

## Postprocessing Deep Neural Network for Performance Improvement of Interictal Epileptiform Discharge Detection

Anguelova, Galia V.; Baines, Patricia M.

**DOI**

[10.1055/s-0045-1806851](https://doi.org/10.1055/s-0045-1806851)

**Publication date**

2025

**Document Version**

Final published version

**Published in**

International Journal of Epilepsy

**Citation (APA)**

Anguelova, G. V., & Baines, P. M. (2025). Postprocessing Deep Neural Network for Performance Improvement of Interictal Epileptiform Discharge Detection. *International Journal of Epilepsy*, 11(2), 71-77. <https://doi.org/10.1055/s-0045-1806851>

**Important note**

To cite this publication, please use the final published version (if applicable). Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.



# Postprocessing Deep Neural Network for Performance Improvement of Interictal Epileptiform Discharge Detection

Galia V. Anguelova<sup>1,2</sup>  Patricia M. Baines<sup>3</sup>

<sup>1</sup> Department of Neurology, Haaglanden Medical Centre, The Hague, The Netherlands

<sup>2</sup> Department of Clinical Neurophysiology, Stichting Epilepsie Instellingen Nederland (SEIN), Zwolle, The Netherlands

<sup>3</sup> Department of BioMechanical Engineering, Delft University of Technology, Delft, The Netherlands

**Address for correspondence** Galia V. Anguelova, MD, PhD, MSc, Stichting Epilepsie Instellingen Nederland (SEIN), Dokter Denekampweg 20, 8025 BV Zwolle, The Netherlands (e-mail: ganguelova@sein.nl).

Int J Ep

## Abstract

**Objective** Automated detection of interictal epileptiform discharges (IEDs) on electroencephalographic (EEG) data aims to reduce the time and resources spent on visual analysis by experts (the gold standard) with algorithms that match or outperform experts. In this study, we aimed to further improve IED detection performance of a deep neural network based algorithm with a simpler second-level postprocessing deep learning network, a new approach in this field.

**Materials and Methods** Seventeen interictal ambulatory EEGs were used, 15 with focal and 2 with generalized epilepsy in patients of aged 4 to 80 years (median: 19 years; 25th–75th percentile: 14–32 years). Two-second nonoverlapping epochs with a 0.99 or higher IED probability were selected by a previously developed VGG-C convolutional neural network (CNN) as input for the second-level postprocessing CNN we developed. Our CNN was tested on the resulting 580 EEG epochs after 80/20 training/validation with 3,049 epochs.

**Results** Model accuracy was 86% for the validation set and 60% for the test set. The first-level CNN selected 37% true IEDs, and with the addition of our second-level postprocessing CNN, this increased to 38%. Doubling input data of the second-level CNN, and making its architecture more complex, as well as less complex, did not improve performance.

**Conclusion** We were unable to reproduce the previously reported performance of the first-level CNN, and adding the postprocessing CNN did not improve IED detection.

## Keywords

- ▶ deep learning
- ▶ epilepsy
- ▶ EEG
- ▶ interictal epileptiform discharges
- ▶ IED

## Introduction

Interictal epileptiform discharges (IEDs) are electroencephalographic (EEG) patterns associated with an increased

likelihood of epileptic seizures.<sup>1,2</sup> The gold standard for IED detection in an EEG is visual analysis by experts<sup>3</sup>; however, this requires extensive analysis time, among several drawbacks.<sup>4</sup> For this reason computer-assisted IED

DOI <https://doi.org/10.1055/s-0045-1806851>.  
ISSN 2213-6320.

