Acta Med. Okayama, 2022
Vol. 76, No. 6, pp. 617-624
Copyright© 2022 by Okayama University Medical School.

Acta Medica Okayama
http://escholarship.lib.okayama-u.ac.jp/amo/

Original Article

Artificial Intelligence-based Detection of Epileptic Discharges from Pediatric Scalp Electroencephalograms: A Pilot Study

Katsuhiro Kobayashi*, Takashi Shibata, Hiroki Tsuchiya, and Tomoyuki Akiyama

Department of Child Neurology, Okayama University Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama 700-8558, Japan

We developed an artificial intelligence (AI) technique to identify epileptic discharges (spikes) in pediatric scalp electroencephalograms (EEGs). We built a convolutional neural network (CNN) model to automatically classify steep potential images into spikes and background activity. For the CNN model' training and validation, we examined 100 children with spikes in EEGs and another 100 without spikes. A different group of 20 children with spikes and 20 without spikes were the actual test subjects. All subjects were ≥3 to <18 years old. The accuracy, sensitivity, and specificity of the analysis were >0.97 when referential and combination EEG montages were used, and <0.97 with a bipolar montage. The correct classification of background activity in individual patients was significantly better with a referential montage than with a bipolar montage ($p=0.0107$). Receiver operating characteristic curves yielded an area under the curve >0.99, indicating high performance of the classification method. EEG patterns that interfered with correct classification included vertex sharp transients, sleep spindles, alpha rhythm, and low-amplitude ill-formed spikes in a run. Our results demonstrate that AI is a promising tool for automatically interpreting pediatric EEGs. Some avenues for improving the technique were also indicated by our findings.

Key words: neural network, deep learning, electroencephalogram, children, spike

Scalp electroencephalograms (EEGs) are indispensable for the diagnosis and management of epilepsy and other neurological disorders. However, reviewing and interpreting EEGs is time-consuming and requires years of training for epileptologists and/or electroencephalographers (EEGers). Technology that would allow the automated screening of EEGs to detect epileptic discharges or spikes would thus be useful in reducing the human labor. A number of teams have attempted to develop tools to automatically review EEGs [1-4], but none of the systems have been widely adopted in clinical practice. Against this background, various research groups have attempted to adapt artificial intelligence (AI) technologies such as artificial neural networks to medical applications [5,6], including the analysis of EEG data [7-10].

Deep learning with convolutional neural networks (CNNs) is particularly useful for image classification. Despite encouraging results, the application of AI techniques to EEG analysis is still in its infancy in terms of ictal and interictal EEG analyses [4,11-14]. In particular, the application of AI technology to the analysis of pediatric EEGs is very rare, and the reported number of children with EEG spikes in previous AI studies is very small [15]. In other studies, the subjects were limited to patients suffering from childhood epilepsy with centrotemporal spikes, a unique type of epilepsy syndrome

Received March 19, 2022; accepted May 9, 2022.
*Corresponding author. Phone: +81-86-235-7372; Fax: +81-86-235-7377
E-mail: k_koba@md.okayama-u.ac.jp (K. Kobayashi)

Conflict of Interest Disclosures: No potential conflict of interest relevant to this article was reported.

with rolandic spikes which are morphologically stereotyped and generally easy to identify [16, 17].

Pediatric EEGs differ from adult EEGs in various aspects, including age-dependent changes of background activity, patterns of epileptic discharges, and a tendency for contamination with many artifacts (*e.g.*, electrode potential jumps, muscle activity, ocular movements, and movement artifacts). We hope to develop a method to automatically detect spikes in pediatric EEGs based on deep learning with CNNs, although this might be more challenging than a comparable system for adult EEGs. The present pilot study is an initial attempt to create such a system. Our aim is to exploit the power of CNNs and apply it to image processing by converting EEG data into images and exposing them to the network, in contrast to most of the previous attempts in which time-series and/or spectral EEG data were used [12, 18-21]. EEG image data have been used with CNNs for the detection of seizures or ictal EEG patterns [14] but not for interictal EEGs, to the best of our knowledge. Herein, we investigated a simple application of an already established neural network model that may work for this purpose. We also attempted to identify which EEG montage setting is the most appropriate for this AI technique.

