into four categories. In the first group (I), users select and extract characteristics of the EEG which are used to determine the probability of a sample including an IED. Examples include maximum spectral frequency, amplitude of a transient, similarity to a template, where classification is based on passing a particular user-defined threshold. Methods in the second category (II) use explicitly defined features as in the first group but classifiers ‘learn’, using statistical modeling, from the input, without the use of a pre-defined classification rule.

Expert systems (III) in IED detection are applied to the output of an algorithm from the first two categories. These methods cannot be used as a standalone approach, as they do not use raw data or features as input, but they can refine the classification of the previous algorithms.

Deep learning (IV) is a standalone approach that aggregates the feature extraction and classification steps, eliminating the need for a priori feature definition. These techniques use raw data as input and find its own representation of the feature space, which is then used to classify samples, constituting an end-to-end method.

Table 1 - Approaches to automated IED detection. More details are shown in Supplementary Table S1.

| Approach                  | Reference                                                                 |
|---------------------------|---------------------------------------------------------------------------|
| Mimetic Methods (I)       | (Gabor and Seyal, 1992; Gotman, 1982; Gotman and Gloor, 1976; Gotman and Wang, 1991; Guedes de Oliveira et al., 1983; |
| Feature/Classifier Type                                      | References                                                                 |
|-------------------------------------------------------------|-----------------------------------------------------------------------------|
| Thresholding of Frequency Features (I)                      | (Chaibi et al., 2015; Goelz et al., 2000; Indiradevi et al., 2008; Latka et al., 2003) |
| Morphological Filtering (I)                                | (Juozapavicius et al., 2011; Xu et al., 2007; Xu et al., 2006)              |
| Template Matching (I)                                       | (El-Gohary et al., 2008; Horak et al., 2015; Ji et al., 2011a; Ji et al., 2011b; Jing et al., 2016; Lodder et al., 2013; Lodder and van Putten, 2014; Nonclercq et al., 2012; Pietilä et al., 1994; Sankar and Natour, 1992; Thomas et al., 2017; Vijayalakshmi and Abhishek, 2010) |
| Traditional Machine Learning Classifiers (II)               | (Acir and Güzeliş, 2004; Acir et al., 2005; Bagheri et al., 2018; Chavakula et al., 2013; Derya Ubeyli, 2008; Güler and Ubeyli, 2007; Guo et al., 2011; Horak et al., 2015; Inan and Kuntalp, 2007; Iscan et al., 2011; Kelleher et al., 2010; Lima et al., 2010; Martinez-Vargas et al., 2011; Nonclercq et al., 2012; Orhan et al., 2011; Thomas et al., 2018; Thomas et al., 2017; Wang et al., 2011; Zhou et al., 2013) |
| Small Artificial Neural Networks (II)                       | (Abibullaev et al., 2010; Artameeyanant et al., 2012; Boos et al., 2010; Castellaro et al., 2002; Derya Ubeyli, 2008; Feucht et al., 1997; Gabor and Seyal, 1992; Güler and Ubeyli, 2005, 2007; Güler et al., 2005; Guo et al., 2009; Halford et al., 2013; James et al., 1999; Johansen et al., 2016; Kalayci and Ozdamar, 1995; Ko and Chung, 2000; Kurth et al., 2000; Kutlu et al., 2009; Liu et al., 2002; Nigam and Graupe, 2004; Orhan et al., 2011; Ozdamar and Kalayci, 1998; Pang et al., 2003; Park et al., 1998; Sezer et al., 2012; Song and Zhang, 2013; Srinivasan et al., 2005; Suresh and Balasubramanyam, 2013; Tarassenko et al., 1998; Übeyli, 2009a, 2009b, 2010; Webber et al., 1994; Wilson et al., 1999) |
| Expert Systems (III)                                        | (Argoud et al., 2006; Benlamri et al., 1997; Black et al., 2000; Dingle et al., 1993; Park et al., 1998; Tzallas et al., 2006a; Wang et al., 2011) |
| Deep Learning (IV)                                          | (Jing et al., 2020; Johansen et al., 2016; Lourenço et al., 2019; Tjepkema-Cloostermans et al., 2018) |

4.1. Pre-defined Features and Classifier (I)
Several types of EEG characteristics and classification strategies have been used for IED detection throughout the years. Early methods, such as mimetic approaches (section 4.1.1) classified a sample as having an IED if pre-chosen morphological (i.e. shape related) features were above a certain value. The same principle was applied to features in the frequency domain (section 4.1.2). Other approaches such as morphological filtering (section 4.1.3) and template matching (section 4.1.4) will also be discussed.

### 4.1.1. Mimetic Methods

Mimetic methods were the first approach to automated IED detection. These techniques aim to emulate the visual analysis process and thus resort mostly to morphological characteristics of the EEG, since these are the ones used by experts when searching for IEDs (Gotman, 1982; Gotman and Gloor, 1976; Gotman and Wang, 1991, 1992; Guedes de Oliveira et al., 1983; Hostetler et al., 1992).

The transient is divided into segments (i.e. the section between two extrema of amplitude), as shown in Figure 2. A minimum difference between segments is often imposed to reduce the impact of noise (third panel, Figure 2). Waves are defined as the set of two adjacent segments in opposite directions, with half-waves being the segments on either side of the maximum amplitude. From the half-waves, features such as relative amplitude, duration, slope and sharpness are extracted (Benlamri et al., 1997; Gotman, 1982; Gotman and Gloor, 1976; Guedes de Oliveira et al., 1983; Nigam and Graupe, 2004; Srinivasan et al., 2005; Tzallas et al., 2006a).

After extracting these features, physiologically acceptable values are used as threshold to decide whether a sample contains an IED or not. These are typically the same thresholds used when performing visual analysis.

**Figure 2**
The first published method was based on amplitude, duration and 'sharpness' of the half-waves (Gotman and Gloor, 1976). This algorithm was later updated to detect IEDs in different states (active wakefulness, quiet wakefulness, desynchronized EEG, phasic EEG and slow EEG), increasing generalization (Gotman and Wang, 1991, 1992). Another mimetic algorithm was partially implemented in hardware to support on-line detection, which was relevant progress at the date of publication (1980s) (Guedes de Oliveira et al., 1983).

In most of these studies, separation of training and test sets was not properly reported. Furthermore, sensitivity and specificity were not the chosen metrics for performance assessment.

Mimetic methods appear to have limited ability to distinguish IEDs from transients and artefacts with similar morphology, leading to a high number of false positives. While some form of artefact rejection was used in several studies (Gotman and Gloor, 1976; Gotman and Wang, 1991; Guedes de Oliveira et al., 1983), non-artefactual phenomena (i.e. physiological transients) still led to misclassifications. One study reported an average of 41 true and 39 false detections in 6 h of recordings, showcasing the high proportion of false positives (Gotman, 1982). Therefore, this type of approach lost popularity and gave way to other types of methods.

4.1.2. Thresholding of Frequency Features

Towards the end of the 1990s, frequency-based characteristics started being used for IED detection (Chaibi et al., 2015; Goelz et al., 2000; Indiradevi et al., 2008; Latka et al., 2003). Similar to mimetic methods, a threshold value was applied to these features to decide whether a sample contained an IED or not.