© 2025. Indian Epilepsy Society. All rights reserved.  
This is an open access article published by Thieme under the terms of the Creative Commons Attribution-NonDerivative-NonCommercial-License, permitting copying and reproduction so long as the original work is given appropriate credit. Contents may not be used for commercial purposes, or adapted, remixed, transformed or built upon. (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)  
Thieme Medical and Scientific Publishers Pvt. Ltd., A-12, 2nd Floor, Sector 2, Noida-201301 UP, India

detection with algorithms that match or outperform experts have been developed, aiming to reduce the time and resources spent on visual analysis.<sup>5</sup> Automated IED detection is complex due to IEDs' similarity to normal transients.<sup>6</sup> Several approaches have been used for automated IED detection, and one of the more recent approach is deep learning.<sup>5,7-10</sup> An advantage of deep learning is that a predefinition of IED features is not necessary.<sup>7</sup>

One prerequisite for the eventual implementation in clinical practice for such a deep neural network is a sufficiently high performance that can rival with expert visual analysis. One previous study using 50 EEGs and a deep neural network for IED detection has shown a sensitivity of 47% and a 98% specificity.<sup>7</sup> To increase performance, first, the deep neural network was made more complex and, second, the scarce IED input samples were increased through temporal shifting and using different montages, leading to a sensitivity increase from 63 to 96%, with a specificity of 99%.<sup>8</sup> Jing and colleagues have shown that their SpikeNet deep neural network can exceed expert performance, although with more resources: 9,571 EEGs.<sup>10</sup>

Another approach to increase performance is the application of a second-level consecutive deep neural network as a postprocessing step. This approach has been used in two other fields (electronics and nephrology), as a way of noise reduction in datasets, suggesting improvement in deep neural network performance,<sup>11,12</sup> but has not yet been applied in IED detection. The benefit of such a two-step approach would be that the first deep neural network can be trained to focus and perform excellently on filtering out artefacts. In the second step, other features may be more important to discern the actual IEDs, possibly improving overall performance and because the data are less noisy, possibly achieved with a shallower network and less inputs. The cognitive model behind this can be viewed as the synergy between a general practitioner and a neurologist: the general practitioner is the first-level network, filtering a different patient population for the neurologist, the second-level network, making detection of the purely neurological conditions easier. Therefore, the aim of our current study was to investigate whether such a deep learning postprocessing step improves the performance of IED detection in EEGs with limited resources.

## Materials and Methods

### EEG Data and Preprocessing

We used 17 interictal 24-hour ambulatory EEGs randomly selected from the digital database of the Medisch Spectrum Twente, in the Netherlands, which were not previously used for training of deep neural networks. All EEGs were obtained as part of routine care, and anonymized before analysis. The Medical Ethical Committee Twente waived the need for informed consent for EEG monitoring acquired as part of routine care (K24-07). The EEGs were retrospectively accessed on April 14, 2022 and July 3 and 6, 2023, and the identity of the participants was kept unknown to the authors. There were 15 EEGs with focal epilepsy and 2