Subjects and Methods

Subjects. We enrolled 240 pediatric patients in the study who visited Okayama University Hospital during the period from January 2018 to March 2021 and who were ≥ 3 years and < 18 years old at the time of scalp EEG recording. Of these children, 120 showed spikes on EEGs and the remaining 120 did not, as determined through the consensus of two experienced epileptologists (explained below). Demographic data of the participants are listed in Table 1.

This study, which was based on the use of previous

Table 1 Demographic and electroencephalogram (EEG) data

	Epileptic discharges		Background	
	Training/validation	Test	Training/validation	Test
No. of participants (male/female)	100 (53/47)	20 (10/10)	100 (54/46)	20 (10/10)
Median age (range)	9.9 (3.3–17.3)	10.7 (3.3–17.5)	10.5 (3.4–17.6)	10.6 (3.0–17.9)
Disorder				
Epilepsy	88	16		
Epilepsy with seizures suppressed	6	1	54	13
Febrile seizures			6	2
Behavior disorders without epilepsy			15	1
Various neurological disorders	6	3	15	3
Suspected seizures (epilepsy unlikely)			10	1
No. of EEG data images				
Epileptic discharges	3,025⇒6,050*	640		
Background#			6,400⇒12,800*	1,280
Alpha rhythm			370⇒740*	83
Vertex sharp transients			301⇒602*	53
Sleep spindles			561⇒1,122*	50
Artifacts			437⇒874*	35

#Background data included artifacts.
*Number doubled due to addition of data with temporal jitter.

clinical records in the Department of Child Neurology, was approved by the Okayama University Ethics Committee (Ken #2109-029).

EEG recording and methods of analysis. EEGs were recorded using the Neurofax system (Nihon-Kohden, Tokyo) designed for the diagnoses of epilepsy and other neurological disorders (Table 1) with the international 10-20 electrode system and a sampling rate of 500 Hz. We extracted a 20-min interictal data segment including both waking and sleep phases from the EEG record of each patient.

As a spike is transient with pointed peak(s), the main component of which is generally negative [22], the EEG data were pre-screened for the detection of all steep negative potentials (local negative peaks with a potential gradient ≥ 25 μV within a 50-ms range and/or ≥ 50 μV within a 100-ms range at both ends) in a referential montage with reference to the average of all electrode potentials. Clear spikes were then identified using both referential and bipolar montages based on the consensus of two epileptologists.

The data of up to 32 spikes (the highest negative peak in cases of spread to multiple channels; minimal temporal separation, 3 sec) were exported as 224-by-224-pixel gray TIFF images in a three-channel referential montage alone, a two-channel bipolar montage alone, and a combination of referential and bipolar montages (process illustrated in Fig. 1A, B) from each patient with spikes. Another image was created for each selected spike with a temporal jitter of 0.5 sec (random choice of the forward or backward temporal shift) to avoid overfitting. A bandpass EEG filter (1.5-80 Hz) was used to create the images.

