

Efficient use of clinical EEG data for deep learning in epilepsy

Catarina da Silva Lourenço^{a,*}, Marleen C. Tjepkema-Cloostermans^{a,b},
Michel J.A.M. van Putten^{a,b}

^a Department of Clinical Neurophysiology, Institute for Technical Medicine, University of Twente, Technical Medical Centre, Enschede, the Netherlands

^b Neurocentrum, Medisch Spectrum Twente, Enschede, the Netherlands



ARTICLE INFO

Article history:

Accepted 22 January 2021

Available online 26 March 2021

Keywords:

Deep learning

Interictal epileptiform discharges

Data augmentation

Convolutional neural networks

Electroencephalogram

HIGHLIGHTS

- Augmenting datasets improves the performance of neural networks for interictal epileptiform discharge detection.
- Time shifting and different montages can reduce the need for annotated data.
- Deep learning may cause a fundamental shift in clinical EEG analysis.

ABSTRACT

Objective: Automating detection of Interictal Epileptiform Discharges (IEDs) in electroencephalogram (EEG) recordings can reduce the time spent on visual analysis for the diagnosis of epilepsy. Deep learning has shown potential for this purpose, but the scarceness of expert annotated data creates a bottleneck in the process.

Methods: We used EEGs from 50 patients with focal epilepsy, 49 patients with generalized epilepsy (IEDs were visually labeled by experts) and 67 controls. The data was filtered, downsampled and cut into two second epochs. We increased the number of input samples containing IEDs through temporal shifting and using different montages. A VGG C convolutional neural network was trained to detect IEDs.

Results: Using the dataset with more samples, we reduced the false positive rate from 2.11 to 0.73 detections per minute at the intersection of sensitivity and specificity. Sensitivity increased from 63% to 96% at 99% specificity. The model became less sensitive to the position of the IED in the epoch and montage.

Conclusions: Temporal shifting and use of different EEG montages improves performance of deep neural networks in IED detection.

Significance: Dataset augmentation can reduce the need for expert annotation, facilitating the training of neural networks, potentially leading to a fundamental shift in EEG analysis.

© 2021 International Federation of Clinical Neurophysiology. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Interictal epileptiform discharges (IEDs) are transient patterns that reflect an increased likelihood of epileptic seizures (Pillai and Sperling, 2006; Smith, 2005). IEDs are present in about half of the routine EEG recordings of epilepsy patients, rising to 80% in sleep recordings (Halford, 2009). Visual analysis of EEG signals by experts is currently the gold standard in IED detection (Lodder et al., 2014), but this approach entails several disadvantages. The learning curve is long, review times are significant and

specialized personnel is required. Furthermore, intra and inter-rater variability leads to error rates up to 25% (Benbadis and Lin, 2008).

Automating IED detection can reduce the resources spent on visual analysis, time to diagnosis and the misdiagnosis rate. Several approaches have been developed for this purpose. Most are based on 'pre-chosen' features, which might limit the algorithm's ability to learn how to detect these transients and seem to justify the moderate performance of these algorithms. More recently, end-to-end deep learning approaches have been used (Tjepkema-Cloostermans et al., 2018; Lourenço et al., 2020; Johansen et al., 2016; Jing et al., 2019), which can learn their own representation of the feature space from raw data (LeCun et al., 2015).

Convolutional neural networks (CNNs) have been able to accurately detect IEDs and several approaches have been explored. A 3-

* Corresponding author.

E-mail addresses: c.dasilvalourenco@utwente.nl (C. da Silva Lourenço), M.Tjepkema-Cloostermans@mst.nl (M.C. Tjepkema-Cloostermans), m.j.a.m.vanputten@utwente.nl (M.J.A.M. van Putten).

layer CNN trained on raw data achieved an area under the Receiver Operating Characteristic curve (AUC) of 0.95 (Johansen et al., 2016). A two-dimensional CNN led to an AUC of 0.94, with 47.4% sensitivity at 98.0% specificity (Tjepkema-Cloostermans et al., 2018). A known CNN architecture from the literature (VGG) was also used for IED detection, achieving an AUC of 0.96 (Lourenço et al., 2020). Another architecture (SpikeNet) was able to outperform commercially available software (Persyst 13 (Scheuer et al., 2017)) in IED detection (Jing et al., 2019). The authors were also able to train the network to classify whole EEGs with an AUC of 0.85.