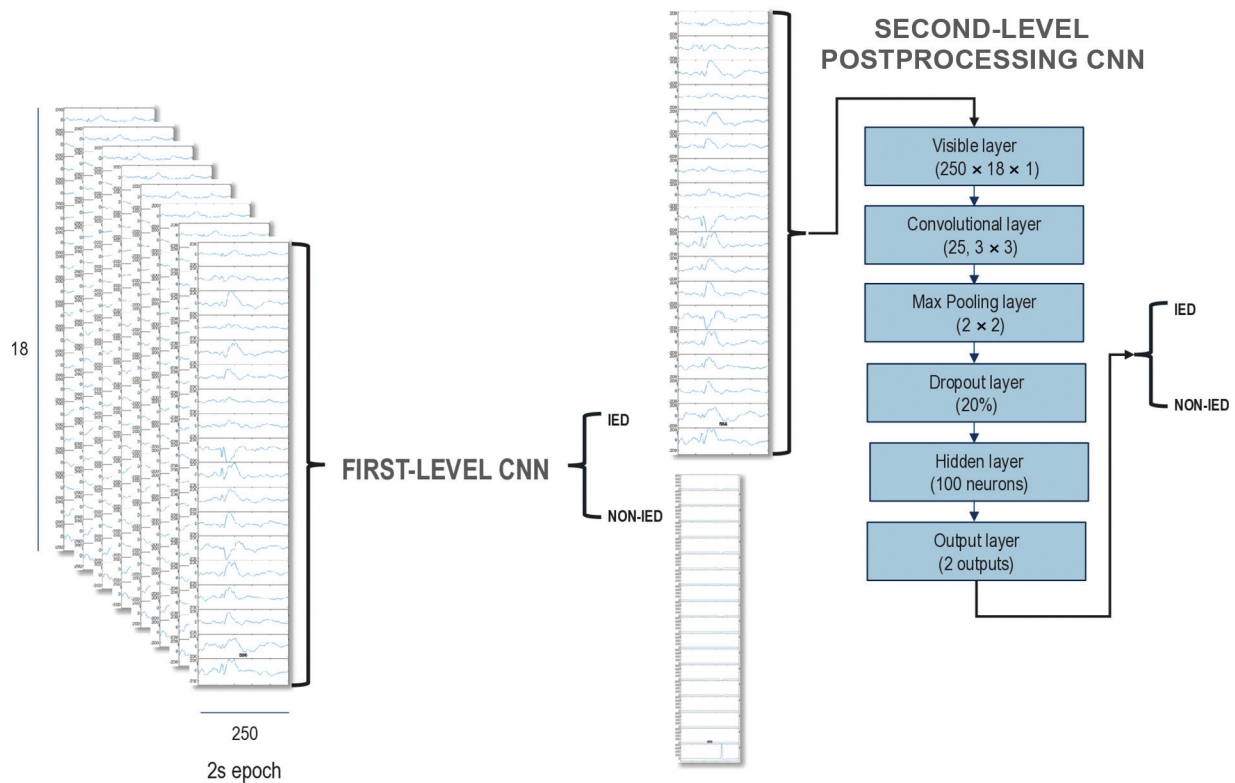
with generalized epilepsy. Patients aged 4 to 80 years, with a median of 19 years and 25th to 75th percentile of 14 to 32 years.

The EEG data were filtered in the 0.5- to 30-Hz range, down-sampled to 125 Hz, and split into nonoverlapping epochs of 2 seconds in Matlab R2021a (The MathWorks, Inc., Natick, MA, United States) to prevent using datapoints more than once, resulting in an  $18 \times 250$  matrix for each epoch in the longitudinal bipolar montage. These epochs were used as input for the VGG-C-based convolutional neural network (CNN) previously developed by da Silva Lourenço and colleagues,<sup>8</sup> which was trained on both routine and long-term ambulatory registrations, and normal as well as EEGs with IEDs. The model architecture details and prediction capabilities of the first-level CNN are reported in previous work.<sup>8</sup> The output of this deep neural network was the probability that each of the epochs contains an IED. The epochs that corresponded to an IED probability of at least 0.99 were selected. This resulted in 3,629 epochs, which were used as the input for the second postprocessing deep neural network, developed for this study. Thus, the first-level VGG-C-based CNN was used to preselect or filter the EEG epoch inputs for the second postprocessing CNN. This process is illustrated in a flowchart in ►Fig. 1.

For this second postprocessing deep neural network, we used supervised learning in order to assess its errors and improve its performance. Thus, each epoch selected from the first deep neural network was visually labeled by one of the authors (G.V.A.), assigning a score of 1 for epochs containing an IED and 0 for those not containing an IED (non-IED). This was performed in a MATLAB App developed for this purpose with a graphical user interface (GUI) as shown in ►Supplementary Fig. S1 (available in the online version). Examples of each of those epochs (IED and non-IED) are shown in ►Fig. 2. Data were divided into an 80/20 training/validation and a test set, where epochs from a particular patient were used for either training/validation or testing, resulting in 14 EEGs (3,049 epochs) for training/validation and 3 EEGs (580 epochs) for testing. The 80/20 training/validation was applied due to the limited data and its class imbalance. A total of 3,629 EEG epochs with 1,136 true IEDs was deemed a sufficient sample size based on a previous study by Cho and colleagues that suggested that 1,000 inputs per prediction class showed good performance in a deep learning CNN much more complex than ours applied on medical images.<sup>13</sup> This is also comparable to previous IED detection studies such as Tjepkema-Cloostermans and colleagues,<sup>7</sup> who used 50 EEGs, including a combination of routine 20-minute recordings and long-term ambulatory registrations, corresponding to 1,478 IEDs for their training set, an amount comparable to our 1,136 IEDs.