In each child without spikes on EEG, steep negative

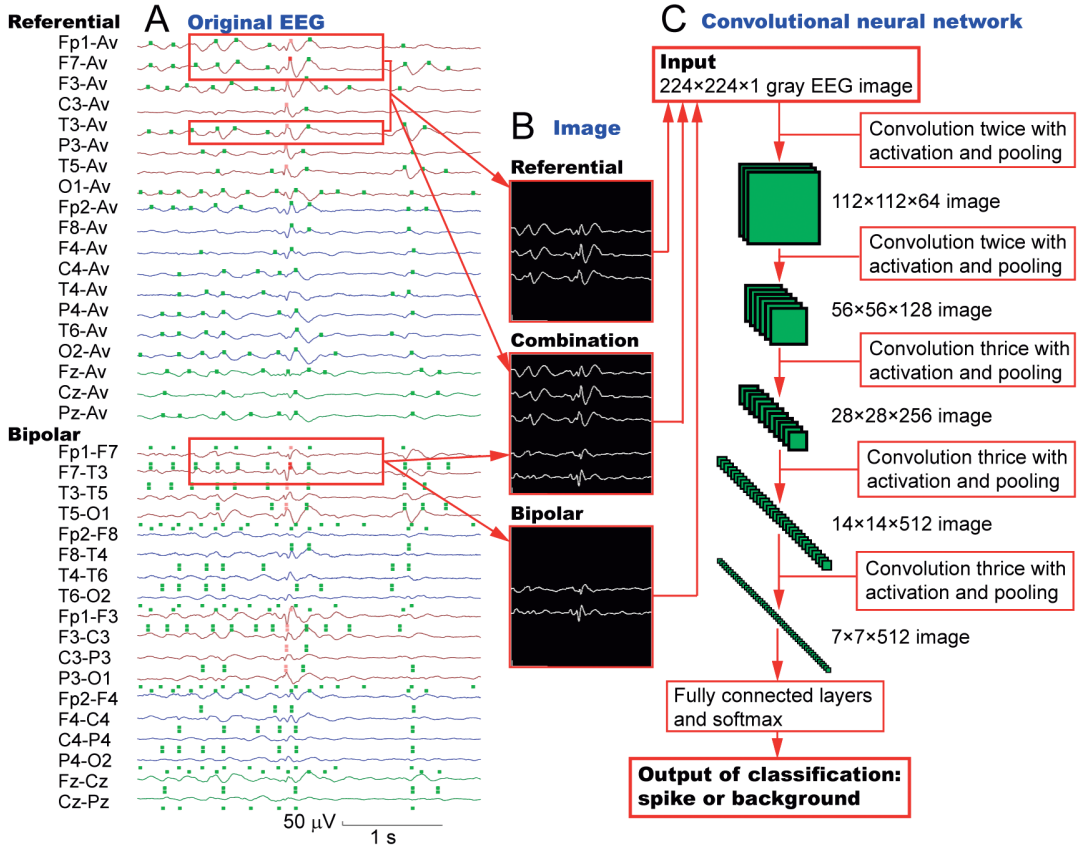

Fig. 1 Schematic illustration of the neural network used in this study. (A) Representative original EEG data in a referential montage (*top*) and a bipolar montage (*bottom*). *Green dots*: background peaks. *Red dots*: highest spike peaks. *Pink dots*: surrounding spike peaks. Parts of EEG traces including the highest spike peaks indicated by *red rectangles* were exported as images in a referential montage, a combination montage, and a bipolar montage (B). These images were processed through a convolutional neural network (CNN) for training, and the resulting output indicated whether the activity was a spike or background activity (C). Av, average reference.

potentials were similarly detected from the background, and a maximum of 64 potentials (minimal temporal separation, 3 sec) were exported as data images along with an additional identical number of jittered data images. Artifacts were not excluded, so that the network could be trained to differentiate artifact-laden background data from spikes. There were more background images than spike images because background activity generally shows an abundance of different patterns. Discrimination between spike and background data by human experts is necessary to build the present type of model (supervised machine learning) irrespective of the data format used (image, time-series, or spectral EEG). Human involvement is also necessary to test the performance of the model.

Computations were performed by a program written in-house for MATLAB (ver. 9.11 [R2021b]; MathWorks, Natick, MA, USA) with the Deep Learning Toolbox including the associated pretrained VGG-16 network model <https://www.robots.ox.ac.uk/~vgg/research/very_deep/> (accessed April, 2022), a type of CNN suited to image recognition. CNN models extract features from matrix-type data by a multi-layer application of small pattern-patches (convolution with filters) to output classification. Generally, CNN models are trained with a large amount of data to attain the optimal parameters. In this study, by application of the transfer learning method to this VGG-16 model, the network was trained on spikes from 100 patients and background data from another set of 100 children (randomly selected 20% data for validation; initial learning rate: 0.0001; number of epochs: 50; batch size: 64; activation function: Rectified Linear Unit [ReLU]) (Fig. 1C). The GPU (graphics processing unit) used was a GeForce RTX3060 (Nvidia, Santa Clara, CA, USA).

Test data were drawn from 20 patients with spikes (32 spike images per person without jitter) and 20 children without spikes (64 background images per person without jitter) who did not overlap with those used for the training/validation of the model. The cases of the patients used for gathering the test data were selected so as to have an even distribution in terms of age and sex (shown in Table 1) and to include a variety of spike types.

