Epilepsy Research 178 (2021) 106815

Contents lists available at ScienceDirect

Epilepsy Research

journal homepage: www.elsevier.com/locate/epilepsyres

Deep learning-based diagnosis of temporal lobe epilepsy associated with hippocampal sclerosis: An MRI study

Yosuke Ito [a,*], Masafumi Fukuda [a], Hitoshi Matsuzawa [b,g], Hiroshi Masuda [a], Yu Kobayashi [c], Naoya Hasegawa [d], Hiroki Kitaura [e], Akiyoshi Kakita [e], Yukihiko Fujii [f]

[a] Department of Functional Neurosurgery, Epilepsy Center, NHO Nishiniigata Chuo Hospital, Japan
[b] Center for Integrated Human Brain Science, Niigata University, Japan
[c] Department of Child Neurology, Epilepsy Center, NHO Nishiniigata Chuo Hospital, Japan
[d] Department of Epilepsy, Epilepsy Center, NHO Nishiniigata Chuo Hospital, Japan
[e] Department of Pathology, Brain Research Institute, Niigata University, Japan
[f] Department of Neurosurgery, Brain Research Institute, Niigata University, Japan
[g] Center for Advanced Medicine and Clinical Research, Kashiwaba Neurosurgical Hospital

ARTICLE INFO

Keywords:
Mesial temporal lobe epilepsy
Machine learning
Convolutional neural network
Temporal lobe epilepsy with hippocampal sclerosis
Fine-tuning
Artificial intelligence

ABSTRACT

Purpose: The currently available indicators—sensitivity and specificity of expert radiological evaluation of MRIs—to identify mesial temporal lobe epilepsy (MTLE) associated with hippocampal sclerosis (HS) are deficient, as they cannot be easily assessed. We developed and investigated the use of a novel convolutional neural network trained on preoperative MRIs to aid diagnosis of these conditions.

Subjects and methods: We enrolled 141 individuals: 85 with clinically diagnosed mesial temporal lobe epilepsy (MTLE) and hippocampal sclerosis International League Against Epilepsy (HS ILAE) type 1 who had undergone anterior temporal lobe hippocampectomy were assigned to the MTLE-HS group, and 56 epilepsy clinic outpatients diagnosed as nonepileptic were assigned to the normal group. We fine-tuned a modified CNN (mCNN) to classify the fully connected layers of ImageNet-pretrained VGG16 network models into the MTLE-HS and control groups. MTLE-HS was diagnosed using MRI both by the fine-tuned mCNN and epilepsy specialists. Their performances were compared.

Results: The fine-tuned mCNN achieved excellent diagnostic performance, including 91.1% [85%, 96%] mean sensitivity and 83.5% [75%, 91%] mean specificity. The area under the resulting receiver operating characteristic curve was 0.94 [0.90, 0.98] (DeLong's method). Expert interpretation of the same image data achieved a mean sensitivity of 73.1% [65%, 82%] and specificity of 66.3% [50%, 82%]. These confidence intervals were located entirely under the receiver operating characteristic curve of the fine-tuned mCNN.

Conclusions: Deep learning-based diagnosis of MTLE-HS from preoperative MR images using our fine-tuned mCNN achieved a performance superior to the visual interpretation by epilepsy specialists. Our model could serve as a useful preoperative diagnostic tool for ascertaining hippocampal atrophy in patients with MTLE.

1. Introduction

The classic presentation of a mesial temporal lobe epilepsy (MTLE) seizure is initially characterized by feelings of fear and anxiety associated with a rising sensation in the chest, followed by the onset of motor symptoms and impairment of consciousness, which takes several minutes to be restored (Wieser and Epilepsy, 2004). Cognitive dysfunction and/or psychiatric symptoms may also co-occur (Wieser and Epilepsy, 2004). Most patients with MTLE can become resistant to treatment with antiepileptic drugs (Semah et al., 1998). In comparison, surgical resection has reportedly eliminating seizures in 70–80% of patients with MTLE. In these patients, the resection specimens show

Abbreviations: ATL, anterior temporal lobectomy; AUC, area under the ROC curve; CI, confidence interval; CNN, convolutional neural network; FLAIR, fluid-attenuated inversion recovery; HS, hippocampal sclerosis;; ILAE, International League Against Epilepsy; mCNN, modified CNN; MRI, magnetic resonance imaging; MTLE, mesial temporal lobe epilepsy; ROC, receiver operating characteristic; SVM, support vector machine; TLE, temporal lobe epilepsy.