### Postprocessing Deep Learning Model

A two-dimensional (2D) postprocessing CNN was implemented in Python 3.10 using Keras 2.6.0, Tensorflow 2.8.0 and scikit-learn 1.0.2 (►Fig. 3). EEG epochs were used as CNN input, processed as an  $18$  (channels)  $\times$   $250$  (timepoints) matrix. The CNN applied 25 2D convolutional filters with a



**Fig. 1** Postprocessing deep learning neural network process flowchart. Second-level postprocessing deep neural network (convolutional neural network [CNN] developed in this study) for improvement of EEG interictal epileptiform discharge (IED) detection by a first-level deep neural network (VGG-C-based CNN previously developed by da Silva and colleagues).

receptive field of  $3 \times 3$  on each epoch and down-sampled the data further with a  $2 \times 2$  max pooling layer. A dropout layer of 20% was used to prevent overfitting. The data were flattened and forwarded to a hidden layer comprising 100 neurons. Stochastic optimization was performed using an Adam optimizer, with default parameters: learning rate = 0.001,  $\beta_1 = 0.9$ ;  $\beta_2 = 0.999$ , and  $\epsilon = 10^{-7}$ . We used binary cross-entropy as a loss function and a batch size of 50. The ratio of IED to non-IED epochs was 1,136:2,493 and was used as a weight factor in the model. The model provides the probability for IED presence for each epoch as output.

### Performance Evaluation

Model performance was evaluated as the accuracy for the second-level postprocessing CNN for the validation and test set, and the sensitivity and specificity for the test set using Python 3.10. The receiver operating characteristic (ROC) curve and corresponding area under the curve (AUC) were calculated using Matlab R2021a. Additionally, the percentage of epochs correctly labeled as containing an IED (positive predictive value) was calculated in the test set for the first-level VGG-C-based CNN, and the second-level postprocessing CNN, using the same IED probability threshold of 0.99.

We checked that input data for the second-level CNN were sufficient by increasing the number of input epochs from 1,483 to 3,629 (with 1,136 true IEDs). Additionally, we checked that the model architecture was optimal by determining model