Given the high number of trainable parameters, large volumes of data are needed to train deep neural networks appropriately. For supervised learning, this data must be labeled so that the network has a way of assessing its errors and iteratively improve its performance, getting closer to the gold standard (i.e. label provided by experts) (Taylor and Nitschke, 2017; Perez and Wang, 2017). This means that every IED in an EEG recording should be labeled by experts (preferably several to establish consensus, given the high variability), which limits the availability of such data. In turn, this creates a bottleneck in the development of generalizable algorithms of this type. Increasing the number of input samples using augmentation techniques or temporal shifting, can potentially improve performance.

Data augmentation techniques use existing samples to create new ones, aiming to improve the accuracy or robustness of a classifier. When working with images, cropping, padding, flipping and other transformations are often used (Taylor and Nitschke, 2017; Perez and Wang, 2017). However, given that EEGs are time-series, many of these approaches are not applicable, as they would interfere with the temporal component of the signal. The spatial component of the EEG (i.e. the channels) is also relevant in visual analysis and possibly in the training of the classifier. Since several montages can be used by experts to identify IEDs, we hypothesize that a transformation in what concerns the order of the channels could contribute to a more robust classifier.

Temporal shifting, i.e. shifting the window of acquisition of the samples in time, creates novel context for the network. As such, it also increases the number of samples available for training and can be applied to EEG signals without loss of temporal continuity, adding novel temporal context to the IEDs.

We build on previous work (Lourenço et al., 2020) and explore the impact of applying temporal shifting and using data in different montages to increase the number of IED samples used in the training process of a convolutional neural network.

2. Methods

2.1. EEG data and pre-processing

We used EEG data from 166 patients, randomly selected from the digital database of the Medisch Spectrum Twente, in the Netherlands (Lourenço et al., 2020). All EEGs were obtained as part of routine care, and anonymized before analysis. As EEG is part of routine care, the Medical Ethical Committee Twente waived the need for informed consent for continuous EEG monitoring. Interictal EEGs (with IEDs) from patients with focal (50 patients) and generalized (49 patients) epilepsy were included, along with normal EEGs from 67 healthy controls. The IEDs were visually labeled by experts (MvP and MTC).

EEG data was filtered in the 0.5–30 Hz range and downsampled to 125 Hz, aiming to reduce artefacts and data dimensionality. The signals were split into 2 s non-overlapping epochs. These steps were implemented in Matlab R2019a (The MathWorks, Inc., Natick, MA).

2.2. Dataset creation

Four different datasets were created using the original and augmented data (see Table 1). Set A was comprised of the aforementioned data, using the longitudinal bipolar montage (DB). Set B included the same data in three different montages: longitudinal bipolar, source derivation (SD) and common average (G19). Since the last two montages include 19 channels instead of the 18 present in the DB montage, the last channel of the SD and G19 samples was discarded. This led to a threefold multiplication of the number of samples (since the same transformation was applied to IED and non-IED samples). Fig. 1 shows an example of an epoch in the three montages. New samples containing IEDs were generated by shifting the time window of the epoch by 0.5 s, 1 s and 1.5 s, as can be seen in Fig. 2. This resulted in a fourfold multiplication of the number of samples with IEDs, which were used in Set C. To create Set D, data was time-shifted in each montage (DB, SD and G19).

Data was randomized and a 80/20 split into a training/validation and test set was applied. All epochs from a particular patient were used either for training or testing. Fivefold cross validation was applied on the training/validation set.

2.3. Deep learning models

A VGG C convolutional neural network (see Supplementary Material - Supplementary Fig. S1), was implemented in Python 3.4 using Keras 2.1.2, Tensorflow 1.4.0 and a CUDA-enabled NVIDIA GPU (GTX-1080), running on CentOS 7. Stochastic optimization was performed using an Adam optimizer (Kingma and Ba, 2014) with a learning rate of $2 * 10^{-5}$, $\beta_1 = 0.91$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$. A sparse categorical cross entropy function was employed to estimate the loss. A batch size of 64 and weights of 100:1 were used (100 corresponding to the positive class, i.e. samples with IEDs) for sets A through C. Technical details can be found in the Supplementary Material - Technical aspects regarding the implementation of the algorithm.

2.4. Performance evaluation

Model performance was evaluated based on the average Receiver Operating Characteristic (ROC) curves, built for all cross-validation sets, based on 1001 discretizations. The corresponding area under the curve (AUC) and Confidence Intervals (CIs) at 95% were calculated. Sensitivity, specificity and false positive detection rates per minute were assessed at the intersection of sensitivity and specificity and at 99% specificity. These routines were implemented in Matlab R2019a (The MathWorks, Inc., Natick, MA).