Features such as spectral entropy and power were used in several approaches (Iscan et al., 2011; Martinez-Vargas et al., 2011; Srinivasan et al., 2005; Tarassenko et al., 1998), as well other characteristics derived from Hilbert and Walsh transforms (Adjouadi et al., 2004; Feucht et al.,
Most studies used features derived from discrete or continuous wavelet transforms, which decompose a signal into different levels of resolution (Abibullaev et al., 2010; Adeli et al., 2007; Argoud et al., 2006; Artameeyanant et al., 2012; Calvagno et al., 2000; Chaibi et al., 2015; Chavakula et al., 2013; Goelz et al., 2000; Güler and Ubeyli, 2005, 2007; Guo et al., 2011; Guo et al., 2009; Halford et al., 2013; Indiradevi et al., 2008; Kalayci and Ozdamar, 1995; Latka et al., 2003; Lima et al., 2010; Liu et al., 2002; Orhan et al., 2011; Park et al., 1998; Sezer et al., 2012; Song and Zhang, 2013; Suresh and Balasubramanyam, 2013; Übeyli, 2008, 2009b; Wang et al., 2011; Zhou et al., 2013). Goelz et al. compared both types of transforms on the same dataset, concluding that continuous wavelet transforms led to better detection of isolated transients (Goelz et al., 2000).

Still, there was a lack of studies adequately assessing the algorithm’s performance. While one publication reported 91.7% sensitivity at 89.3% specificity, there was no separation between training and test sets, making these results unreliable (Indiradevi et al., 2008). Furthermore, choosing a suitable wavelet for EEG decomposition and computing an appropriate threshold was not trivial, and other types of algorithms started being developed to circumvent these shortcomings associated with parameter and threshold choices.

4.1.3. Morphological Filtering

Morphological filtering an alternative strategy used for IED detection. In these approaches, a structuring element is combined with the signal using a set operator (intersection, union, among others) (Juozapavicius et al., 2011; Xu et al., 2007; Xu et al., 2006). This changes the shape of the signal, separating the spikes from the background. An amplitude threshold can then be applied to the difference between the original signal and the background, aiming to detect IEDs.

Xu et al. developed a method based on the combination of different morphological operators (Xu et al., 2007). While it performed better than standard filter types typically used in
morphological operations, the presence of artefacts with morphological characteristics similar to IEDs led to a large number of false detections in all of the authors' approaches.

Juozapavicius et al. reached a similar conclusion. While their algorithm detected IEDs, it also identified other spikes (e.g. physiological spikes during sleep) as well as false positives in noisy recordings (Juozapavicius et al., 2011). This suggested that, while more complex, morphological filtering suffered from the same drawbacks as mimetic methods, due to the similarity of IEDs to other transients and artefacts.

4.1.4. Template Matching

Template matching is another strategy for IED detection, which started being used in the end of the 1990s (El-Gohary et al., 2008; Horak et al., 2015; Ji et al., 2011a; Ji et al., 2011b; Jing et al., 2016; Lodder et al., 2013; Lodder and van Putten, 2014; Nonclercq et al., 2012; Pietilä et al., 1994; Sankar and Natour, 1992; Thomas et al., 2017; Vijayalakshmi and Abhishek, 2010). It consists of comparing each IED candidate pattern to a set of templates, classifying it as an epileptiform transient if there is enough similarity (see Figure 3).

Figure 3

The similarity between each EEG segment and each template can be quantified in terms of distance (Jing et al., 2016; Sankar and Natour, 1992; Thomas et al., 2017), mean squared error (El-Gohary et al., 2008) or correlation (Horak et al., 2015; Vijayalakshmi and Abhishek, 2010). While these methods are more complex than feature thresholding, they do include a thresholding step since one needs to specify how small the distance must be or how large the correlation must be for a candidate pattern to be considered an IED. The threshold value has a high impact on IED recognition: a lower similarity threshold will lead to the detection of more epochs as IEDs, leading to more false positives (and vice versa).
The template pools used in these approaches must be large and representative enough to capture the inter-patient variability of IED patterns. To account for this, authors tended to use several multi-channel templates (El-Gohary et al., 2008; Ji et al., 2011a). Other ways of increasing robustness included using additional templates for the background patterns of the EEG (Thomas et al., 2017) and increasing the template pool by finding epochs similar to the existing templates (Lodder et al., 2013).

Self-adapting systems were also proposed to improve template matching (Lodder and van Putten, 2014). In this work, the system started by detecting possible IEDs based on the available templates and showing the prospective detections to the clinicians. If the detection was correct, the IED could be used as a template, increasing the pool of available. Using 241 IEDs, it was possible to detect a third of the transients after one iteration with the clinicians (in which the clinicians confirm correct classifications). This number rose 95% after 15 iterations, showing the impact of the expert input. However, as the authors point out, the system was optimized in terms of sensitivity and there was a trade-off with specificity, leading to a higher number of false positives, ranging from 0.24 to 6.6 per minute.

While this self-adapting system can contribute to the creation of larger template pools, the new templates will always be morphologically related to the previous ones (since the system detected the epochs through a previous template). Thus, if there is any type of morphology lacking in the original template pool, it will be very complicated for the system to bridge that gap on its own, constituting a limitation of this method.

Several template matching approaches were hindered by the use of very small datasets, which could not capture the morphological diversity of IEDs (El-Gohary et al., 2008; Ji et al., 2011a; Nonclercq et al., 2012). For instance, one algorithm trained on data from 6 patients yielded 97% sensitivity at 78% specificity when there were clear IEDs, but the performance dropped
to 31% sensitivity at 33% specificity in recordings where the discharges were less clear (i.e. closer to the clinical reality) (Pietilä et al., 1994).

4.2. Pre-defined Features and Machine Learning (II)

Machine learning (ML) methods take in pre-determined features and use them for classification. They include traditional machine learning classifiers (section 4.2.1) and small artificial neural networks (section 4.2.2). Unlike the previous approaches, these algorithms learn from the input features instead of being completely defined a priori. This strategy aimed to reduce user-defined parameters and typically use a large number of characteristics of the EEG for classification.

4.2.1. Traditional Machine Learning Classifiers

Traditional machine learning classifiers are a set of algorithms based on concepts such as point neighborhood, clustering or class division that adapt their parameters according to the input features, learning from them to classify new data.

K-nearest neighbors (KNN) is an example of an ML classifier proposed for IED detection (Guo et al., 2011; Iscan et al., 2011; Martinez-Vargas et al., 2011; Zhou et al., 2013). After feature extraction, for each sample, the algorithm finds a certain number (k) of training samples that are closest to it in the feature space - its k-nearest neighbors. Then, a voting scheme leads to classification: the sample is classified as 'having an IED' if most of its neighbors have an IED.