Accuracy (the sum of true positive [spike] and true negative [background activity] outputs of classification divided by the total number of test images), sensitivity (the number of true positive outputs of classification divided by the number of spike images for the test), specificity (the number of true negative outputs of classification divided by the number of background images for the test), and the receiver operating characteristic (ROC) curve with the associated area under the curve (AUC) were computed for each data set in the referential montage alone, the bipolar montage alone, and the combination of referential and bipolar montages. The threshold of classification was 0.5, and the optimum threshold was also computed based on the ROC curve for each montage. The Wilcoxon signed-rank test was used to compare the correct classification of spike and background images in individual patients, with Bonferroni correction for multiple comparisons; significance was defined as a p-value $< 0.05/3$ with three analyses). We used the JMP Japanese ver. 11 software (SAS Institute Japan, Tokyo) for these analyses.

Results

The accuracy, sensitivity, and specificity were each > 0.97 when the referential and combination montages were used with a threshold of 0.5, and they were each < 0.97 with the bipolar montage (Table 2). A correct classification of spikes and background activity in individual patients is shown in Fig. 2. Although montage-related differences were not observed regarding spikes, the use of the bipolar montage resulted in significantly

Table 2 Accuracy, sensitivity, and specificity of classification

	Threshold at 0.5			Optimum threshold for each montage			
Montage	Accuracy	Sensitivity	Specificity	Optimum threshold	Accuracy	Sensitivity	Specificity
Combination	0.97708	0.97969	0.97578	0.99387	0.97917	0.95781	0.98984
Referential	0.97917	0.97188	0.98281	0.93950	0.98281	0.96406	0.99219
Bipolar	0.96823	0.96719	0.96875	0.99555	0.97297	0.94219	0.98828

lower correct classification rates compared to the use of the referential montage in the case of background activity ($p = 0.0107$).

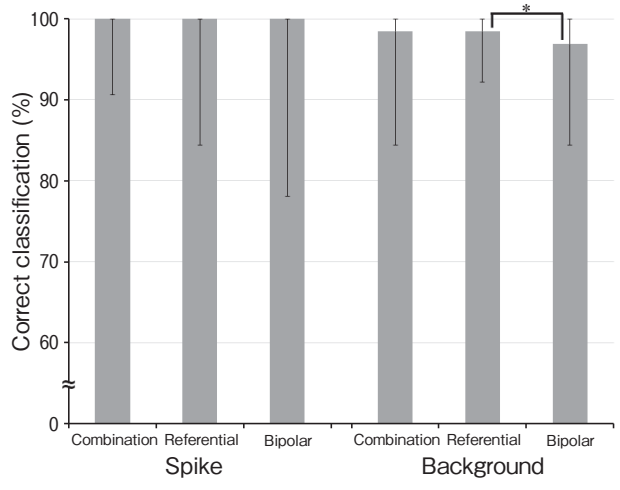

Fig. 2 Correct classification of spike and background images. The median output counts of the correct classification of spike images (true positive) and of background images (true negative) in individual patients are shown as percentages with bars indicating the ranges. While there was no montage-related difference regarding spikes, the use of the bipolar montage resulted in significantly poorer correct classification compared to the referential montage in the case of background data. $*p = 0.0107$, $(< 0.05/3)$.

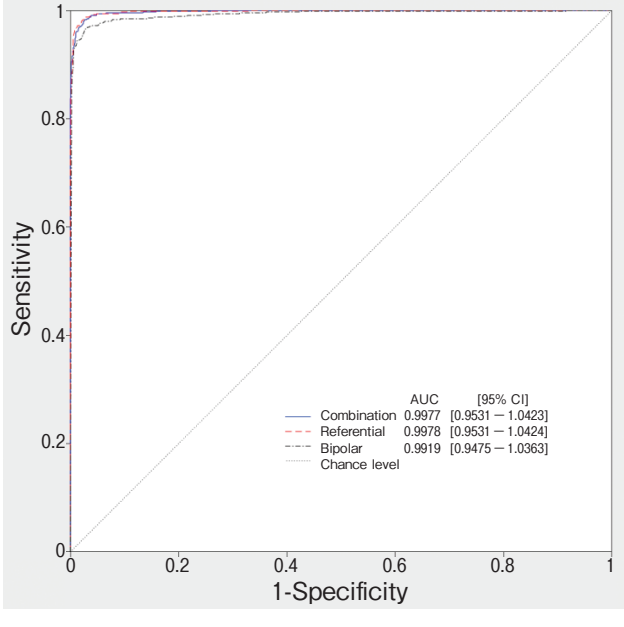

Fig. 3 Receiver operating characteristic (ROC) curve and area under the curve (AUC) of the test classification.