* Corresponding author.
E-mail address: y0226@kc4.so-net.ne.jp (Y. Ito).

https://doi.org/10.1016/j.eplepsyres.2021.106815
Received 28 July 2021; Received in revised form 18 October 2021; Accepted 10 November 2021
Available online 21 November 2021

histopathological features of hippocampal sclerosis (HS) (Thom et al., 2010; Wiebe et al., 2001). Determining whether a patient should be diagnosed with mesial temporal lobe epilepsy with HS (MTLE-HS) is crucial to selecting the optimal treatment plan (Cersósimo et al., 2011). In the 2010 International League Against Epilepsy (ILAE) classification system, mesial temporal lobe epilepsy with hippocampal sclerosis (MTLE-HS) was classified as a clear, specific syndrome (Berg et al., 2010). In the 2017 ILAE epilepsy type classification, it was classified as a focal epilepsy (Scheffer et al., 2017). The ILAE Special Committee is considering classifying it as an epilepsy syndrome that develops at various ages. It has been stated that diagnostic imaging is important because MTLE-HS is often drug-resistant but can be placed in complete remission by epilepsy surgery (Riney, 2021a, 2021b).

Brain magnetic resonance imaging (MRI) is the most useful imaging modality for detecting HS (Berkovic et al., 1991a, 1991b; Jackson et al., 1990); patients tend to have better seizure outcomes after a resection when hippocampal atrophy is confirmed to be unilateral on preoperative MR images (Jutila et al., 2002). However, radiologists may not always make an accurate interpretation by visual inspection performed using an epilepsy-specific standard protocol. One study found that 86% of affected cases were overlooked (Von Oertzen et al., 2002). Moreover, scans acquired from healthy individuals can exhibit features very similar to the pathological hallmarks of MTLE-HS, which makes it challenging to diagnose this disorder in some cases: e.g., hippocampal signal abnormalities are detectable on fluid-attenuated inversion recovery (FLAIR) and T2-weighted images in approximately one-third of healthy adults (Labate et al., 2010), as is a unilateral enlargement of the temporal horn (Menzler et al., 2010). Different techniques for diagnosing MTLE-HS have been investigated that do not involve human interpretation. Some of these methods use hippocampal volume and voxel-based morphometry, while others utilize T2 relaxation time (Coan et al., 2014; Hakimi et al., 2019; Mueller et al., 2012; Riederer et al., 2020). However, such quantitative methods are not generally applied in routine practice because they require preprocessing steps, such as segmenting and statistically processing the MR images or defining the regions of interest in which to measure the T2 relaxation time.

Deep learning is a form of artificial intelligence that is especially effective at classifying and recognizing images, which has recently led many research groups to experiment with the clinical applications of this technology (Ehteshami Bejnordi et al., 2017; Esteva et al., 2017; Gulshan et al., 2016). Convolutional neural networks (CNNs) are a family of deep learning algorithms that owe their superior classification performance to their ability to extract key visual features beneficial for classification in the learning process using a training dataset (LeCun et al., 2015).

In this study, an ImageNet-pretrained VGG-16 network was modified and then fine-tuned on training data consisting of preoperative MR images from confirmed MTLE-HS patients, in addition to control MRIs acquired from healthy subjects. For fine-tuning, the modified CNN (mCNN) was initialized with the pretrained (VGG-16) weights and configured to allow all nodes to be updated during training. The diagnostic performance of the fine-tuned CNN was then validated using MR images not included in the training dataset. Finally, to investigate the utility and limitations of deep learning-based imaging diagnostics, the model's performance was compared with that of epilepsy specialists interpreting the same images.

2. Subjects and methods

2.1. Study population

The study protocol was approved by the Ethics Committee of NHO Nishiniigata Chuo Hospital. The study was conducted in accordance with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. In this study, we implemented the opt-out method of consent for the participants, which guaranteed them the opportunity to decline to have their diagnostic information publicized as part of the study.

2.2. Patients with MTLE-HS

From the database of patients who had undergone surgery at the NHO Nishiniigata Chuo Hospital between 2000 and 2019, we selected individuals who had been diagnosed with temporal lobe epilepsy based on their clinical symptoms, as well as EEG and image data. Of the selected patients, we excluded those whose MRI showed gross lesions such as tumors, cerebrovascular malformation, cerebral dysplasia, postencephalitic changes, and traumatic changes. Of the patients with temporal lobe epilepsy whose MRI did not show gross lesions, we selected 114 who had undergone anterior temporal lobe hippocampectomy, and whose resected hippocampal specimens allowed HS International League Against Epilepsy (ILAE) classifications (Fig. 1).