Table 1

Datasets used for training/validation of the neural network, with # IED samples being the number of input samples containing Interictal Epileptiform Discharges. The montages mentioned in the table are longitudinal bipolar (DB), source derivation (SD) and common average (G19).

Training Set	Data Description	# IED samples
Set A	Original Dataset in DB montage	2220
Set B	Dataset augmented by using three different montages: DB + SD and G19 montages	8496
Set C	Dataset augmented by using temporal shifting: DB + 3 time shifts	6808
Set D	Dataset augmented by using three different montages and temporal shifting: DB + SD and G19 montages + 3 time shifts	25302

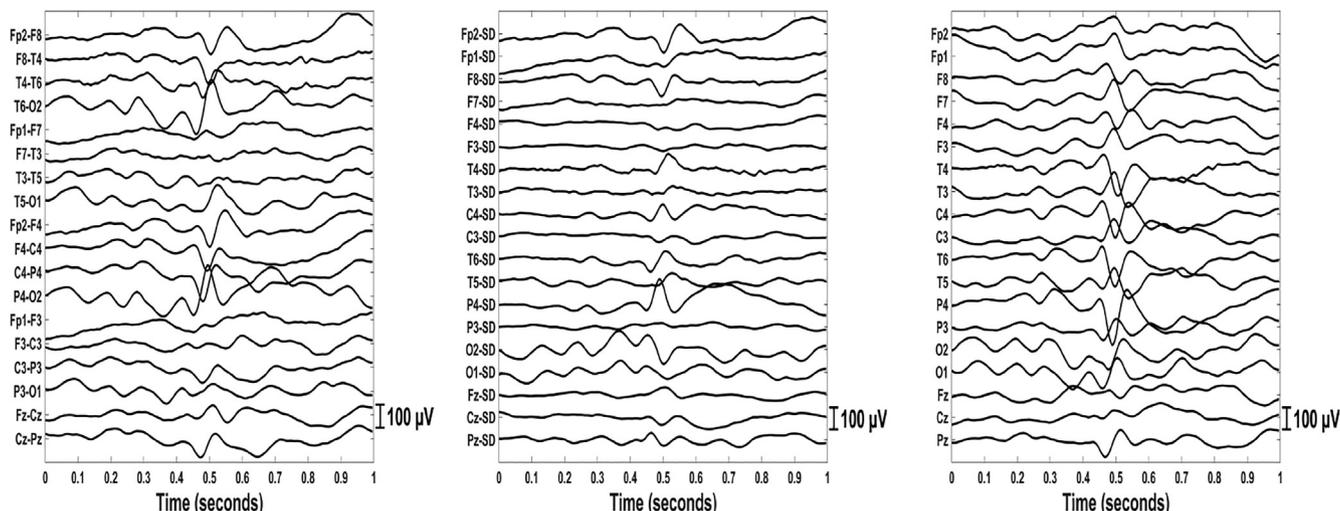


Fig. 1. Example of an EEG sample in three different montages. Shown is an Interictal Epileptiform Discharge (IED) in the right centro-parieto-temporal region. Left: longitudinal bipolar (DB), Middle: source derivation (SD), Right: common average (G19).