The AUC was generally high when the combination and referential montages were used (0.99772 and 0.99778, respectively), and it was slightly lower with the bipolar montage (0.99189) (Fig. 3). When the optimum threshold obtained from the ROC curve was used for each montage, the accuracy and specificity improved slightly, but the sensitivity worsened (Table 2).

There were a total of 30 false-negative outputs of classification (13, 18, and 21 outputs of misclassification as background with overlap in the analyses of the combination, referential, and bipolar montages, respectively, among 640 assessments of spike images) and 59 false-positive outputs of classification (31, 22, and 40 outputs of misclassification as spikes with overlap in the analyses of the combination, referential, and bipolar montages, respectively, among 1,280 assessments of background images) with a threshold of 0.5. Correct spike classification tended to fail in cases of low-amplitude ill-formed spikes, particularly epileptic discharges with a rather long duration (termed "sharp waves") in a run (Fig. 4A) that accounted for 25 false-negative outputs of classification. Misclassification of the remaining five spike images appeared to be due to noise contamination.

Correct background classification also failed in cases of slow waves, particularly runs of slow waves with rather steep morphology (Fig. 4B), which accounted for 39 false-positive outputs of classification. Misclassification of the background occurred in two images, which included sleep spindles associated with slow waves (Fig. 4C), seven images including vertex sharp transients (Fig. 4D), seven images including alpha activity with rather deformed/irregular morphology associated with slow waves and/or noise, and four images including artifacts.

The time needed to train the network ranged from 2,578 to 2,591 min, and the mean time to test 1,000 images was 28.6 sec.

Discussion

We used an established CNN model to recognize spikes in pediatric EEGs without any additional data processing, and we obtained acceptable results. The accuracy, sensitivity, specificity, and AUC obtained with our method were generally high, and they were largely comparable to the results of various other AI studies involving adults [19], a combination of adults