Clustering methods are another ML approach, of which k-means is the most common (Inan and Kuntalp, 2007; Nonclercq et al., 2012; Orhan et al., 2011). This algorithm partitions the training data into a number (k) of clusters. Then, each new sample is assigned the cluster with the closest center. More complex clustering techniques have been applied to IED detection. When comparing five different types of clustering (k-means, k-medoids, fuzzy C-means, agglomerative clustering and affinity propagation), a study showed that affinity propagation
yielded the best results (Thomas et al., 2017). This method does not require previous specification of the number of clusters. Instead, this is determined as the algorithm runs, since it finds 'exemplars', which are training samples representative of clusters.

Another widely used ML classifier is the support vector machine (SVM) (Acir and Güzelioğlu, 2004; Acir et al., 2005; Bagheri et al., 2018; Chavakula et al., 2013; Güler and Ubeyli, 2007; Iscan et al., 2011; Kelleher et al., 2010; Lima et al., 2010; Thomas et al., 2018). Given the training data, SVMs try to build a barrier (hyperplane) between each class, maximizing the margin between classes. A new sample is classified based on which side of the 'fence' it is, in the parameter space. Authors have compared the performance of SVM to a multilayer perceptron (MLP, described in section 4.2.2) on the same dataset, leading to the conclusion that the SVM shows superior accuracy in IED detection (Derya Ubeyli, 2008; Güler and Ubeyli, 2007).

One study compared the results of a traditional SVM and one of its variants, least-squares SVM (LS-SVM), to other ML classifiers (Iscan et al., 2011). The LS-SVM reconfigures the problem so that it is solved using a set of linear equations instead of quadratic ones. Ultimately, the LS-SVM surpassed all other ML classifiers tested in this dataset. Using a cascade of SVMs rather than a single one also led to increased precision and sensitivity (Bagheri et al., 2018). This was congruent with the previously mentioned study where affinity propagation beat k-means, leading to the conclusion that updated versions of traditional ML algorithms can improve IED detection (Thomas et al., 2017).

Comparing these algorithms not trivial, as several authors only provided qualitative results (Horak et al., 2015; Lima et al., 2010) or did not separate training and test sets (e.g. (Kelleher et al., 2010; Orhan et al., 2011)). One study reported an area under the receiver operating characteristic curve (AUC) of 0.93 on a patient-specific classifier, indicating that this
performance would drop when applied to another patient, showcasing the need for external validation (Kelleher et al., 2010). Furthermore, several studies reported accuracy as the only performance metric (e.g. (Güler and Ubeyli, 2007; Martinez-Vargas et al., 2011; Ubeyli, 2008)). Accuracy ranged from 75% to 99% but, as explained in section 3, this is not indicative of a good IED detector.

4.2.2. Small Artificial Neural Networks

Artificial neural networks (ANNs) are architectures of artificial neurons organized in layers, which have been used for IED detection since 1992 (Gabor and Seyal, 1992). They are similar to traditional ML classifiers in their ability to learn from data, but ANNs present a specific type of architecture and mathematical framework that distinguishes them from other ML approaches. This section will focus on ANNs with a shallow/small structure and that use features as input (i.e. are not used as end-to-end methods, as opposed to the ANNs discussed in section 4.4.).

Regarding the input of these ANNs, initially, morphological features were used (Boos et al., 2010; Castellaro et al., 2002; Gabor and Seyal, 1992; James et al., 1999; Webber et al., 1994; Wilson et al., 1999). Given the morphological similarities of IED to other EEG patterns, towards the end of the 1990s, frequency-based features were explored (Abibullaev et al., 2010; Adjouadi et al., 2004; Artameeyanant et al., 2012; Feucht et al., 1997; Güler and Ubeyli, 2005, 2007; Guo et al., 2011; Guo et al., 2009; Halford et al., 2013; Haydari et al., 2011; Liu et al., 2002; Orhan et al., 2011; Park et al., 1998; Sezer et al., 2012; Song and Zhang, 2013; Srinivasan et al., 2005; Suresh and Balasubramanyam, 2013; Tarassenko et al., 1998; Übeyli, 2009a; Wang et al., 2011).

In terms of architecture, the ANNs used for IED detection can be broadly divided into feed-forward and recurrent networks. One of the most widely used types of feed-forward ANN for
IED detection is the multilayer perceptron (MLP) – a simple type of network composed of stacked perceptrons (see Figure 4) (Abibullaev et al., 2010; Acir et al., 2005; Argoud et al., 2006; Boos et al., 2014; Castellaro et al., 2002; Feucht et al., 1997; Gabor and Seyal, 1992; Güler and Ubeyli, 2007; Kalayci and Ozdamar, 1995; Ko and Chung, 2000; Kutlu et al., 2009; Liu et al., 2002; Ozdamar and Kalayci, 1998; Pang et al., 2003; Park et al., 1998; Sezer et al., 2012; Tarassenko et al., 1998; Ubeyli, 2008; Übeyli, 2008, 2009b; Webber et al., 1994). Perceptrons are the building blocks of neural networks: the output of a perceptron is linked to the input of the next layer's perceptron, creating a network. Even within MLPs, there is architecture variability - the number of layers, nodes and activation function, among others, are differentiating parameters.

Figure 4

Aside from MLPs, other feed-forward ANN architectures have been used to automate IED detection (James et al., 1999; Kurth et al., 2000; Nigam and Graupe, 2004; Song and Zhang, 2013; Ubeyli, 2008; Übeyli, 2008, 2010; Wilson et al., 1999). Several authors compared the performance of MLPs and their own custom architectures (Güler et al., 2005; Ubeyli, 2008; Übeyli, 2008). Custom architectures outperformed the MLP in all the studies, showing that it is likely that more complex algorithms are necessary to fully capture and process the information given as input.

Recurrent neural networks (RNNs) differ from feed-forward networks by allowing information to travel forward and backward in the network through feedback loops. These have also been an option for IED detection, reaching similar performances (Güler et al., 2005; Sezer et al., 2012; Srinivasan et al., 2005; Übeyli, 2009a).

Some authors combined ANNs with traditional ML classifiers (Acir et al., 2005; Inan and Kuntalp, 2007; Thomas et al., 2018). One study used a neural network as a feature extractor
and the output of this step was fed to an SVM, which performed EEG-level classification (Thomas et al., 2018). Another study used two perceptrons to pre-classify EEG samples into definite IEDs, definite non-IEDs and possible IEDs (Acir et al., 2005). An SVM was then used to separate epochs in the third group. This study used data from 19 patients for training and 10 for testing, reaching 89.1% sensitivity at 85.9% specificity.

Unlike this last example, many studies using ANNs only reported accuracy (e.g. (Güler and Ubeyli, 2005; Güler et al., 2005; Übeyli, 2009a, 2009b, 2010)). Other studies used limited datasets, which does not contribute to generalization (Kalayci and Ozdamar, 1995; Kurth et al., 2000; Ozdamar and Kalayci, 1998; Song and Zhang, 2013; Tarassenko et al., 1998). Furthermore, not separating the training and test sets leads to unreliable results, since overtraining is one of the main possible drawbacks of ANNs, given the high number of trainable parameters. These parameters can be overly fitted to a training set, thus not being generalizable. In turn, running an overtrained model on an independent set would lead to a much lower performance. For instance, it is not likely that 100% sensitivity and specificity could be reached on an independent set (Sezer et al., 2012). Examples of studies that performed a proper train/test split and reported relevant metrics showed that it is possible to overcome the issue of overtraining and, with different morphological features and ANNs, sensitivity could range from 55.3% to 89.9%, at approximately 99% specificity (James et al., 1999; Wilson et al., 1999). While these results represent an improvement from previously described approaches, they mostly fell short of human performance (Wilson et al., 1999).