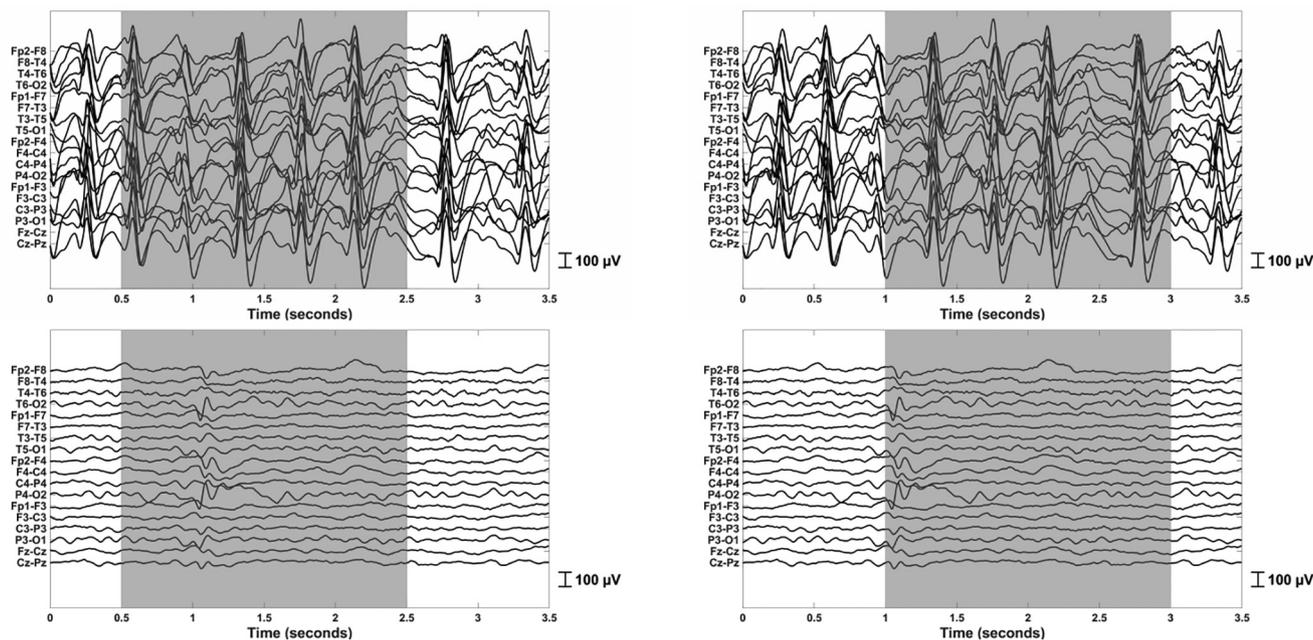


Fig. 2. Example of a temporal shift of 0.5 s applied to two EEG samples. The top panels show the inclusion of new Interictal Epileptiform Discharges (IEDs) while the bottom panels illustrate the addition of novel context to the sample containing the IED.

2.5. Occlusion technique

Occlusion is a network visualization technique used to assess the relative importance of each part of the input samples in the network’s classification (Zeiler and Fergus, 2014). Typically, the sample is divided into small sections and each of these is iteratively occluded. The difference between the prediction of the full sample and the prediction of the sample with the occluded section is calculated. When plotting this difference, warmer colours were used for larger difference values, indicating higher importance in classification. More details can be found in (Lourenço et al., 2020).

3. Results

The VGG C architecture was trained with the four different datasets. With the original dataset (Set A), the intersection

between sensitivity and specificity occurred at 93% with a false positive rate of 2.11 (0.85–3.38) false detections per minute. With data from different montages (Set B), using 1001 thresholds, it was not possible to completely match sensitivity and specificity. The closest values for these parameters were 90% and 93%, respectively. Using time-shifting (Set C), the intersection was at 96% and combining the two techniques (Set D), it was possible to achieve an intersection at 97%. The corresponding false positive rate for Set D was 0.73 (0.00–1.54) false detections per minute.

At 99% specificity, the sensitivity improved from 63% for the original dataset (Set A) to 96% for the augmented dataset (Set D, using both time shifting and different montages). In Table 2, we summarize performance for the various datasets. Fig. 3 shows the ROC curve for the models trained with Set A and Set D, with the corresponding areas under the curve (AUC) of 0.95 and 0.98.

Focusing on the classification of EEGs from individual patients of the test set (Supplementary Tables S2 and S3), it was possible

Table 2

Average sensitivity (Sens), specificity (Spec) and false positives per minute (FP/min) for the VGG trained with the different datasets. The 95% Confidence Intervals are also shown. The results for the intersection of sensitivity and specificity are shown, along with the corresponding false positive (FP) rate. These results are followed by the sensitivity at 99% specificity for each set. Set A: original dataset, Set B: augmented dataset using three montages, Set C: augmented dataset using temporal shifting, Set D: augmented dataset with both using different montages and temporal shifting.

	FP/min (95% CI)	Sens (95% CI)	Spec (95% CI)
Set A	2.1 (0.9–3.4)	92.7 (88.7–96.8) 63.5 (53.6–73.3)	92.8 (88.4–97.1) 99.1 (98.6–99.5)
Set B	1.9 (1.3–2.6)	89.5 (82.6–96.4) 74.2 (67.9–80.5)	93.5 (91.2–95.7) 99.0 (98.5–99.5)
Set C	1.2 (0.7–1.6)	96.1 (93.1–99.0) 81.1 (73.5–88.7)	96.1 (94.4–97.7) 99.0 (98.8–99.4)
Set D	0.7 (0.0–1.5)	97.5 (94.9–100.0) 95.9 (91.6–100.0)	97.5 (94.7–100.0) 99.0 (97.8–100.0)