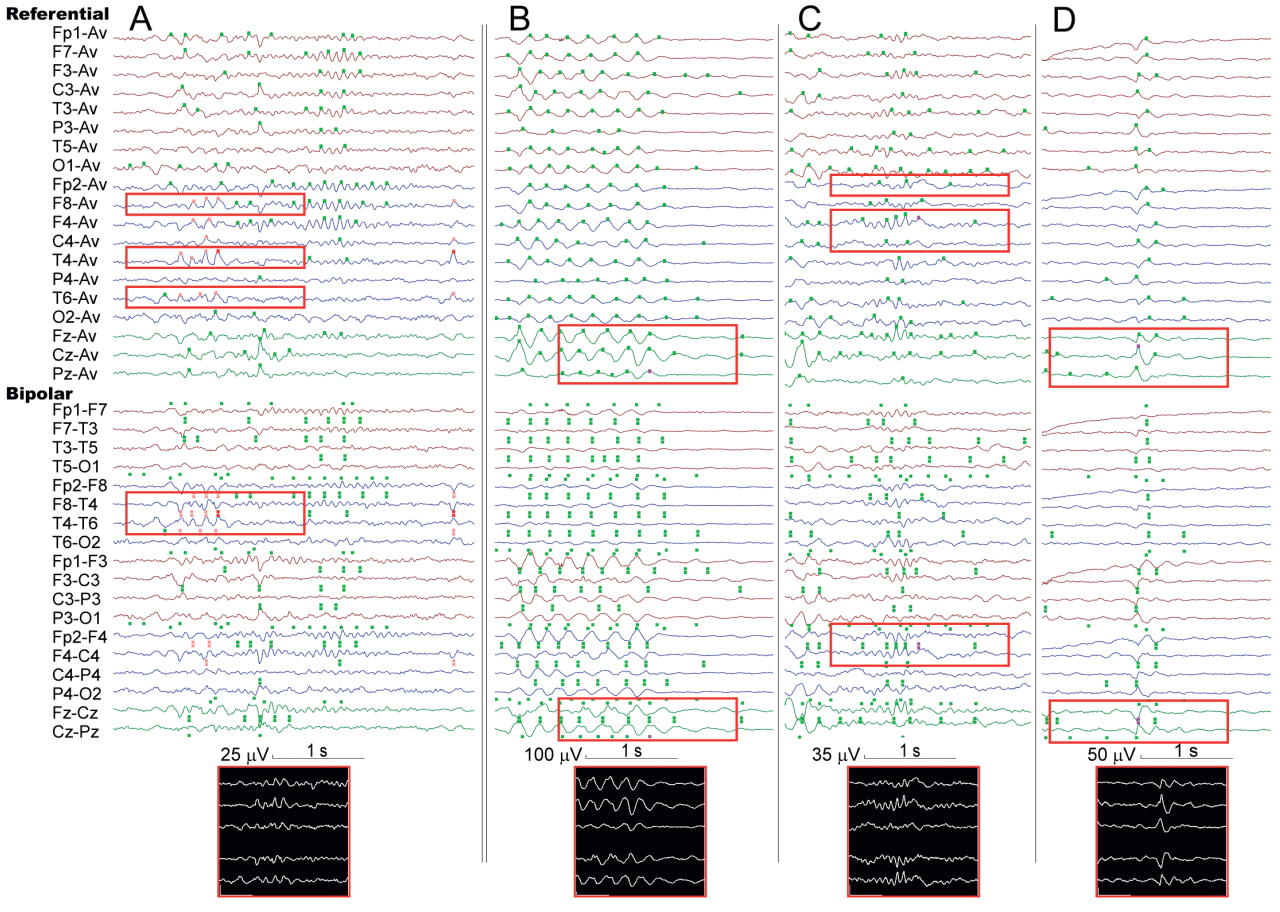

Fig. 4 Representative false-negative and false-positive classifications. (**A**) Mis-classification of a run of sharp waves over T4 as the background. Mis-classification of background grapho-elements as spikes. (**B**) slow waves in a burst, (**C**) sleep spindles associated with slow waves, and (**D**) a vertex sharp transient. Purple dots: background peaks that happened to be selected for the test. *Green dots*: background peaks. *Red dots*: highest spike peaks. *Pink dots*: surrounding spike peaks.

and children [20,21], children [16,17], and participants of unknown ages [11,18]. It is notable that Jing *et al.* reported expert-level performance of their network model (SpikeNet) [23]. A straightforward comparison of performances among these proposed methods is difficult because the EEG data used are different. In most of the previous studies, EEG data in the form of time-series numerical values were used. Because EEG data are a time series of potential values, it may be mathematically more natural and precise to use time-series data than to use image-transformed data. However, our present method based on an image data analysis may have an under-recognized and attractive feature: it has potential similarity to human EEG reviewing, as epileptologists and EEGers see EEGs as images of traces for pattern recognition and not as runs of numerical values. When the present model is completed, both the process of picking out candidate steep potentials from EEG data and the spike/background classification can be fully automated, as is the case with other AI methods.

Epileptologists and EEGers learn to tell spikes from steep potentials buried in background activity through experience; however, they do not know exactly what is the essential morphological difference between spikes and background steep potentials, and therefore the mathematical formulation of this differentiation remains to be established. As in many EEG studies, we needed the consensus of experienced epileptologists to select spikes — due to the lack of a formal objective definition of spikes — regardless of the data format used. CNNs are thought to model real neural responses in the

primary visual cortex [24,25], although the processing in CNNs is largely a black box. If a CNN model really works for EEG reviewing and the reasons for such success are identified in the future, then our new model might provide some clue to what human eyes recognize in the morphology of EEGs.

The present results provide several clues suggesting how our method can be improved. First, regarding montage selection, the use of a referential or combination montage gave better results than the use of the bipolar montage. A subtraction of potential values between adjacent EEG channels in a referential montage yields potential values in a bipolar montage. Therefore, all data in a bipolar montage are included in a referential montage, and the data in a combination montage include duplicate information. Image panels in a referential montage may contain EEG data information most efficiently in a limited area compared to other montages. In childhood, however, the EEGs in a referential montage are generally prone to deformation due to artifacts, and both referential and bipolar montages are often necessary for precise interpretation by humans. EEG images in only a bipolar montage may not be ideal, but further research is necessary to determine whether a referential montage is better than a combination montage.

Second, there is a trade-off between the sensitivity and specificity of the analysis, and this is affected by the threshold selection. In this study, the background data appeared to have more influence on the model than the spike data, probably because the number of background images was double that of spike images. Consequently, our selection of the optimum threshold led to decreased sensitivity (true spike classification) and increased specificity (true background classification) compared to the initial threshold of 0.5. The selection of the threshold may depend on subjective decisions regarding the relative importance of spikes and background.