4.3. Expert Systems (III)

Expert systems (ESs) are fundamentally different from the previously described approaches, as they can only be used in conjunction with one of the classifiers described in sections 4.1. and 4.2. ESs for IED detection take the results of previous methods as input and refine the
decision (accepting or discarding the detections) by incorporating spatial and/or temporal context or previous knowledge given to the system. This is highly dependent on the human expert, since clinicians decide what information is relevant in this final classification step.

Mimetic methods have been used in conjunction with ESs (Benlamri et al., 1997; Black et al., 2000; Dingle et al., 1993). One approach applied an ES after thresholding morphological features to incorporate spatial context to reject artefacts (e.g. muscle spikes, eye blinks, electrode movement). Then, a combination of spatial and temporal context was used to approve the candidate patterns (Dingle et al., 1993).

ESs have also been coupled with ANNs (Argoud et al., 2006; Park et al., 1998; Tzallas et al., 2006a). One study used an ANN to pre-classify EEG segments into categories (e.g. spikes, muscle activity, eye blinks or sharp alpha activity) which were then classified by the ES using contextual information regarding synchronicity and channel adjacency (Tzallas et al., 2006b).

Studies using ESs suffer the same limitations as the previous ones, regarding the lack of separation of train and test sets (Dingle et al., 1993; Park et al., 1998), as well as improper performance metrics (Benlamri et al., 1997; Tzallas et al., 2006a). Furthermore, since ESs must be coupled with other classifiers, judging their performance comparatively is impossible. While they may aid several other algorithms as a final step, they cannot be implemented in the clinics on their own. Therefore, their contribution to classification should be studied on a case basis, comparing the algorithm with and without the ES.

4.4. Deep Learning (IV)

Deep learning methods encompass neural networks with a large number of layers (i.e. depth), distinguishing them from the smaller ANNs discussed in section 4.2.2. Typically, these networks are used as end-to-end methods: they take in raw data, learn their own representation of the feature space and classify new samples. This constitutes an important difference
regarding the methods in the previous sections, which need pre-defined features. Making the models feature-independent increases the likelihood that the algorithm will be able to capture more information from the EEG signal as it is not being forced to focus on a limited number of user-defined characteristics.

Before the rise of deep learning in computer science, authors compared the use of raw and parameterized data for IED detection (Ko and Chung, 2000; Webber et al., 1994). Studies showed below random performance with raw data, leading authors to conclude that it was impossible to use ANNs as end-to-end methods. However, with the development of novel architectures and availability of computing power, the paradigm changed.

Convolutional neural networks (CNNs) are a type of ANN that is often used in deep learning approaches. The core building block of CNNs are convolutional layers, which extract information using filters that are iteratively convolved with the input. The use of CNNs in IED detection is growing (Jing et al., 2020; Johansen et al., 2016; Lourenço et al., 2019; Tjepkema-Cloostermans et al., 2018), as these have shown potential in fields such as image analysis, but also in health-related applications. Figure 5 shows an example of a CNN architecture that was initially developed for image analysis and has been adapted for IED detection (Lourenço et al., 2019).

**Figure 5**

Applying CNNs to raw EEGs, it was possible to achieve areas under the receiver operating characteristic curve (AUCs) higher than 0.9, indicative of algorithms with very good performance (Jing et al., 2020; Johansen et al., 2016; Lourenço et al., 2019; Tjepkema-Cloostermans et al., 2018). Jing et al. (Jing et al., 2020) tested their CNN against commercially available software (Persyst 13, which uses extracted features and feed-forward NN rules (Scheuer et al., 2017)) and achieved an AUC 0.1 higher than this program on the same test set.
CNNs also yielded high sensitivity and specificity, reaching intersections of these metrics at 93% (Lourenço et al., 2019).

While these performances are impressive, the increased complexity of the methods (in terms of the number of trainable parameters) reduces the explainability of the decisions. CNNs are often called 'black-boxes', as it is not easy to understand what goes on inside the model - the information that the network retrieves from the input and its relative importance towards the final classification is not explicit. However, steps have been taken to increase the transparency of these techniques.

Tjepkema et al. (Tjepkema-Cloostermans et al., 2018) showed that two-dimensional (2D) CNNs worked better than one-dimensional (1D) CNNs for IED detection, when the input was given as a matrix of several channels over time. The difference is that 2D-CNNs have two-dimensional filters while 1D-CNNs have vector filters. This means that the 2D-CNN is able to capture information from several channels in each application of a filter kernel, while the 1D-CNN restrains its information retrieval to one channel. Since the 2D-CNNs yielded better performances, this implies that using more spatial context (in this case, using more than one channel in each application of the filter kernel) is of added benefit for classification.

Through the application of network visualization techniques, da Silva Lourenço et al. were able to show that the CNN was detecting IEDs in the samples (see Figure 6) (Lourenço et al., 2019). This proved that the network was classifying the samples based on the same visual stimulus an expert would use, rather than complementary features.

**Figure 6**

The number of studies applying deep networks to IED detection is still scarce. However, the field is developing fast and there is an increase in healthcare applications of deep networks.
Furthermore, the high sensitivity and specificity values obtained by several authors contributes to the indication that we should expect more of these approaches in the near future.

5. General Considerations

Despite the efforts towards automating IED detection, clinical use is still scarce. Previously, this was mostly due to the sub-par performance of the methods when compared to the gold standard. However, the current main hurdles are related to the small and incomparable datasets and performance measures.

Most authors used different datasets to train their algorithms, hindering result comparison. There is a lack of public standardized datasets with a large number of EEGs from patients in different age ranges, reviewed by a significant number of experts. This is mostly because expert opinion is expensive, and the organizations that are able to obtain such a dataset are typically not willing to share it. To change this situation, state-supported studies should be funded to allow the creation of this type of dataset and subsequent homogenization of model evaluation.

Furthermore, some algorithms are trained with very small datasets, which are not representative of IED diversity (El-Gohary et al., 2008; Ji et al., 2011a; Latka et al., 2003; Radmehr and Anisheh, 2013; Tarassenko et al., 1998; Tzallas et al., 2006b). Some authors do not use control data (from healthy subjects, e.g. (Gabor and Seyal, 1992; Inan and Kuntalp, 2007; Kurth et al., 2000)) and several works do not separate training and test sets (e.g. (Nigam and Graupe, 2004; Pang et al., 2003; Song and Zhang, 2013; Thomas et al., 2017)). This makes it impossible to judge the generalization ability of the algorithm, rendering the results unreliable.

Additionally, while some studies reported sensitivity and specificity, as well as false positive rates, several studies did not present these values (e.g. (Boos et al., 2010; Guedes de Oliveira et al., 1983; Hostetler et al., 1992; Ko and Chung, 2000)). Approaches reporting qualitative results also lack information needed to objectively assess the method and compare it to others.
(Calvagno et al., 2000; Ji et al., 2011a; Juozapavicius et al., 2011; Suresh and Balasubramanyam, 2013; Vijayalakshmi and Abhishek, 2010).

