Not one size fits all: influence of EEG type when training a deep neural network for interictal epileptiform discharge detection

Catarina da Silva Lourenço MSc. [1], Marleen C. Tjepkema-Cloostermans PhD [1,2], Michel J. A. M. van Putten MD PhD [1,2]

[1]Department of Clinical Neurophysiology, Institute for Technical Medicine, University of Twente, Technical Medical Centre, Enschede, The Netherlands

[2]Department of Neurology and Neurophysiology, Hospital Medisch Spectrum Twente (MST), Enschede, The Netherlands

Correspondence: Catarina da Silva Lourenço, Technical Medical Centre, Hallenweg 15, 7522 NB, Enschede, The Netherlands, Email: c.dasilvalourenco@utwente.nl; Michel van Putten, Technical Medical Centre, Hallenweg 15, 7522 NB, Enschede, The Netherlands, Email: m.j.a.m.vanputten@utwente.nl

Abstract

Objective: Deep learning methods have shown potential in automating interictal epileptiform discharge (IED) detection in electroencephalograms (EEGs). However, it is known that these algorithms are dependent on the type of data used for training and this is not currently explored in EEG analysis applications. We aim to explore the difference in performance of artificial neural networks on routine and ambulatory EEG data.

Methods: We trained the same neural network on three datasets: 166 routine EEGs (VGGC–R), 75 ambulatory EEGs (VGGC–R) and a combination of the two data types (VGGC-C, 241 EEGs total). These networks were tested on 34 routine EEGs and 33 ambulatory recordings. Sensitivity, specificity and false positive rate (FPR) were calculated at a 0.99 probability threshold.

Results: The VGGC-R led to 84% sensitivity at 99% specificity on the routine EEGs, but its sensitivity was only 53% on ambulatory EEGs, with FPR > 3 FP/min. The VGGC-Cand VGGC-A yielded sensitivities of 79% and 60%, respectively, at 99% specificity on ambulatory data, but their sensitivity was under 60% for routine EEGs.

Conclusion: We show that the VGGC-R should be used for routine recordings and the VGGC-C should be used for ambulatory recordings for IED detection.

Significance: As different networks work better for different types of data, algorithms should be trained with the same type of EEG data they will be applied to, either routine or ambulatory.

Introduction

Interictal epileptiform discharges (IEDs) in electroencephalograms (EEGs) reflect an increased likelihood of seizures and their characteristics can assist in the identification and classification of epilepsy syndromes [1, 2]. Visual analysis of EEG recordings by experts is currently the gold standard in IED detection, and a crucial part of epilepsy diagnosis, as well as monitoring the effect of therapy [3].

Visual analysis is a time and resource-consuming endeavor, as EEG review times are long and trained experts are required. Aside from the long learning curve, inter and intra-subject variability constitute further drawbacks of visual analysis [4, 5]. Given the increasing number of ambulatory and even ultra-long EEG recordings of months to years, made possible by technological developments, the resources needed for this task have increased exponentially [6-8].

Automating IED detection could reduce the number of work hours invested in visual analysis, streamlining procedures in the clinic [2, 9]. This is particularly useful when a large volume of data needs to be analyzed. There have been several approaches aiming to automate IED detection [10]. These range from thresholding of morphological or frequency features [11, 12] to traditional machine learning methods [13, 14] and, more recently, deep learning techniques [15-20]. The popularity of deep learning methods in the medical field has grown in the past years, with applications from predicting mortality from echocardiographic video [21] to skin cancer classification [22]. In EEG analysis, and specifically in epilepsy, deep learning has been applied to tasks such as IED detection, seizure detection and seizure prediction [10, 23-25].

The type and quality of the data used for training and validation of deep learning algorithms is critical, as it also defines the boundary conditions for their applicability. As neural networks learn from experience, they will inevitably make mistakes when exposed to types of data that were not included in the training process [26]. This has been explored in fields such as computer vision or natural language processing, where strategies such as transfer learning are employed to extend model

applicability in new contexts [27, 28]. This limitation may also apply to deep learning approaches for IED detection, but this has been left unexplored in current EEG literature. Most studies that study deep networks for IED detection use routine EEG recordings for training, with an average duration of 20-30 minutes [17-20], while in other works data length and acquisition conditions are not clearly specified [16].