to reach a sensitivity of 100% in 5 out of 20 patients with the model trained on Set A and in 17 patients with the model trained on Set D. The average sensitivity for IED detection of the model trained with Set A on the individual test EEGs was 69.24%, and the corresponding specificity was 93.23%. For the model trained with Set D, these values were 94.4% and 89.58%, respectively. It is also possible to assess the performance of the model on focal and generalized IEDs separately. The model trained with Set A detected focal IEDs with a sensitivity of 54.74% at 91.88% specificity, while it was able to detect generalized IEDs with a sensitivity of 83.74% at 94.59% specificity. For the model trained with Set D, the corresponding values were 89.92% sensitivity at 87.71% specificity for focal discharges and 98.89% sensitivity at 91.45% specificity for generalized IEDs. The models trained with Sets A and D were able to classify normal EEG samples from healthy controls with an average specificity of 99.21% and 99.61%, respectively.

Figs. 4 and 5 illustrate the results of occlusion on a sample containing a focal IED before and after temporal shifting by one second, classified by the models trained with Set A and D. Fig. 4 illustrates that, before shifting, there is an isolated red patch, associated with a correct classification by the model trained with Set A, whereas on the second panel there is a more diffuse red area and a consequent misclassification of the sample.

In Fig. 5, the same sample shown in Fig. 4 was both time-shifted and re-referenced three times, followed by classification by the model trained on Set D. It is possible to see isolated red patches

on all panels, regardless of channel order or the position of the IED, and all versions of the sample were correctly classified by the network.

4. Discussion

We studied the effect of temporal shifting and the use of different montages to increase the number of samples used to train a VGG C network for IED detection.

Datasets B through D, which included data obtained by shifting the acquisition window or changing the montage (or both), led to a decrease in the false positive rate at the intersection of sensitivity and specificity when compared to the model trained with Set A. This shows that more training samples lead to an improvement of the model’s performance, in accordance with previous findings (Taylor and Nitschke, 2017; Perez and Wang, 2017). Furthermore, the models trained with Sets C and D also led to a higher sensitivity at 99% specificity when compared to Set A.

The model trained with set C (time-shifted data) leads to a lower false positive rate (1.91 versus 1.15 false detections per minute) and higher sensitivity at 99% specificity (74.20% versus 81.11%) when compared to the one trained with different montages (set B). This is due to the difference in difficulty of the task itself. Since Set B contains samples with different channel orders, the model needs to learn to ‘see’ what IEDs look like in each of them. With Set C, this is not necessary as all the epochs are in the DB montage and only the position of the IEDs changes.

The model trained with set D was able to outperform the other three models in false positive rate, intersection of sensitivity and specificity and also sensitivity at 99% specificity (cf Table 2). While the AUC value for this model was also higher (0.98) (cf Fig. 3), this improvement was not as resounding as in the previously mentioned variables since the AUC of the model trained with Set A was already quite high (0.95).

The improvements of the model trained with Set D were also shown in the classification of EEGs from individual patients. This model achieved a higher average specificity when classifying EEG samples without IEDs and it detected all the IEDs present in 17 out of the 20 epilepsy patients of the test set, while the model trained with Set A was only able to reach 100% sensitivity in 5 patients. Considering patients with focal and generalized epilepsy separately, the average sensitivity of IED detection increased for both types of epilepsy. Focal IEDs were detected with 35% more sensitivity, while the increase for generalized discharges was

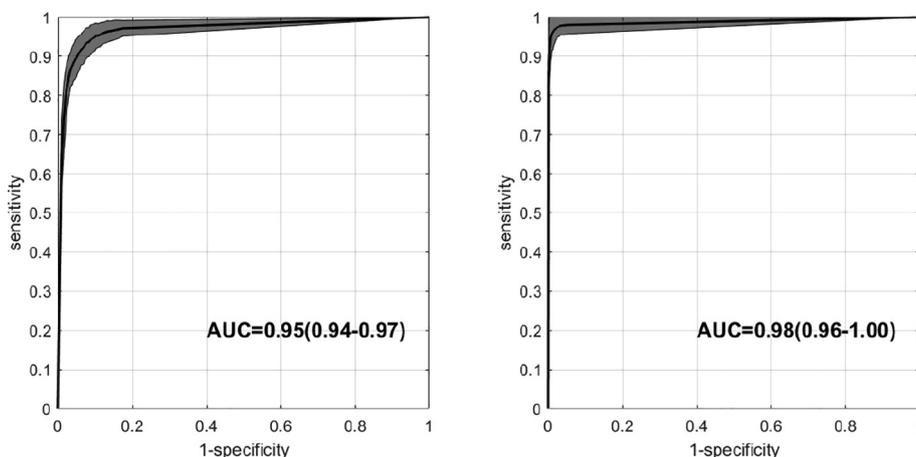


Fig. 3. Receiver Operating Characteristic (ROC) curves for the VGG model trained with Set A (left) and Set D (right) on the test set. The 95% Confidence Interval of the ROC curves is shown as a shaded area. The right panel shows a higher area under the ROC curve (AUC).