In fact, out of all the approaches reviewed in this paper, only six fulfilled two basic requirements: separating the training and test sets and reporting sensitivity and specificity or false positive rate (Acir and Güzelis, 2004; James et al., 1999; Lodder et al., 2013; Lourenço et al., 2019; Tjepkema-Cloostermans et al., 2018; Wilson et al., 1999). Still, even these papers were not directly comparable due to the difference in the datasets.

On a more global perspective, however, over the past forty years, there have been significant developments in the automation of IED detection. Authors went from trying to mimic the behavior of experts to creating more complex and diverse approaches. Figure 7 provides a chronological overview of the first use of the different types of methods.

**Figure 7**

False detections associated with morphological features were circumvented by using other characteristics and the development of fields such as machine learning led to a wider breadth of algorithms that could be applied to IED detection. Furthermore, more data became available and the computer power needed to process such data was accessible, allowing the training of more generalizable algorithms. As time progressed, there was a shift from older techniques, with moderate results, to newer approaches such as machine learning, with better results and deep learning, with impressive outcomes.

**6. Conclusion**

The methods used for automated IED detection have evolved over the past 45 years, with a resulting increase in performance. While the diversity of datasets and performance measures has hindered direct comparison of the studies, the pervasiveness of machine and deep learning
techniques has increased over the past years. The results of these approaches, which seem to surpass those of other methods, further contribute to increasing interest in this line of research.

It is expected that the upward trend of machine and deep learning in IED detection will continue. It is also likely that these algorithms will become an integral part of the assessment of clinical EEG, reducing variability, subjectivity and time to diagnosis.

Conflict of Interests

M.J.A.M. van Putten is co-founder of Clinical Science Systems, a supplier of EEG systems for Medisch Spectrum Twente. Clinical Science Systems offered no funding and was not involved in the design, execution, analysis, interpretation or publication of the study. The remaining authors have disclosed that they do not have any conflicts of interest.

Acknowledgements

This research was funded by the Epilepsiefonds Foundation, grant number WAR16-08.
References

Abibullaev, B., Seo, H. D., & Kim, M. S. (2010). Epileptic spike detection using continuous wavelet transforms and artificial neural networks. *Int. J. Wavelets Multiresolution Inf. Process.*, 8(01), 33-48.

Acir, N., & Güzelis, C. (2004). Automatic spike detection in EEG by a two-stage procedure based on support vector machines. *Comput Biol Med, 34*(7), 561-575. doi:10.1016/j.compbiomed.2003.08.003

Acir, N., Oztura, I., Kuntalp, M., Baklan, B., & Güzelis, C. (2005). Automatic detection of epileptiform events in EEG by a three-stage procedure based on artificial neural networks. *IEEE Trans Biomed Eng, 52*(1), 30-40. doi:10.1109/tbme.2004.839630

Adeli, H., Ghosh-Dastidar, S., & Dadmehr, N. (2007). A wavelet-chaos methodology for analysis of EEGs and EEG subbands to detect seizure and epilepsy. *IEEE Trans Biomed Eng, 54*(2), 205-211. doi:10.1109/tbme.2006.886855

Adjouadi, M., Sanchez, D., Cabrerozo, M., Ayala, M., Jayakar, P., Yaylali, I., et al. (2004). Interictal spike detection using the Walsh transform. *IEEE Trans Biomed Eng, 51*(5), 868-872. doi:10.1109/tbme.2004.826642

Andrzejak, R. G., Lehnertz, K., Mormann, F., Rieke, C., David, P., & Elger, C. E. (2001). Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: dependence on recording region and brain state. *Phys Rev E Stat Nonlin Soft Matter Phys, 64*, 061907. doi:10.1103/PhysRevE.64.061907

Argoud, F. I., De Azevedo, F. M., Neto, J. M., & Grillo, E. (2006). SADE3: an effective system for automated detection of epileptiform events in long-term EEG based on context information. *Med Biol Eng Comput, 44*(6), 459-470. doi:10.1007/s11517-006-0056-y

Artameeyanant, P., Chiracharit, W., & Chamnongthai, K. (2012). Spike and epileptic seizure detection using wavelet packet transform based on approximate entropy and energy with artificial neural network. Paper presented at the The 5th 2012 Biomedical Engineering International Conference.

Bagheri, E., Jin, J., Dauwels, J., Cash, S., & Westover, M. B. (2018). CLASSIFIER CASCADE TO AID IN DETECTION OF EPILEPTIFORM TRANSIENTS IN INTERICTAL EEG. In *Proc IEEE Int Conf Acoust Speech Signal Process, 2018*, 970-974. doi:10.1109/icassp.2018.8461992

Benlamri, R., Batouche, M., Rami, S., & Bouanaka, C. (1997). An automated system for analysis and interpretation of epileptiform activity in the EEG. *Comput Biol Med, 27*(2), 129-139. doi:10.1016/s0010-4825(96)00040-6

Black, M. A., Jones, R. D., Carroll, G. J., Dingle, A. A., Donaldson, I. M., & Parkin, P. J. (2000). Real-time detection of epileptiform activity in the EEG: a blinded clinical trial. *Clin Electroencephalogr, 31*(3), 122-130. doi:10.1177/155005940003100304

Boos, C. F., do Carmo Vitarelli Pereira, M., Marques Argoud, F. I., & Mendes de Azevedo, F. (2010). Morphological descriptors for automatic detection of epileptiform events. In *Annu Int Conf IEEE Eng Med Biol Soc, 2010*, 2435-2438. doi:10.1109/iembs.2010.5626339

Boos, C. F., Scolaro, G. R., & Azevedo, F. M. (2014). *Comparison between Artificial Neural Networks and Discriminant Functions for Automatic Detection of Epileptiform Discharges*. Paper presented at the XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013.

Calvagno, G., Ermanli, M., Rinaldo, R., & Sartoretti, F. (2000). A multiresolution approach to spike detection in EEG. Paper presented at the 2000 ICASSP.

Castellar, C., Favaro, G., Castellar, A., Casagrande, A., Castellar, S., Puthenparampil, D. V., et al. (2002). An artificial intelligence approach to classify and analyse EEG traces. *Neurophysiol Clin, 32*(3), 193-214. doi:10.1016/s0987-7053(02)00302-7
Chaibi, S., Lajnef, T., Ghrob, A., Samet, M., & Kachouri, A. (2015). A Robustness Comparison of Two Algorithms Used for EEG Spike Detection. Open Biomed Eng J, 9, 151-156. doi:10.2174/1874120701509010151

Chavakula, V., Sánchez Fernández, I., Peters, J. M., Popli, G., Bosl, W., Rakrade, S., et al. (2013). Automated quantification of spikes. Epilepsy Behav, 26(2), 143-152. doi:10.1016/j.yebeh.2012.11.048

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educ Psychol Meas, 20(1), 37-46.

de Curtis, M., Jefferys, J. G. R., & Avoli, M. (2012). Interictal epileptiform discharges in partial epilepsy. In: Jasper's Basic Mechanisms of the Epilepsies. 4th edition: National Center for Biotechnology Information (US).