Third, the misclassified background EEG patterns that we observed included vertex sharp transients, sleep spindles, and alpha rhythms, which have a specific mode of appearance and distribution: vertex sharp transients occur dominantly over the vertex (Cz) during non-rapid eye movement (NREM) sleep; sleep spindles emerge over the bilateral frontal and/or central regions during NREM sleep, and alpha rhythms appear dominantly over the bilateral occipital regions when the subject closes his or her eyes during a waking state. We may thus build separate models that are designed for specific regions and incorporate arousal-level information.

In addition, various types of EEG grapho-elements other than spikes might well be differentiated through the network model. Although an attempt to include information on age and sleep stage has been described for a model of adult EEG data [19], such an attempt may be challenging for pediatric EEG data because children often keep their eyes open (with blocking of alpha rhythms) during wakefulness, which can make the staging challenging. In addition, such region-, stage- and grapho-element-specific models would require an enormous amount of EEG data for model training.

Fourth, epileptologists and EEGers generally see whole EEGs in both temporal and spatial aspects to identify various patterns, which makes brief single/few-channel traces insufficient for review. Some spikes are difficult to recognize even by human eyes when EEG data from surrounding/contralateral regions are missing. There are models that employ all-channel time-series EEG data [11], but the inclusion of all-channel data in an image would expand its size and might render it somewhat difficult to spot the location of an abnormality. More investigation is needed in this regard.

There are several study limitations to address. We pre-selected steep negative potentials as candidate spikes, for convenience. Spikes with very low amplitude may thus have been missed, and spikes with a predominant positive component may seldom be observed [22]. There are reports of methods to pre-select candidate spikes that would be more sophisticated than our simple method [15,16], plus methods that do not involve pre-selection [11]. Our present analyses did not include data from infants or young children (< 3 years of age) with immature EEG patterns, which may require a different scheme. The time consumed for modeling and testing may be improved if a program written in Python is used with a better GPU. We selected the VGG-16 over other types of CNNs as it allows a relatively large image size, but there might be another CNN model that works better for EEG analyses. Our selection of the patients for tests depended on human decisions and might thus have included some bias. We need to involve many more patients not only from our institution (Okayama University Hospital) but also from other hospitals to improve the efficacy of our

method.

There is much room to improve the methodology, but the results of the present pilot study demonstrate the possibility that pediatric EEGs can be interpreted with AI using currently available CNN models. We hope that our model will become a useful clinical screening tool to reduce human labor in reviewing scalp EEGs.

Acknowledgments. K. Kobayashi was supported by a Grant-in-Aid from the Ministry of Education, Culture, Sports, Science and Technology, Japan (JSPS KAKENHI, no. 21K07754) and by a Health and Labour Research Grant from the Ministry of Health, Labour and Welfare, Japan (no. JPMH20FC1039).