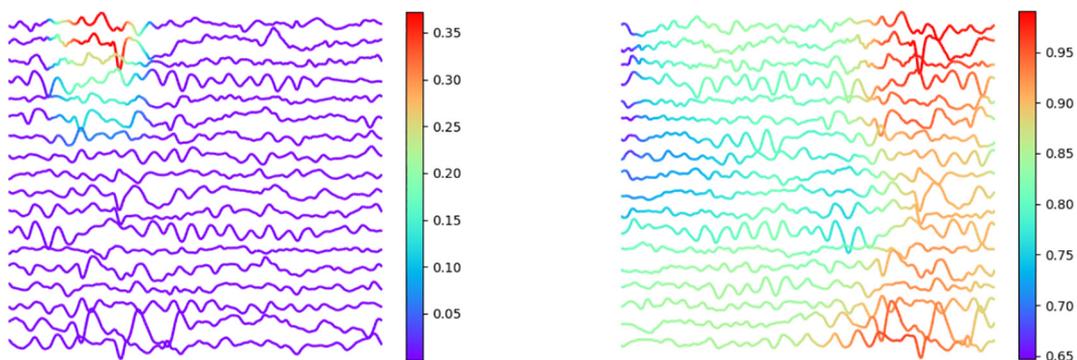


Fig. 4. Probability heatmaps obtained with occlusion on an EEG sample with a focal Interictal Epileptiform Discharge (IED), classified by the model trained with Set A (original dataset). Warmer colors indicate higher importance for classification. The scale shows the difference between the probability assigned to the epoch and what is obtained when a patch is occluded. This scale is relative to the model and sample and, as such, should not be compared between panels. On the left panel, the IED is correctly identified by the model and the sample is classified as a True Positive. On the right panel, the same sample is shown after 1 s shifting. This panel shows a large and non-specific red patch, indicating that the network was not able to see the IED in this position. The sample was misclassified as a False Negative.