Derya Ubeyli, E. (2008). Analysis of EEG signals by combining eigenvector methods and multiclass support vector machines. Comput Biol Med, 38(1), 14-22. doi:10.1016/j.compbiomed.2007.06.002

Dingle, A. A., Jones, R. D., Carroll, G. J., & Fright, W. R. (1993). A multistage system to detect epileptiform activity in the EEG. IEEE Trans Biomed Eng, 40(12), 1260-1268. doi:10.1109/10.250582

El-Gohary, M., McNames, J., & Elsas, S. (2008). User-guided interictal spike detection. Annu Int Conf IEEE Eng Med Biol Soc, 2008, 821-824. doi:10.1109/iembs.2008.4649280

Feucht, M., Hoffmann, K., Steinberger, K., Witte, H., Benninger, F., Arnold, M., et al. (1997). Simultaneous spike detection and topographic classification in pediatric surface EEGs. Neuroreport, 8(9-10), 2193-2197. doi:10.1097/00001756-199707070-00021

Fisher, R. S., Acevedo, C., Arzimanoglou, A., Bogacz, A., Cross, J. H., Elger, C. E., et al. (2014). ILAE official report: a practical clinical definition of epilepsy. Epilepsia, 55(4), 475-482. doi:10.1111/epi.12550

Gabor, A. J., & Seyal, M. (1992). Automated interictal EEG spike detection using artificial neural networks. Electroencephalogr Clin Neurophysiol, 83(5), 271-280. doi:10.1016/0013-4694(92)90086-w

Goelz, H., Jones, R. D., & Bones, P. J. (2000). Wavelet analysis of transient biomedical signals and its application to detection of epileptiform activity in the EEG. Clin Electroencephalogr, 31(4), 181-191. doi:10.1177/1550059400310406

Gotman, J. (1982). Automatic recognition of epileptic seizures in the EEG. Electroencephalogr Clin Neurophysiol, 54(5), 530-540. doi:10.1016/0013-4694(82)90038-4

Gotman, J., & Gloor, P. (1976). Automatic recognition and quantification of interictal epileptic activity in the human scalp EEG. Electroencephalogr Clin Neurophysiol, 41(5), 513-529. doi:10.1016/0013-4694(76)90063-8

Gotman, J., & Wang, L. Y. (1991). State-dependent spike detection: concepts and preliminary results. Electroencephalogr Clin Neurophysiol, 79(1), 11-19. doi:10.1016/0013-4694(91)90151-s

Gotman, J., & Wang, L. Y. (1992). State dependent spike detection: validation. Electroencephalogr Clin Neurophysiol, 83(1), 12-18. doi:10.1016/0013-4694(92)90127-4

Guedes de Oliveira, P., Queiroz, C., & Lopes da Silva, F. (1983). Spike detection based on a pattern recognition approach using a microcomputer. Electroencephalogr Clin Neurophysiol, 56(1), 97-103. doi:10.1016/0013-4694(83)90011-1

Güler, I., & Ubeyli, E. D. (2005). Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients. J Neurosci Methods, 148(2), 113-121. doi:10.1016/j.jneumeth.2005.04.013

Güler, I., & Ubeyli, E. D. (2007). Multiclass support vector machines for EEG-signals classification. IEEE Trans Inf Technol Biomed, 11(2), 117-126. doi:10.1109/titb.2006.879600

Güler, N. F., Übeyli, E. D., & Güler, I. (2005). Recurrent neural networks employing Lyapunov exponents for EEG signals classification. Expert Syst. Appl., 29(3), 506-514.
Guo, L., Rivero, D., Dorado, J., Munteanu, C. R., & Pazos, A. (2011). Automatic feature extraction using genetic programming: An application to epileptic EEG classification. *Expert Syst. Appl.*, 38(8), 10425-10436.

Guo, L., Rivero, D., Seoane, J. A., & Pazos, A. (2009). Classification of EEG signals using relative wavelet energy and artificial neural networks. In *Proceedings of the first ACM/SIGEVO Summit on Genetic and Evolutionary Computation* (pp. 177-184).

Halford, J. J., Schalkoff, R. J., Zhou, J., Benbadis, S. R., Tatum, W. O., Turner, R. P., et al. (2013). Standardized database development for EEG epileptiform transient detection: EEGnet scoring system and machine learning analysis. *J Neurosci Methods*, 212(2), 308-316. doi:10.1016/j.jneumeth.2012.11.005

Haydari, Z., Zhang, Y., & Soltanian-Zadeh, H. (2011). *Semi-automatic epilepsy spike detection from EEG signal using genetic algorithm and wavelet transform*. Paper presented at the 2011 IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBM).

Horak, P. C., Meisenhelter, S., Testorf, M. E., Connolly, A. C., Davis, K. A., & Jobst, B. C. (2015). *Implementation and evaluation of an interictal spike detector*. Paper presented at the Image Reconstruction from Incomplete Data VIII.

Hostetler, W. E., Doller, H. J., & Homan, R. W. (1992). Assessment of a computer program to detect epileptiform spikes. *Electroencephalogr Clin Neurophysiol*, 83(1), 1-11. doi:10.1016/0013-4694(92)90126-3

Inan, Z. H., & Kuntalp, M. (2007). A study on fuzzy C-means clustering-based systems in automatic spike detection. *Comput Biol Med*, 37(8), 1160-1166. doi:10.1016/j.compbiomed.2006.10.010

Indiradevi, K. P., Elias, E., Sathidevi, P. S., Dinesh Nayak, S., & Radhakrishnan, K. (2008). A multi-level wavelet approach for automatic detection of epileptic spikes in the electroencephalogram. *Comput Biol Med*, 38(7), 805-816. doi:10.1016/j.compbiomed.2008.04.010

Iscan, Z., Dokur, Z., & Demiralp, T. (2011). Classification of electroencephalogram signals with combined time and frequency features. *Expert Syst. Appl.*, 38(8), 10499-10505.

James, C. J., Jones, R. D., Bones, P. J., & Carroll, G. J. (1999). Detection of epileptiform discharges in the EEG by a hybrid system comprising mimetic, self-organized artificial neural network, and fuzzy logic stages. *Clin Neurophysiol*, 110(12), 2049-2063. doi:10.1016/s1388-2457(99)00168-6

Ji, Z., Sugi, T., Goto, S., Wang, X., & Nakamura, M.. (2011a). *Multi-channel template extraction for automatic EEG spike detection*. Paper presented at the The 2011 IEEE/ICME International Conference on Complex Medical Engineering.

Ji, Z., Wang, X., Sugi, T., Goto, S., & Nakamura, M. (2011b). *Automatic spike detection based on real-time multi-channel template*. Paper presented at the 2011 4th International Conference on Biomedical Engineering and Informatics (BMEI).