References

1. Gotman J and Wang LY: State dependent spike detection: validation. Electroencephalogr Clin Neurophysiol (1992) 83: 12–18.
2. Kurth C, Gilliam F and Steinhoff BJ: EEG spike detection with a Kohonen feature map. Ann Biomed Eng (2000) 28: 1362–1369.
3. Lodder SS, Askamp J and van Putten MJAM: Inter-ictal spike detection using a database of smart templates. Clin Neurophysiol (2013) 124: 2328–2335.
4. da Silva Lourenço C, Tjepkema-Cloostermans MC and van Putten MJAM: Machine learning for detection of interictal epileptiform discharges. Clin Neurophysiol (2021) 132: 1433–1443.
5. Topol EJ: High-performance medicine: the convergence of human and artificial intelligence. Nat Med (2019) 25: 44–56.
6. Abbasi B and Goldenholz DM: Machine learning applications in epilepsy. Epilepsia (2019) 60: 2037–2047.
7. Gabor AJ and Seyal M: Automated interictal EEG spike detection using artificial neural networks. Electroencephalogr Clin Neurophysiol (1992) 83: 271–280.
8. Patnaik LM and Manyam OK: Epileptic EEG detection using neural networks and post-classification. Comput Methods Programs Biomed (2008) 91: 100–109.
9. Wilson SB, Turner CA, Emerson RG and Scheuer ML: Spike detection II: automatic, perception-based detection and clustering. Clin Neurophysiol (1999) 110: 404–411.
10. James CJ, Jones RD, Bones PJ and Carroll G: Detection of epileptiform discharges in the EEG by a hybrid system comprising mimetic, self-organized artificial neural network, and fuzzy logic stages. Clin Neurophysiol (1999) 110: 2049–2063.
11. da Silva Lourenço C, Tjepkema-Cloostermans MC and van Putten MJAM: Efficient use of clinical EEG data for deep learning in epilepsy. Clin Neurophysiol (2021) 132: 1234–1240.
12. Clarke S, Karoly PJ, Nurse E, Seneviratne U, Taylor J, Knight-Sadler R, Kerr R, Moore B, Hennessy P, Mendis D, Lim C, Miles J, Cook M, Freestone DR and D'Souza W: Computer-assisted EEG diagnostic review for idiopathic generalized epilepsy. Epilepsy Behav (2021) 121(Pt B): 106556.
13. Acharya UR, Oh SL, Hagiwara Y, Tan JH and Adeli H: Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. Comput Biol Med (2018) 100: 270–278.
14. Emami A, Kunii N, Matsuo T, Shinozaki T, Kawai K and Takahashi H: Seizure detection by convolutional neural network-based analysis of scalp electroencephalography plot images. Neuroimage Clin (2019) 22: 101684.
15. Mera-Gaona M, López DM, Vargas-Canas R and Miño M: Epileptic spikes detector in pediatric EEG based on matched filters and neural networks. Brain Inform (2020) 7: 4.
16. Fukumori K, Yoshida N, Sugano H, Nakajima M and Tanaka T: Epileptic spike detection using neural networks with linear-phase convolutions. IEEE J Biomed Health Inform (2022) 26: 1045–1056.
17. Xu Z, Wang T, Cao J, Bao Z, Jiang T and Gao F: BECT spike detection based on novel EEG sequence features and LSTM algorithms. IEEE Trans Neural Syst Rehabil Eng (2021) 29: 1734–1743.
18. Prasanth T, Thomas J, Yuvaraj R, Jing J, Cash SS, Chaudhari R, Leng TY, Rathakrishnan R, Rohit S, Saini V, Westover BM and Dauwels J: Deep Learning for Interictal Epileptiform Spike Detection from scalp EEG frequency sub bands. Annu Int Conf IEEE Eng Med Biol Soc (2020) 3703–3706.
19. van Leeuwen KG, Sun H, Tabaeizadeh M, Struck AF, van Putten MJAM and Westover MB: Detecting abnormal electroencephalograms using deep convolutional networks. Clin Neurophysiol (2019) 130: 77–84.
20. Fürbass F, Kural MA, Gritsch G, Hartmann M, Kluge T and Beniczky S: An artificial intelligence-based EEG algorithm for detection of epileptiform EEG discharges: Validation against the diagnostic gold standard. Clin Neurophysiol (2020) 131: 1174–1179.
21. Tjepkema-Cloostermans MC, de Carvalho RCV and van Putten MJAM: Deep learning for detection of focal epileptiform discharges from scalp EEG recordings. Clin Neurophysiol (2018) 129: 2191–2196.
22. Niedermeyer E: Abnormal EEG patterns: epileptic and paroxysmal; in Electroencephalography. Basic Principles, Clinical Applications, and Related Fields, Niedermeyer E and Lopes da Silva F eds, 5th Ed, Lippincott Williams & Wilkins, Philadelphia (2005) pp 255–280.
23. Jing J, Sun H, Kim JA, Herlopian A, Karakis I, Ng M, Halford JJ, Maus D, Chan F, Dolatshahi M, Muniz C, Chu C, Sacca V, Pathmanathan J, Ge W, Dauwels J, Lam A, Cole AJ, Cash SS and Westover MB: Development of expert-level automated detection of epileptiform discharges during electroencephalogram interpretation. JAMA Neurol (2020) 77: 103–108.
24. Cadena SA, Denfield GH, Walker EY, Gatys LA, Tolias AS, Bethge M and Ecker AS: Deep convolutional models improve predictions of macaque V1 responses to natural images. PLoS Comput Biol (2019) 15: e1006897.
25. Zhang Y, Lee TS, Li M, Liu F and Tang S: Convolutional neural network models of V1 responses to complex patterns. J Comput Neurosci (2019) 46: 33–54.