Neural networks trained on EEG data recorded in a hospital environment may perform differently if applied to EEGs recorded in other situations. Routine data is acquired in a more controlled environment and under expert supervision, resulting in less artefacts. Ambulatory data usually contains more and more diverse artefacts, ranging from loose electrodes to electrical interference, chewing, movement, among others. Furthermore, ambulatory EEGs contain periods of sleep, which are not typically present in routine recordings. This means that rhythms and patterns typical of sleep such as sleep spindles, K-complexes and vertex waves, some of which can be misidentified as epileptiform abnormalities, are widely present in ambulatory recordings but not in routine EEGs [29, 30]. It is therefore unclear whether the performance of a neural network trained with routine EEG data is maintained when applied to ambulatory recordings. This is clinically highly relevant, as such recordings are increasingly available and used, among others, for IED detection [6-8].

To the best of our knowledge, there have been no studies comparing the performance of algorithms in different types of EEG data. Here, we compare the performance of three deep neural networks trained with different types of EEG data on a test set comprised of routine and ambulatory EEG recordings. We previously reported the performance of the network trained on routine data on an independent test set [15]. The two other algorithms have not been previously published. Further, the test sets used on this study are different and do not overlap with the data used in [15].

Methods

EEG data and pre-processing

We used data from 308 patients, randomly selected from the database of the Medisch Spectrum Twente, in the Netherlands. Recordings from 159 epilepsy patients were included, as well as 149 EEGs that were classified as normal. In the recordings from epilepsy patients, the IEDs were visually labeled by experts (MvP and MTC). As EEG is part of routine care, the Medical Ethical Committee Twente waived the need for informed consent. All EEGs were anonymized before analysis.

Of the 308 recordings, 200 were routine EEGs (average duration of 20 minutes) and 108 were ambulatory EEGs (average duration of 20 hours). Routine and ambulatory recordings have significant differences, as ambulatory EEGs include e.g. periods of sleep, chewing, and other artefacts such as electrodes with poor contact, among others, as shown in Figure 1.

EEG data was filtered in the 0.5–30 Hz range in the longitudinal bipolar montage and downsampled to 125 Hz, aiming to reduce artefacts and data dimensionality. The signals, in the longitudinal bipolar montage, were split into 2 s epochs. These steps were implemented in Matlab R2021b (The MathWorks, Inc., Natick, MA).