Jing, J., Dauwels, J., Rakthanmanon, T., Keogh, E., Cash, S. S., & Westover, M. B. (2016). Rapid annotation of interictal epileptiform discharges via template matching under Dynamic Time Warping. *J Neurosci Methods*, 274, 179-190. doi:10.1016/j.jneumeth.2016.02.025

Jing, J., Sun, H., Kim, J. A., Herlopian, A., Karakis, I., Ng, M., et al. (2020). Development of Expert-Level Automated Detection of Epileptiform Discharges During Electroencephalogram Interpretation. *JAMA Neurol*, 77(1), 103-108. doi:10.1001/jamaneurol.2019.3485

Johansen, A. R., Jin, J., Maszczyn, T., Dauwels, J., Cash, S. S., & Westover, M. B. (2016). *EPILEPTIFORM SPIKE DETECTION VIA CONVOLUTIONAL NEURAL NETWORKS*. *Proc IEEE Int Conf Acoust Speech Signal Process*, 2016, 754-758. doi:10.1109/icassp.2016.7471776

Juozapavicius, A., Bacevicius, G., Bugelskis, D., & Samaitiene, R. (2011). EEG analysis–automatic spike detection. *Nonlinear Anal-Model*, 16(4), 375-386.

Kalayci, T., & Ozdamar, O. (1995). Wavelet preprocessing for automated neural network detection of EEG spikes. *IEEE ENG MED BIOL*, 14(2), 160-166.
Kelleher, D., Temko, A., Nash, D., McNamara, B., & Marnane, W. (2010). SVM detection of epileptiform activity in routine EEG. *Annu Int Conf IEEE Eng Med Biol Soc, 2010*, 6369-6372. doi:10.1109/ems.2010.5627297

Ko, C. W., & Chung, H. W. (2000). Automatic spike detection via an artificial neural network using raw EEG data: effects of data preparation and implications in the limitations of online recognition. *Clin Neurophysiol, 111*(3), 477-481. doi:10.1016/s1388-2457(99)00284-9

Kurth, C., Gilliam, F., & Steinhoff, B. J. (2000). EEG spike detection with a Kohonen feature map. *Ann Biomed Eng, 28*(11), 1362-1369. doi:10.1114/1.1331312

Kutlu, Y., Kuntalp, M., & Kuntalp, D. (2009). Optimizing the performance of an MLP classifier for the automatic detection of epileptic spikes. *Expert Syst. Appl.*, 36(4), 7567-7575.

Latka, M., Was, Z., Kozik, A., & West, B. J. (2003). Wavelet analysis of epileptic spikes. *Phys Rev E Stat Nonlin Soft Matter Phys, 67*, 052902. doi:10.1103/PhysRevE.67.052902

Lima, C. A., Coelho, A. L., & Eisencraft, M. (2010). Tackling EEG signal classification with least squares support vector machines: a sensitivity analysis study. *Comput Biol Med, 40*(8), 705-714. doi:10.1016/j.compbiomed.2010.06.005

Liu, H. S., Zhang, T., & Yang, F. S. (2002). A multistage, multimethod approach for automatic detection and classification of epileptiform EEG. *IEEE Trans Biomed Eng, 49*, 1557-1566. doi:10.1109/tbme.2002.805477

Lodder, S. S., & van Putten, M. J. (2014). A self-adapting system for the automated detection of interictal epileptiform discharges. *PloS One, 9*(1), e85180. doi:10.1371/journal.pone.0085180

Lopes da Silva, F., Blanes, W., Kalitzin, S. N., Parra, J., Suffczynski, P., & Velis, D. N. (2003). Epilepsies as dynamical diseases of brain systems: basic models of the transition between normal and epileptic activity. *Epilepsia, 44 Suppl 12*, 72-83. doi:10.1111/j.0013-9580.2003.12005.x

Loureço, C., Tjepkema-Cloostermans, M. C., Teixeira, L. F., & van Putten, M. J. A. M. (2019). *Deep Learning for Interictal Epileptiform Discharge Detection from Scalp EEG Recordings*. Paper presented at the Mediterranean Conference on Medical and Biological Engineering and Computing.

Martinez-Vargas, J. D., Avendano-Valencia, L. D., Giraldo, E., & Castellanos-Dominguez, G. (2011). *Comparative analysis of Time Frequency Representations for discrimination of epileptic activity in EEG Signals*. Paper presented at the 2011 5th International IEEE/EMBS Conference on Neural Engineering.

Nigam, V. P., & Graupe, D. (2004). A neural-network-based detection of epilepsy. *Neural Res, 26*(1), 55-60. doi:10.1179/016164104773026534

Nowack, W. J. (1997). Epilepsy: a costly misdiagnosis. *Clin Electroencephalogr, 28*(4), 225-228. doi:10.1177/15505949702800407

Orhan, U., Hekim, M., & Ozer, M. (2011). EEG signals classification using the K-means clustering and a multilayer perceptron neural network model. *Expert Syst. Appl.*, 38(10), 13475-13481.

Ozdamar, O., & Kalayci, T. (1998). Detection of spikes with artificial neural networks using raw EEG. *Comput Biomed Res, 31*(2), 122-142. doi:10.1006/cbmr.1998.1475
Pang, C. C., Upton, A. R., Shine, G., & Kamath, M. V. (2003). A comparison of algorithms for detection of spikes in the electroencephalogram. *IEEE Trans Biomed Eng, 50*(4), 521-526. doi:10.1109/tbme.2003.809479

Park, H. S., Lee, Y. H., Kim, N. G., Lee, D. S., & Kim, S. I. (1998). Detection of epileptiform activities in the EEG using neural network and expert system. *Stud Health Technol Inform, 52*, 1255-1259.

Pietilä, T., Vapaakoski, S., Nousiainen, U., Värri, A., Frey, H., Häkkinen, V., et al. (1994). Evaluation of a computerized system for recognition of epileptic activity during long-term EEG recording. *Electroencephalogr Clin Neurophysiol, 90*(6), 438-443. doi:10.1016/0013-4694(94)00134-1

Radmehr, M., & Anisheh, S. M. (2013). EEG spike detection using stationary wavelet transform and time-varying autoregressive model. *Int. J. Comput. Appl, 83*(13).

Sankar, R., & Natour, J. (1992). Automatic computer analysis of transients in EEG. *Comput Biol Med, 22*(6), 407-422. doi:10.1016/0010-4825(92)90040-t

Scheuer, M. L., Bagic, A., & Wilson, S. B. (2017). Spike detection: Inter-reader agreement and a statistical Turing test on a large data set. *Clin Neurophysiol, 128*(1), 243-250. doi:10.1016/j.clinph.2016.11.005

Sezer, E., İşik, H., & Saracoğlu, E. (2012). Employment and comparison of different Artificial Neural Networks for epilepsy diagnosis from EEG signals. *J Med Syst, 36*(1), 347-362. doi:10.1007/s10916-010-9480-5

Song, Y., & Zhang, J. (2013). Automatic recognition of epileptic EEG patterns via extreme learning machine and multiresolution feature extraction. *Expert Syst. Appl., 40*(14), 5477-5489.