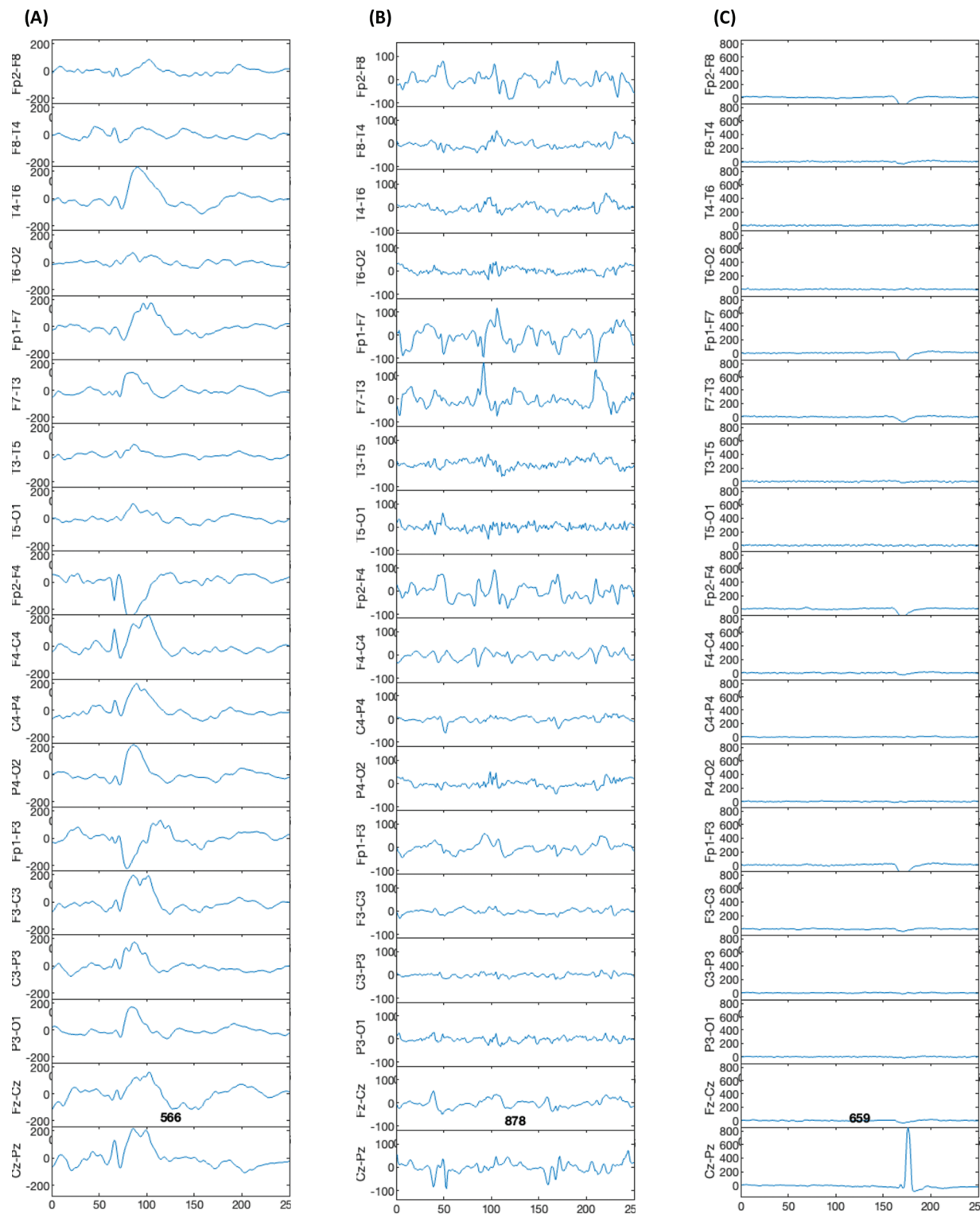
accuracy after, first, making the model more complex and, second, less complex. First, we added (1) a second convolutional or (2) a hidden dense layer with 50 neurons. Second, we (1) removed the dense layer of 100 neurons or (2) reduced the number of neurons to 20 (from 100).

The MATLAB and Python code and our dataset of selected anonymous EEG epochs are shared in a publicly available repository at the German Neuroinformatics Node/G-Node (GIN), with doi:10.12751/g-node.swrz7z.

## Results

The accuracy of the model for the validation set was 86%. The model accuracy for the test set was 60%, with a sensitivity of 0.89 and specificity of 0.11. The ROC curve is shown in **Fig. 4** with an AUC of 0.56. The percentage of epochs correctly labeled as containing an IED was 38% (10 of 26 epochs) for the second-level postprocessing CNN. This was 37% (215 of 580 epochs) in the data preselected by the first-level VGG-C-based CNN.

Doubling the number of input epochs for the second-level CNN did not improve model performance. Making the model architecture more complex by adding (1) a second convolutional or (2) a hidden dense layer with 50 neurons did not improve the performance. Making the model architecture less complex by (1) removing the dense layer of 100 neurons or (2) reducing the number of neurons to 20 (from 100) showed a deterioration of accuracy (53%).