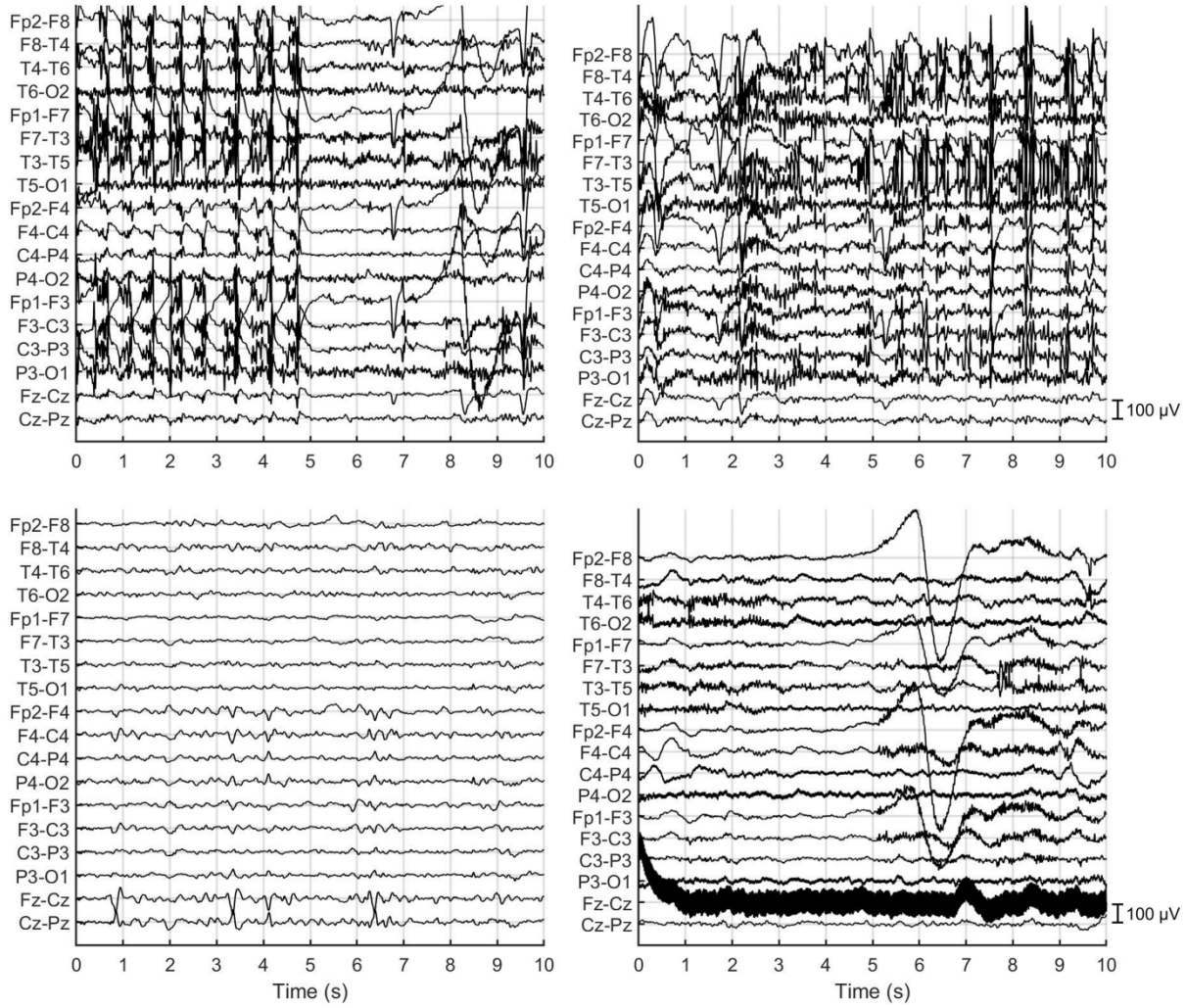

Figure 1. Examples of patterns that can be present in ambulatory EEG recordings. The top panels show two examples of chewing, the bottom left panel includes vertex waves (clearly visible in channels Fz-Cz and Cz-Pz) and the bottom right panel shows an example of an electrode (Fz, visible in Fz-Cz) with poor contact. Another artefact in this recording is the large deflection arising from Fp2. Epoch length is 10 s, bipolar montage, filter settings 0.5-30 Hz.

Dataset Creation

The EEG data was first separated into independent training and test sets, ensuring that all data epochs from a particular patient were used either for training or for testing. Training data was further divided into three datasets: set R, containing only the routine data; set A, comprised of the ambulatory data

and set C, a combination of the previous two datasets. Dataset R had EEGs from 166 subjects, while dataset A contained recordings from 75 subjects. The combination set, C, had data from 241 patients. This is summarized in Table 1.

Training data was augmented to increase the number of training samples, as described in [15]. In short: we used three different EEG montages: longitudinal bipolar (DB), laplacian (SD) and common average (G19). In the SD and G19 montages, the last channel was removed to maintain 18 channels across all samples. Further, to increase the number of IEDs in the dataset, the epochs that contained an IED were shifted with 0 s, 0.5 s, 1 s and 1.5 s. This resulted in the multiplication of the number of IEDs in the datasets by approximately 12 fold (see Table 1).

Table 1. Overview of training datasets: R, comprised of routine data; A, comprised of ambulatory EEGs and C, including both routine and ambulatory recordings. The number of normal EEGs and EEGs with interictal discharges is shown, as well as the number of IEDs in the recordings. The last row shows the number of IEDs after re-montaging in three different montages and shifting the data acquisition window by 0.5 seconds, resulting in a multiplication of approximately 12 times. It was not an exact 12-fold increase due to some epochs coinciding with end of file, for example.