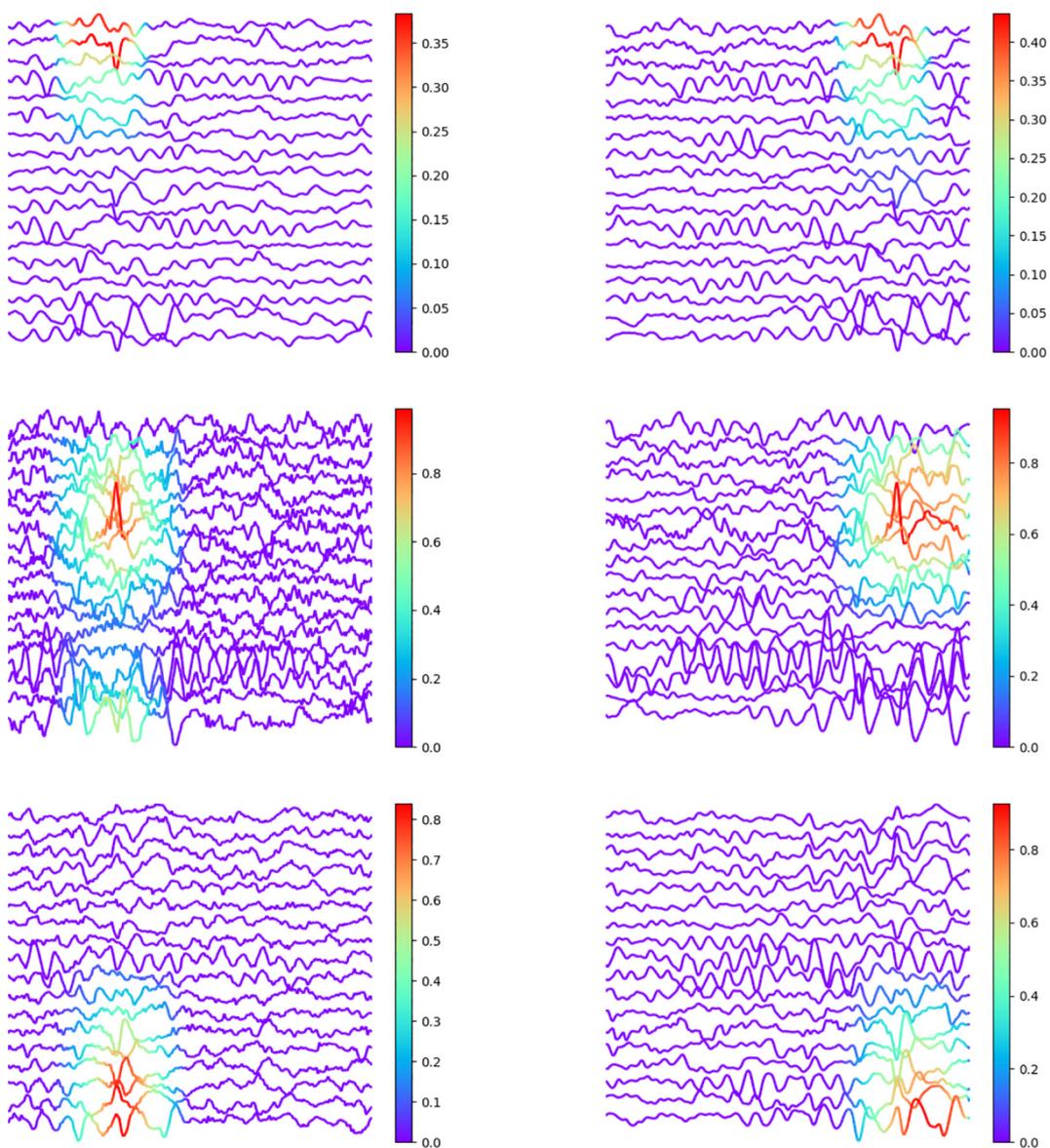


Fig. 5. Probability heatmaps obtained with occlusion on the same EEG epoch with a focal Interictal Epileptiform Discharge (IED) shown in Fig. 4, classified by the model trained with Set D. Warmer colours indicate higher importance for classification. The scale shows the difference between the probability assigned to the epoch and what is obtained when a patch is occluded. This scale is relative to the model and sample and, as such, should not be compared between panels. All the versions of the sample were classified as True Positives by the network and the IED was identified. Top: longitudinal bipolar montage (DB), Middle: source derivation (SD), Bottom: common average (G19); Left: no temporal shift, Right: 1 s. temporal shift.

15%. This is most likely due to the higher subtlety and complexity of focal discharges, which are typically more difficult to recognize. While the model trained with Set A was already quite satisfactory at detecting generalized discharges (sensitivity of 83.74%), it struggled with focal IEDs (sensitivity of 54.74%). This meant that there was more room for improvement, as noted by the results obtained when training the model with Set D.

Our results are also up to par with other CNN-based approaches described in the literature. Reported AUCs of 0.95 (Johansen et al., 2016) and 0.94, obtained using a subset of Set A for training (Tjepkema-Cloostermans et al., 2018), do not surpass the 0.98 obtained with the VGG trained on Set D. Persyst 13 is a commercially available for IED detection (Scheuer et al., 2017). This algorithm achieved 43.9% sensitivity and 1.65 false detections per minute. Despite the use of a vastly different dataset, our networks (in particular the one trained with Set D), far surpass this sensitivity and also lead to lower false positive rates.

The occlusion technique was used to elucidate which parts of the input are apparently relevant in the network's classification process. As shown in Fig. 4, it is possible to see that the model trained with Set A had a bias towards the position of the IED within the sample, as it was able to correctly identify the transient in the beginning of the sample but not after it was shifted in time, positioning the IED in the last second of the epoch. This is due to the way the experts labeled the IEDs: they identified the transient and tended to start its label roughly 0.5 s before the IED itself. Therefore, many of the training samples had this configuration, leading the network to learn this bias from the experts.

By shifting the data in time, we create samples that do not follow this trend, increasing the variety of the training data. Retraining the VGG with these samples rendered the network insensitive to the location of the IED, making the model more invariant and increasing its discriminative power. Adding re-referenced samples further improved the flexibility of the network by forcing it to 'look' at how IEDs are represented in different channel orders.

Combining these two methods and creating Set D led to a larger increase in the number of samples (there was a twelve-fold increase in samples containing IEDs when compared to Set A) but also in the variety of the samples, as two different approaches were used. This model was able to eliminate the position bias seen in the model trained with Set A. Furthermore, it was possible to confirm that the network learned to identify IED patterns in the three montages used in training, regardless of their position (cf Fig. 5).