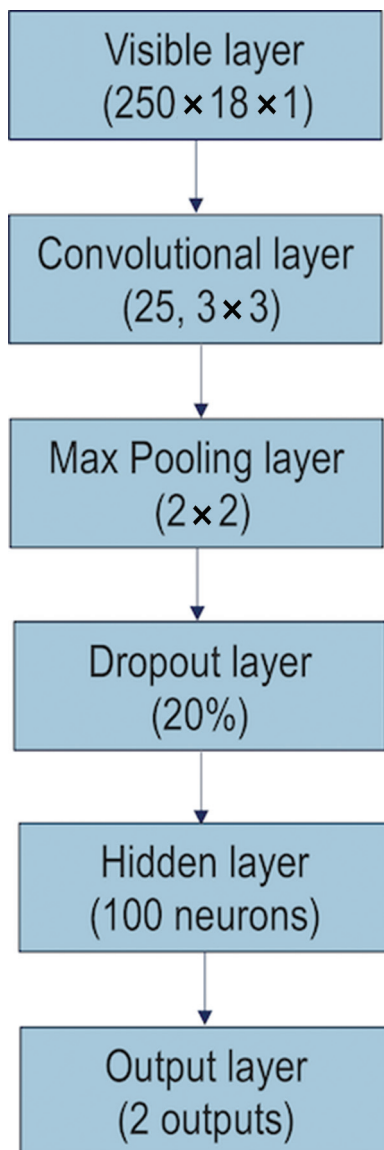
**Fig. 2** EEG epoch examples. Examples of an EEG epoch scored as (A, B) EEG interictal epileptiform discharge (IED) and as (C) not epileptiform (non-IED).

### Discussion

In summary, our major findings showed a model accuracy of 86% for the validation set and 60% for the test set. We also found that the first-level CNN selected 37% true IEDs, and after adding our second-level postprocessing CNN, this increased to 38%. In conclusion, we were unable to reproduce

the previously reported performance of the first-level CNN, and adding the postprocessing CNN did not improve IED detection, considering the model performance with insufficient specificity of 0.11.

Underperformance of a deep learning model in general can be due to (1) insufficient amount of data, (2) the quality of the data, and/or (3) underfitting or overfitting of the



**Fig. 3** Postprocessing deep learning neural network architecture. The total number of parameters is 3,150,552.

model.<sup>14</sup> First, it is unlikely that the sample size we used was insufficient because we doubled the total number of input EEG epochs without any improvement of model performance.<sup>13,14</sup>

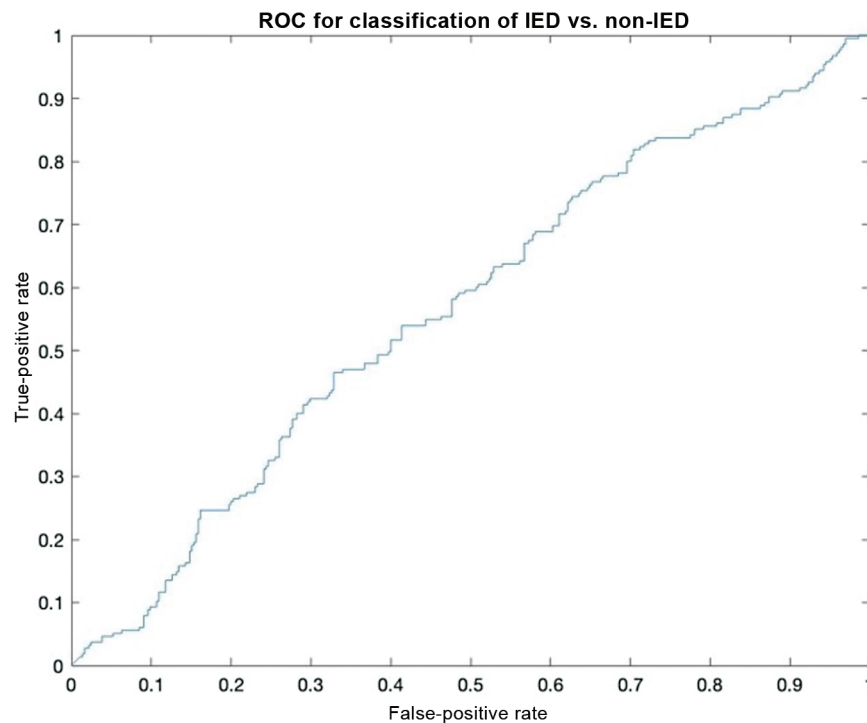
Second, the quality of our input data was mainly affected by the performance of the first-level VGG-C network. This is likely due to the limited number of two IED assessors, who scored the EEGs used for training the first-level CNN. Only 37% of the epochs that were the output of the first-level network, and consequently input for our postprocessing, were correctly labeled as containing an IED in our study. This corresponds with differences in IED interrater agreement in general, which are reported to be 49% (95% confidence interval [CI]: 37–60%).<sup>15</sup> This suggests that using a first-level network in the future that is trained with IEDs labeled by more assessors may present more generalizable and robust overall result. A limitation of our study is that the number of assessors of EEG epochs for the second-level CNN