	R	A	C
# normal EEGs	67	49	116
# EEGs with IEDs	99	26	125
# IEDs	2220	3262	5482
# IEDs after augmentation	25302	37878	63180

Since the test set should be representative of all types of data, we randomly selected 18 routine EEGs with IEDs and 16 routine EEGs that were classified as normal, as well as 16 ambulatory EEGs with IEDs

and 17 ambulatory EEGs that were classified as normal (see Table 2). The test set included over 2.300 IEDs and 255 h of EEG data.

Table 2. Overview of routine and ambulatory test datasets, including normal EEGs and EEGs with IEDs. The number of interictal discharges in the recordings is also shown.

	Routine	Ambulatory
# normal EEGs	16	17
# EEGs with IEDs	18	16
# IEDs	403	1918

Deep Learning models

A modified VGG C [31] convolutional neural network was implemented in Python 3.8 using Tensorflow 2.2.0 and a CUDA–enabled NVIDIA GPU (GTX–1080), running on Rocky Linux 8. The modifications done to the original architecture were mostly related to data dimensionality: the input dimensions were changed to [number of epochs x 250 x 18], with 250 corresponding to 2 seconds of data sampled at 125 Hz and 18 being the number of channels. In the final dense layer, we used two nodes instead of 1000 to fit the binary classification problem.

Stochastic optimization was performed using an Adam optimizer [32] with a learning rate of $2*10^{-5}$, β_1=0.91, β_2=0.999 and ε=10-8. A sparse categorical cross entropy function was employed to estimate the loss and a batch size of 64 was used.

This architecture was trained with the three datasets: R, A and C, as described in the previous section. This originated the three trained networks: VGGC–R, VGGC–A and VGGC–C. Given the class imbalance of the dataset (the number of samples with IEDs - positive class - was much lower than that of normal samples – negative class), class weights were used. For the VGGC-R(routine model), we used weights of 100:1 (100 corresponding to the positive class, i.e. samples with IEDs). This meant that each sample of the positive class had 100 times more importance for the network, when compared to a sample

from a healthy control. For the VGGC-A (ambulatory model) and VGGC-C (combined model), the weights were 25:1 and 50:1, respectively.

Performance Evaluation

The three neural networks described in the previous section were applied to the test set, classifying each 2 second EEG sample with a probability ranging from 0 (no IED) to 1 (IED is present). The Receiver Operating Characteristic (ROC) curves were plotted for each model, for the routine and ambulatory sets. A threshold of 0.99 was applied to the probabilities, with all EEG samples above this probability value being classified as containing an IED. This threshold was chosen based on optimization done on the training set of each network. At this probability threshold we calculated the sensitivity, specificity and false positive rate (FPR) per minute, including their 95% confidence intervals, for all three networks.

Results

The classification performance of the three networks (VGGC-R, VGGC-A and VGGC-C) on a routine and on an ambulatory EEG test set is shown in Figure 2. The shape of the curves provides information on the overall classifying potential of the networks, across thresholds. For this study, we focused on the point in the curve corresponding to a probability threshold of 0.99, displayed in Figure 2 as the blue dot and square for the routine and ambulatory EEGs, respectively.

The sensitivity, specificity and FPR at this threshold are shown in Figure 3. For the routine data, the VGGC-R model showed the highest sensitivity, yielding 84% sensitivity at 99% specificity, with a false positive rate of 0.8 FPs/min on these recordings (cf Figure 3). For the ambulatory data, the VGGC-C led to the highest sensitivity of 79% at 99% specificity, with a false positive rate of 0.4 FPs/min (cf Figure 3).

We also report the sensitivity, specificity and false positive rate of the algorithms on the normal EEGs and on the EEGs with IEDs separately in the Supplementary Material (Tables S1 to S3). The VGGC-R had a FPR of 0.3 FPs/min on normal routine EEGs and 3.2 FPs/min on normal ambulatory EEGs (see Table S1). The FPRs for the VGGC-A and VGGC-C were under 0.8 FPs/min in all the test set sub-partitions (cf Tables S2 and S3). Further, we report the results of the networks for each individual patient (Tables S4 to S9). The VGGC-R and VGGC-C detected IEDs in 17 out of 18 epilepsy patients with routine EEGs and the VGGC-A detected IEDs in 13 out of 18 of the routine EEGs with epileptiform discharges. The three networks detected IEDs in all the ambulatory EEGs containing IEDs.