This improved version of the model is more suitable for clinical application, given its generalization ability and satisfactory classification performance. After training the neural network with more varied samples, obtained through temporal shifting and re-montaging, it is able to identify IEDs in any of the three montages used for training, regardless of their position within the sample. Such a model can be used as an assistive tool for IED detection by clinicians. Using a graphical interface such as an EEG viewer, signals can be fed to the algorithm and the output can be sorted according to the probability of an epoch containing an IED. Showing the samples with higher probability first, the expert can go through as many as necessary to assess whether the patterns are relevant and enough to diagnose the patient.

This would lead to a significant decrease in the time needed to analyse EEGs from prospective epilepsy patients, as experts would be able to 'jump' to potentially relevant segments instead of scanning the whole signal. Logically, the relevance of this type of tool increases when longer EEG recordings are considered, since the analysis of these signals is proportionally more time consuming. Furthermore, situations where the marking of spikes is necessary over an entire recording, such as EEG analysis in presurgical evaluation for epilepsy surgery, can also benefit from this algorithm. In

these cases, after applying the neural network, the clinician would only need to review the classification, which would also lead to workload reduction.

Since the aim is to use the neural network as an assistive tool, it would still be up to the clinician to judge whether the detected spikes are relevant and correspond to IEDs. This is important since it is possible that the EEG recording includes 'epileptiform variants' (i.e. patterns that look like IEDs but are not significant for the diagnosis, such as wicket waves or small sharp spikes) that can be disregarded by the human reviewer.

5. Conclusion

We show that increasing the number of samples in the training set of a neural network through time-shifting and using different montages increases the discriminative power for IED detection and makes the model more generalizable. Combining these two strategies lowered the false positive rate and increased sensitivity when compared to the separate use of time-shifting and re-referencing. Furthermore, it was possible to eliminate a bias in the algorithm towards the location of the IED, shown by occlusion. Using these techniques to multiply available samples by several folds can reduce the bottleneck created by the scarceness of expert annotated data.

Declaration of Competing Interest

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.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.clinph.2021.01.035>.

References

- Benbadis SR, Lin K. Errors in eeg interpretation and misdiagnosis of epilepsy. *Eur Neurol* 2008;59:267–71.
- Halford JJ. Computerized epileptiform transient detection in the scalp electroencephalogram: Obstacles to progress and the example of computerized eeg interpretation. *Clin Neurophysiol* 2009;120:1909–15.
- Jing J, Sun H, Kim JA, Herlopian A, Karakis I, Ng M, et al. Development of expert-level automated detection of epileptiform discharges during electroencephalogram interpretation. *JAMA Neurol* 2019. <https://doi.org/10.1001/jamaneurol.2019.3485>.
- Johansen AR, Jin J, Maszczyk T, Dauwels J, Cash SS, Westover MB. Epileptiform spike detection via convolutional neural networks. In: *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on. IEEE; 2016. p. 754–8.*
- Kingma DP, Ba J. Adam: A method for stochastic optimization; 2014. arXiv preprint arXiv:1412.6980.
- LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436.
- Lodder SS, Askamp J, van Putten MJ. Computer-assisted interpretation of the eeg background pattern: a clinical evaluation. *PLoS One* 2014;9:e85966.
- Lourenço C, Tjepkema-Cloostermans MC, Teixeira LF, van Putten MJAM. Deep learning for interictal epileptiform discharge detection from scalp eeg recordings. In: *Henriques J, Neves N, de Carvalho P, editors. XV Mediterranean Conference on Medical and Biological Engineering and Computing – MEDICON 2019. Cham: Springer International Publishing; 2020. p. 1984–97.*
- Perez L, Wang J. The effectiveness of data augmentation in image classification using deep learning; 2017. arXiv preprint arXiv:1712.04621.

Pillai J, Sperling MR. Interictal eeg and the diagnosis of epilepsy. *Epilepsia* 2006;47:14–22.

Scheuer ML, Bagic A, Wilson SB. Spike detection: Inter-reader agreement and a statistical turing test on a large data set. *Clin Neurophysiol* 2017;128:243–50.

Smith S. Eeg in the diagnosis, classification, and management of patients with epilepsy. *J Neurol Neurosurg Psychiatry Res* 2005;76:ii2–7.

Taylor L, Nitschke G. Improving deep learning using generic data augmentation; 2017. arXiv preprint arXiv: 1708.06020.

Tjepkema-Cloostermans MC, de Carvalho RC, van Putten MJ. Deep learning for detection of focal epileptiform discharges from scalp eeg recordings. *Clin Neurophysiol* 2018;129:2191–6.

Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. In: European conference on computer vision. Springer; 2014. p. 818–33.