was limited to one, and different from the two assessors of the first-level CNN. Also, assessment of IEDs was performed differently for the first- and second-level CNN: the Matlab App GUI was used for the second-level CNN with an extracted EEG epoch with a fixed montage and filter settings, whereas assessment for the first-level CNN was done in the context of the whole EEG. However, systematically scoring differently by the assessor in this study may have explained the poor positive predictive value of the first-level CNN, but not the second-level CNN. We feel it is unlikely data heterogeneity; for example, the inclusion of EEGs of adults and children would have affected performance, as both groups were included in the training of the first- and second-level CNN. Additionally, we used the leave-one-out principle and excluded the child (a 4-year-old) from our test set to see if accuracy would improve, which would be expected if children were not well represented in the training set. However, the accuracy did not change (61%).

Third, we adapted the model architecture, making it more as well as less complex, to check for under- and overfitting. Adding a second convolutional or a hidden dense layer with 50 neurons did not improve the performance on EEG data, suggesting underfitting was not the issue. Overfitting was addressed by adding a dropout layer in the model architecture. To check for overfitting due to a too complex or deep architecture, we additionally (1) removed the dense layer of 100 neurons or (2) reduced the number of neurons to 20 (from 100), both showing a deterioration of accuracy. Validation accuracy was higher than test accuracy most likely because we used the 80/20 training/validation data split. This is standard practice but implies that a part of the epochs from one patient could be in the training set and another part could be in the validation set. For the test set, we ensured that all epochs from a particular patient were only used for the test set. IEDs within the same patient are likely more similar to each other than IEDs from different patients, thus explaining the difference in model accuracy between the validation and the test set. The percentage of correctly labeled IED epochs by the first-level network was low (37%); thus, the input data for the second-level CNN were not as well filtered as expected for a postprocessing model, despite selecting a relatively high IED probability threshold (0.99). Thus, the preselected EEG data were possibly still too noisy for the limited model architecture complexity of the second-level CNN.

The application of a second-level postprocessing deep neural network has successfully been used in the field of electronics and nephrology,<sup>11,12</sup> but it is novel in the field of clinical neurophysiology and automated EEG IED detection in particular. These negative results are important to guide spending limited resources (time and EEG data) in the future. We were unable to reproduce the previous VGG-C (the first-level) network performance.

The steps to further improve IED detection rate with a postprocessing CNN approach may be achieved as follows. First, a different first-level CNN or the same CNN with an increased number of assessors can be used. Additionally, unifying the assessment methods of the first- and second-level CNN and increasing the number of assessors of



**Fig. 4** Receiver operating characteristic (ROC) curve. ROC curve of the second-level postprocessing deep neural network. IED, interictal epileptiform discharge.

the second-level CNN may contribute to overall performance. The second-level CNN architecture could be further optimized, using, for example, metaheuristic algorithms and different optimizers.<sup>16,17</sup> However, we expect this to improve only its computational burden and not performance, because we have shown that altering the architecture of the postprocessing CNN did not improve its performance.