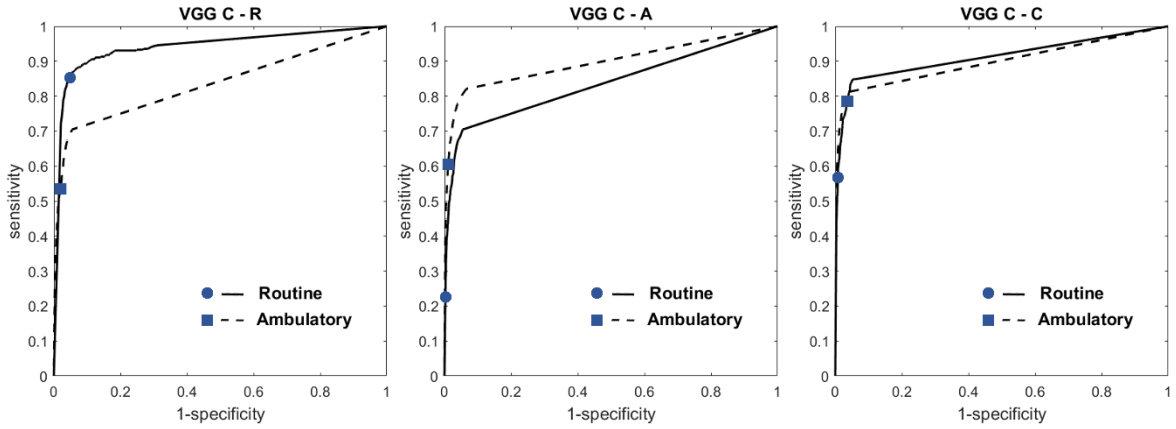

Figure 2. Receiver Operating Characteristic (ROC) curves of the neural networks on the test set with routine and ambulatory EEG data. The left panel shows the VGGC-R, trained on routine data, the middle panel shows the VGGC-A, trained on ambulatory data and the right panel shows the VGGC-C, trained on routine and ambulatory data. The blue dots show the point corresponding to a probability threshold of 0.99 on the routine EEGs and the blue squares show the point corresponding to the same threshold on the ambulatory EEGs.

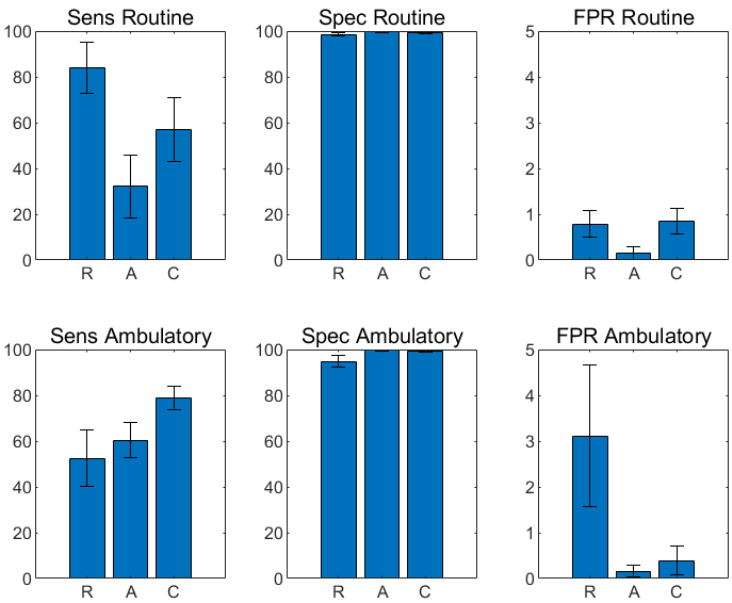

Figure 3. Performance of the of the VGGC-R (R), VGGC-A (A) and VGGC-C (C), thresholded at 0.99, on the routine test set (top) and ambulatory test set (bottom). The panels on the left show the sensitivity (Sens) of the networks, the middle panels report the specificity (Spec) and the panels on the right side illustrate the false positive rate (FPR), in false positives / minute. The 95% confidence intervals are also shown.