Srinivasan, V., Eswaran, C., & Sripaaam, N. (2005). Artificial neural network based epileptic detection using time-domain and frequency-domain features. *J Med Syst, 29*(6), 647-660. doi:10.1007/s10916-005-6133-1

Suresh, H. N., & Balasubramaniam, V. (2013). *Wavelet transforms and neural network approach for epileptical EEG*. Paper presented at the 2013 3rd IEEE International Advance Computing Conference (IACC).

Tarassenko, L., Khan, Y. U., & Holt, M. R. G. (1998). Identification of inter-ictal spikes in the EEG using neural network analysis. *IEE Proc-Sci Meas Technol, 145*(6), 270-278.

Tatum, W. O., Rubboli, G., Kaplan, P. W., Mirsatari, S. M., Radhakrishnan, K., Gloss, D., et al. (2018). Clinical utility of EEG in diagnosing and monitoring epilepsy in adults. *Clin Neurophysiol, 129*(5), 1056-1082. doi:10.1016/j.clinph.2018.01.019

Thomas, J., Comoretto, L., Jin, J., Dauwels, J., Cash, S. S., & Westover, M. B. (2018). EEG Classification Via Convolutional Neural Network-Based Interictal Epileptiform Event Detection. *Annu Int Conf IEEE Eng Med Biol Soc, 2018, 3148-3151*. doi:10.1109/embc.2018.8512930

Thomas, J., Jing, Jin, Dauwels, J., Cash, S. S., & Westover, M. B. (2017). Automated epileptiform spike detection via affinity propagation-based template matching. *Annu Int Conf IEEE Eng Med Biol Soc, 2017, 3057-3060*. doi:10.1109/embc.2017.8037502

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

Tzallas, A. T., Karvelis, P. S., Katsis, C. D., Fotiadis, D. I., Giannopoulos, S., & Konitsiotis, S. (2006a). A method for classification of transient events in EEG recordings: application to epilepsy diagnosis. *Methods Inf Med, 45*(6), 610-621.

Tzallas, A. T., Oikonomou, V. P., & Fotiadis, D. I. (2006b). Epileptic spike detection using a Kalman filter based approach. *Conf Proc IEEE Eng Med Biol Soc, 2006, 501-504*. doi:10.1109/embs.2006.260780

Übeyli, E. D. (2008). Implementing eigenvector methods/probabilistic neural networks for analysis of EEG signals. *Neural Netw, 21*(9), 1410-1417. doi:10.1016/j.neunet.2008.08.005

Übeyli, E. D. (2008). Wavelet/mixture of experts network structure for EEG signals classification. *Expert Syst. Appl., 34*(3), 1954-1962.
Übeyli, E. D. (2009a). Analysis of EEG signals by implementing eigenvector methods/recurrent neural networks. *Digit. Signal Process.*, 19(1), 134-143.

Übeyli, E. D. (2009b). Combined neural network model employing wavelet coefficients for EEG signals classification. *Digit. Signal Process.*, 19(2), 297-308.

Übeyli, E. D. (2010). Lyapunov exponents/probabilistic neural networks for analysis of EEG signals. *Expert Syst. Appl.*, 37(2), 985-992.

Vijayalakshmi, K., & Abhishek, A. M. (2010). Spike detection in epileptic patients EEG data using template matching technique. *Int. J. Comput. Appl*, 2(6), 5-8.

Wang, D., Miao, D., & Xie, C. (2011). Best basis-based wavelet packet entropy feature extraction and hierarchical EEG classification for epileptic detection. *Expert Syst. Appl.*, 38(11), 14314-14320.

Webber, W. R., Litt, B., Wilson, K., & Lesser, R. P. (1994). Practical detection of epileptiform discharges (EDs) in the EEG using an artificial neural network: a comparison of raw and parameterized EEG data. *Electroencephalogr Clin Neurophysiol*, 91(3), 194-204. doi:10.1016/0013-4694(94)90069-8

Wilson, S. B., & Emerson, R. (2002). Spike detection: a review and comparison of algorithms. *Clin Neurophysiol*, 113(12), 1873-1881. doi:10.1016/s1388-2457(02)00297-3

Wilson, S. B., Turner, C. A., Emerson, R. G., & Scheuer, M. L. (1999). Spike detection II: automatic, perception-based detection and clustering. *Clin Neurophysiol*, 110(3), 404-411. doi:10.1016/s1388-2457(98)00023-6

Xu, G., Wang, J., Zhang, Q., Zhang, S., & Zhu, J. (2007). A spike detection method in EEG based on improved morphological filter. *Comput Biol Med*, 37(11), 1647-1652. doi:10.1016/j.compbiomed.2007.03.005

Xu, G., Wang, J., Zhang, Q., & Zhu, J. (2006). *An automatic EEG spike detection algorithm using morphological filter*. Paper presented at the 2006 IEEE International Conference on Automation Science and Engineering.

Zhou, J., Schalkoff, R. J., Dean, B. C., & Halford, J. J. (2013). A study of morphology-based wavelet features and multiple-wavelet strategy for EEG signal classification: results and selected statistical analysis. *Annu Int Conf IEEE Eng Med Biol Soc*, 2013, 5998-6002. doi:10.1109/embc.2013.6610919
Legends

Figure 1 - Examples of interictal epileptiform transients. A) Interictal spike, B) Group of interictal spikes, C) Sharp wave, D) Fast activity (brushes) riding on a spike, E) Paroxysmal slow activity superimposed to slow spikes. Adapted from (Noebels et al., 2012).

Figure 2 - A) Electroencephalographic (EEG) signal, B) Segments of the EEG signal, defined as the section between two extrema of amplitude, C) Sequences of the EEG signal, obtained by imposing a minimum difference between segments to reduce the impact of noise. This allows the algorithm to ‘see’ the wave between A, F and I (Gotman and Gloor, 1976).

Figure 3 - Example of a template (in red). To find matches, characteristics such as correlation coefficients, amplitude and variance are compared between the candidate electroencephalogram (EEG) and the template. By doing this across the EEG signal, it is possible to detect interictal epileptiform discharges, since these patterns will be the most similar to the templates. For example, the EEG segment highlighted in blue is much more similar to the template than the one highlighted in purple (Lodder et al., 2013).

Figure 4 - Basic structure of a perceptron. The inputs, $x_0$ to $x_n$, are combined with varying weights ($w_0$ to $w_n$). This results in a weighed sum that is then passed through a step function (or a more complex activation function), predicting 1 if the result was above a certain threshold and 0 otherwise.

Figure 5 – Example of a convolutional neural network architecture that has been adapted for interictal epileptiform discharge detection (Lourenço et al., 2019). This network includes several blocks of convolutional and pooling layers, ending with fully connected layers.

Figure 6 - Probability heatmap obtained with the occlusion visualization technique on an electroencephalogram sample with a focal interictal epileptiform discharge (IED) (Lourenço et al., 2019). Warmer colors indicate higher importance for classification, showing that the convolutional network detected the IED shape, revealing some information about the network’s decision process.

Figure 7 - Chronological overview of the first report of each type of automated approach for interictal epileptiform discharge detection. The first paper using mimetic methods was published in 1976, followed by artificial neural networks (ANNs), template matching, expert systems (ESs), thresholding of frequency features, machine learning (ML), morphological filtering and deep learning (DL).