#### Funding

None.

#### Conflict of Interest

None declared.

#### Acknowledgments

We thank Marleen C. Tjepkema-Cloostermans and Michel J.A.M. van Putten from the Department of Clinical Neurophysiology and Neurology, Medisch Spectrum Twente, and the Department of Clinical Neurophysiology, University of Twente, Enschede, The Netherlands, for their help with the conceptualization and execution of the project.

#### References

- Smith SJM, Smith S. EEG in the diagnosis, classification, and management of patients with epilepsy. *J Neurol Neurosurg Psychiatry* 2005;76(Suppl 2):ii2–ii7
- Pillai J, Sperling MR. Interictal EEG and the diagnosis of epilepsy. *Epilepsia* 2006;47(Suppl 1):14–22
- Lodder SS, Askamp J, van Putten MJAM. Computer-assisted interpretation of the EEG background pattern: a clinical evaluation. *PLoS One* 2014;9(01):e85966
- Lodder SS, van Putten MJAM. A self-adapting system for the automated detection of inter-ictal epileptiform discharges. *PLoS One* 2014;9(01):e85180
- da Silva Lourenço C, Tjepkema-Cloostermans MC, van Putten MJAM. Machine learning for detection of interictal epileptiform discharges. *Clin Neurophysiol* 2021;132(07):1433–1443
- Wilson SB, Emerson R. Spike detection: a review and comparison of algorithms. *Clin Neurophysiol* 2002;113(12):1873–1881
- Tjepkema-Cloostermans MC, de Carvalho RCV, van Putten MJAM. Deep learning for detection of focal epileptiform discharges from scalp EEG recordings. *Clin Neurophysiol* 2018;129(10):2191–2196
- da Silva Lourenço C, Tjepkema-Cloostermans MC, van Putten MJAM. Efficient use of clinical EEG data for deep learning in epilepsy. *Clin Neurophysiol* 2021;132(06):1234–1240
- Rosenberg Johansen A, Jin J, Maszczyk T, Dauwels J, Cash SS, Brandon Westover M. Epileptiform spike detection via convolutional neural networks. *Proc IEEE Int Conf Acoust Speech Signal Process* 2016:754–758
- Jing J, Sun H, Kim JA, et al. Development of expert-level automated detection of epileptiform discharges during electroencephalogram interpretation. *JAMA Neurol* 2020;77(01):103–108
- Haider A, Wei Y, Liu S, Hwang SH. Pre- and post-processing algorithms with deep learning classifier for Wi-Fi fingerprint-based indoor positioning. *Electronics (Switzerland)* 2019;8(02):195
- Marechal E, Jaugey A, Tarris G, et al. Automatic evaluation of histological prognostic factors using two consecutive convolutional neural networks on kidney samples. *Clin J Am Soc Nephrol* 2022;17(02):260–270
- Cho J, Lee K, Shin E, Choy G, Do S. How much data is needed to train a medical image deep learning system to achieve necessary high

- accuracy? Published online November 19, 2015. doi:doi.org/10.48550/arXiv.1511.06348
- 14 Aggarwal CC. *Neural Networks and Deep Learning: A Textbook*. Cham: Springer; 2023
  - 15 Jing J, Herlopian A, Karakis I, et al. Interrater reliability of experts in identifying interictal epileptiform discharges in electroencephalograms. *JAMA Neurol* 2020;77(01):49–57
  - 16 Lilhore UK, Dalal S, Simaiya S. A cognitive security framework for detecting intrusions in IoT and 5G utilizing deep learning. *Comput Secur* 2024;136:103560
  - 17 Dalal S, Kumar Lilhore U, Faujdar N, et al. Next-generation cyber attack prediction for IoT systems: leveraging multi-class SVM and optimized CHAID decision tree. *J Cloud Comput (Heidelb)* 2023; 12:137