Discussion

EEG recordings are a key element in the (differential) diagnosis of epilepsy. Routine and ambulatory EEGs have significant differences in duration, but also in the number and type of artefacts and other patterns that can easily be mistaken as epileptic. We study the impact of the type of EEG data on the performance of deep neural networks trained for IED detection.

The three modified VGG C networks show significant differences in sensitivity (cf Figure 3) and overall disparities in performance on the routine and ambulatory test sets (cf Figures 2 and 3), with the best results being achieved by the VGGC-R on the routine test set and by the VGGC-C on the ambulatory test set.

The VGGC-R led to 84% sensitivity at 99% specificity on the routine EEGs and had a particularly low false positive rate on normal routine EEGs (0.3 FPs/min, see table S1), showing potential for clinical application. In contrast, the sensitivities of the VGGC-A and VGGC-C were under 60%, which is insufficient for clinical use. This is in line with expectations, since the VGGC-R network was trained on routine data, and thus performed better on data that was acquired on the same conditions as the training dataset.

When tested on the ambulatory recordings, the ROC curves of the VGGC-A and VGGC-C were similar (cf Figure 2). Still, at the 0.99 probability threshold, the VGGC-C showed a significantly higher sensitivity (79% at 99% specificity) for IED detection than the VGGC-A and the VGGC-R (cf Figure 3, Tables S1 to S3). The false positive rate of the VGGC-C on the ambulatory test set (0.4 FPs/min, see Figure 3) was comparable to that of the VGGC-R on the routine test set. The improved performance of the VGGC-C when compared with the VGGC-A in the ambulatory test set likely results from both networks having been exposed to ambulatory data during training, while the VGGC-C 'saw', in addition to this, more examples of IEDs from the routine EEG recordings. The larger and more representative number of training samples led to the satisfactory performance of the VGGC-C, with a high sensitivity to IEDs and

low sensitivity to artefacts and physiological sleep transients (e.g. K-complexes and vertex waves), making this algorithm potentially applicable to detect IEDs in ambulatory EEGs.

On the ambulatory test set, the VGGC-R performed poorly when compared to the VGGC-A and the VGGC–C. This shown by the low sensitivity (53%, cf Figure 3) and the high false positive rate on the ambulatory data (above 3 FPs/min for the normal and EEGs with epileptiform discharges, cf Figure 3 and Table S1). This results from the presence of artefacts and sleep phenomena in the ambulatory EEG data that were not present in the routine data used to train the VGGC–R. Apparently, this network classified such events as abnormal, flagging them as epileptiform discharges. Based on these results, the VGGC-R should not be used in a clinical setting to detect IEDs in ambulatory recordings.

Most studies concerning automated IED detection have used short EEG recordings [17-20]. From these, some algorithms were trained with a low volume of data (e.g. [18], routine EEGs from 5 patients) or used a combination of raw EEGs with additional hand-crafted features for training [19]. One study used EEGs with a median length of 53 minutes to train a convolutional neural network (SpikeNet) [17]. While the authors report only AUC and accuracy, it is possible to infer a false positive rate from the ROC curve. Taking the point where sensitivity=specificity=0.95, the false positive rate is approximately 1.67 FPs/min, assuming 2 s epochs, which is higher than the 0.78 FPs/min from our VGGC-R on the routine dataset. Another study tested an ensemble of one-dimensional convolutional networks on routine EEGs, yielding a sensitivity of 80% with 0.2 FPs/min [20]. This represents a different trade-off between sensitivity and specificity when compared to the VGGC-R, since the sensitivity of our algorithm was larger at the cost of a slightly higher false positive rate.

Other algorithms do not specify the length of the training recordings [16, 25, 33]. DeepSpike [16], a convolutional neural network embedded in the Encevis software, reported a sensitivity of 81.6% at 46.4% specificity for IED detection. The authors used non-epileptiform paroxysmal events instead of normal EEG samples as negative samples for training, which might have contributed to the low specificity of the algorithm, which is subpar to both the VGGC-R on the routine test set and the VGGC-

C on the ambulatory test set. Persyst [33], another commercially available software with an IED detection functionality, reported 43.9% sensitivity and 1.65 false detections per minute on 24h recordings. It is not clear whether this algorithm was trained on routine or ambulatory recordings, but the level of performance suggests that it may have been trained on a routine EEG dataset, only. This performance is similar to our VGGC-R tested on the ambulatory dataset, with 52.5% sensitivity and 3.11 FPs/min. In our data, the poor performance results from the mismatch between the type of training and test data, as the VGGC-R yields a much better performance when tested on routine EEGs.

To the best of our knowledge, this is the first work that specifically addresses the importance of the type of EEGs used for training deep learning algorithms for IED detection. Given the large differences in algorithm performance across data types (cf Figure 3, Tables S1 to S3), we argue that networks for IED detection should preferably be applied to similar data as the network used for training.

All of the networks described in this work were trained with EEGs from adults, which is a limitation of our methods as these algorithms cannot be successfully applied to infant recordings. In this study, we show that algorithm performance depends largely on the type of training data and the patterns that are present in that dataset. Given the differences in patterns in infant and adult EEGs, we suggest that a new network, trained on infant data, is necessary for this type of application. Further, training and testing were carried out with data from one clinical center. This means that the same level of performance is not guaranteed when applied to data from different institutions. This may result from differences in signal acquisition, as well as patient population.

Conclusion

We show that the type of EEG data has a large impact on the performance of deep learning algorithms trained for IED detection. As different networks work better for different types of data, we suggest that algorithms should be trained with the same type of EEG data they will eventually be applied to, either routine or ambulatory. In our study, we show that the VGGC-R should be used for routine recordings and the VGGC-C should be used for ambulatory recordings for the best performance in IED detection.

Conflict of Interests Disclosure

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 no conflicts of interest.

Funding Statement

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

References

1. Pillai, J. and M.R. Sperling, *Interictal EEG and the diagnosis of epilepsy.* Epilepsia, 2006. **47**: p. 14-22.
2. Tatum, W., et al., *Clinical utility of EEG in diagnosing and monitoring epilepsy in adults.* Clinical Neurophysiology, 2018. **129**(5): p. 1056-1082.
3. Smith, S.J., *EEG in the diagnosis, classification, and management of patients with epilepsy.* Journal of Neurology, Neurosurgery & Psychiatry, 2005. **76**(suppl 2): p. ii2-ii7.
4. Benbadis, S.R. and K. Lin, *Errors in EEG interpretation and misdiagnosis of epilepsy.* European neurology, 2008. **59**(5): p. 267-271.
5. Benbadis, S.R. and W.O. Tatum, *Overintepretation of EEGs and misdiagnosis of epilepsy.* Journal of clinical neurophysiology, 2003. **20**(1): p. 42-44.
6. Beniczky, S., et al., *Machine learning and wearable devices of the future.* Epilepsia, 2021. **62**: p. S116-S124.
7. Stirling, R.E., et al., *Seizure forecasting using a novel sub-scalp ultra-long term EEG monitoring system.* Frontiers in Neurology, 2021: p. 1445.
8. Weisdorf, S., et al., *Ultra-long-term subcutaneous home monitoring of epilepsy—490 days of EEG from nine patients.* Epilepsia, 2019. **60**(11): p. 2204-2214.
9. Gavvala, J., et al., *Continuous EEG monitoring: a survey of neurophysiologists and neurointensivists.* Epilepsia, 2014. **55**(11): p. 1864-1871.
10. da Silva Lourenço, C., M.C. Tjepkema-Cloostermans, and M.J. van Putten, *Machine learning for detection of interictal epileptiform discharges.* Clinical neurophysiology, 2021. **132**(7): p. 1433-1443.
11. Goelz, H., R.D. Jones, and P.J. Bones, *Wavelet analysis of transient biomedical signals and its application to detection of epileptiform activity in the EEG.* Clinical electroencephalography, 2000. **31**(4): p. 181-191.
12. Gotman, J. and P. Gloor, *Automatic recognition and quantification of interictal epileptic activity in the human scalp EEG.* Electroencephalography and clinical neurophysiology, 1976. **41**(5): p. 513-529.
13. Guler, I. and E.D. Ubeyli, *Multiclass support vector machines for EEG-signals classification.* IEEE transactions on information technology in biomedicine, 2007. **11**(2): p. 117-126.
14. Lima, C.A., A.L. Coelho, and M. Eisencraft, *Tackling EEG signal classification with least squares support vector machines: a sensitivity analysis study.* Computers in Biology and Medicine, 2010. **40**(8): p. 705-714.
15. da Silva Lourenço, C., M.C. Tjepkema-Cloostermans, and M.J. van Putten, *Efficient use of clinical EEG data for deep learning in epilepsy.* Clinical neurophysiology, 2021. **132**(6): p. 1234-1240.
16. Fürbass, F., et al., *An artificial intelligence-based EEG algorithm for detection of epileptiform EEG discharges: validation against the diagnostic gold standard.* Clinical Neurophysiology, 2020. **131**(6): p. 1174-1179.
17. Jing, J., et al., *Development of expert-level automated detection of epileptiform discharges during electroencephalogram interpretation.* JAMA neurology, 2020. **77**(1): p. 103-108.
18. Johansen, A.R., et al. *Epileptiform spike detection via convolutional neural networks*. in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2016. IEEE.
19. Prasanth, T., et al. *Deep learning for interictal epileptiform spike detection from scalp EEG frequency sub bands*. in *2020 42nd annual international conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. 2020. IEEE.
20. Thomas, J., et al., *Automated detection of interictal epileptiform discharges from scalp electroencephalograms by convolutional neural networks.* International journal of neural systems, 2020. **30**(11): p. 2050030.

21. Attia, Z.I., et al., *An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction.* The Lancet, 2019. **394**(10201): p. 861-867.
22. Esteva, A., et al., *Dermatologist-level classification of skin cancer with deep neural networks.* nature, 2017. **542**(7639): p. 115-118.
23. Acharya, U.R., Y. Hagiwara, and H. Adeli, *Automated seizure prediction.* Epilepsy & Behavior, 2018. **88**: p. 251-261.
24. Nasseri, M., et al., *Ambulatory seizure forecasting with a wrist-worn device using long-short term memory deep learning.* Scientific reports, 2021. **11**(1): p. 1-9.
25. Shoeibi, A., et al., *Epileptic seizures detection using deep learning techniques: a review.* International Journal of Environmental Research and Public Health, 2021. **18**(11): p. 5780.
26. LeCun, Y., Y. Bengio, and G. Hinton, *Deep learning.* nature, 2015. **521**(7553): p. 436-444.
27. Ruder, S., et al. *Transfer learning in natural language processing*. in *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: Tutorials*. 2019.
28. Zhuang, F., et al., *A comprehensive survey on transfer learning.* Proceedings of the IEEE, 2020. **109**(1): p. 43-76.
29. Benbadis, S.R., *What type of EEG (or EEG-video) does your patient need?* 2015, Taylor & Francis. p. 461-464.
30. Otárula, K.A.G., P. Balaguera, and S. Schuele, *Ambulatory EEG to classify the epilepsy syndrome.* Journal of Clinical Neurophysiology, 2021. **38**(2): p. 87-91.
31. Simonyan, K. and A. Zisserman, *Very deep convolutional networks for large-scale image recognition.* arXiv preprint arXiv:1409.1556, 2014.
32. Kingma, D.P. and J. Ba, *Adam: A method for stochastic optimization.* arXiv preprint arXiv:1412.6980, 2014.
33. Scheuer, M.L., A. Bagic, and S.B. Wilson, *Spike detection: Inter-reader agreement and a statistical Turing test on a large data set.* Clinical Neurophysiology, 2017. **128**(1): p. 243-250.