Intracranial electroencephalography electrodes had been implanted in 41 patients—including 14 cases of bilateral placement—to enable long-term electroencephalographic recordings. Each patient had undergone an anterior temporal lobectomy (ATL; 51 left, 63 right) in which 3.0–5.0 cm of tissue behind the temporal pole was surgically removed from the affected hemisphere. The resected specimens—which contained the anterior temporal lobe, hippocampus, and parts of the amygdala—were examined by histopathologists in the pathology department of Niigata University's Brain Research Institute. The histopathologists then graded the extent of HS present according to the international consensus classification scheme established by the ILAE (Blumcke et al., 2013) (Supplementary Table S1). HS ILAE type 1 was the most frequent histopathology (85/114 cases [74.5%]), whereas the grades HS ILAE type 2, HS ILAE type 3, and no HS/gliosis only were assigned in eight (7.0%), four (3.5%), and 17 cases (15%), respectively.

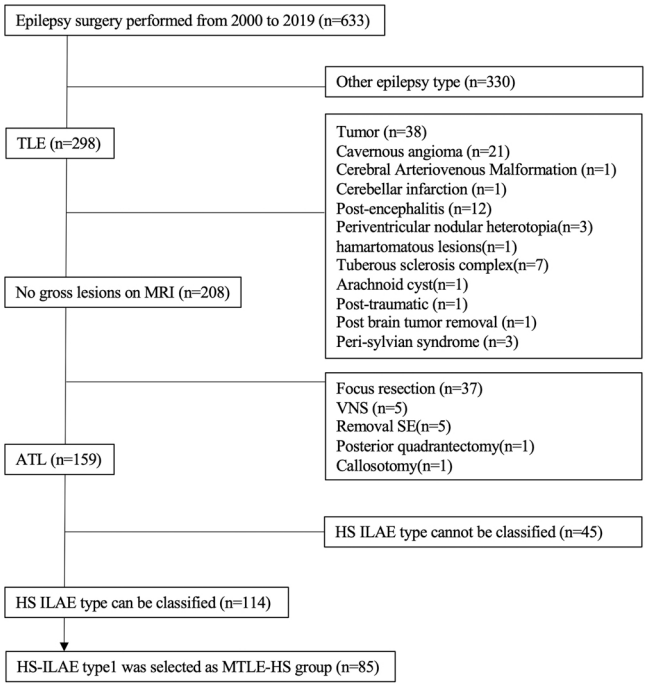

Fig. 1. Patient flowchart. In total, 633 patients underwent epilepsy surgery at Nishi-Niigata Chuo National Hospital from 2000 to 2019. Over 20% were cases of MTLE without gross lesions on MRI (n = 210); ATL was the specific procedure in nearly all of these patients (n = 159). Finally, 114 of these patients whose resected tissue samples could be graded by HS ILAE type were selected and 85 case of ILAE type1 analyzed as the MTLE-HS group. ATL, anterior temporal lobectomy; HS, hippocampal sclerosis; ILAE, International League Against Epilepsy; MRI, magnetic resonance imaging; MTLE, mesial temporal lobe epilepsy; VNS, Vagus Nerve Stimulation; SE, Setting Electrode.

Postoperative seizure outcomes were assessed 1 year after ATL using the Engel Epilepsy Surgery Outcome Scale (Engel, 1993) (Supplementary Table S2). Ratings of Engel class 1, 2, and 3 were achieved in 88, seven, and 17 cases, respectively; an assessment was not possible in two cases. As shown in Fig. 2, cases rated as HS ILAE type 1 based on hippocampal pathology in the resected specimen had significantly better seizure outcomes than did those of other grades (Pearson's chi(1)= 5.4, $p = 0.02$).

Only patients with HS ILAE type 1 were included in the datasets used for subsequent model training and validation. Corresponding to "classic" HS, this pathology accounted for the majority of cases herein reviewed. The resulting group (MTLE-HS, n = 85) had a mean age of 30.5 (range: 11–64) years and a mean age at disease onset of 13.2 (2−53) years; 43 patients were female, and 42 were male. The mean follow-up period was 17.3 years. One case had a history of meningitis. Only MR images taken within 2 years of the procedure were used in this study.

2.3. Healthy control subjects

This study's control group consisted of individuals conclusively determined by an epilepsy specialist to be healthy subjects subsequent to comprehensive assessments (blood work, urinalysis, electrocardiogram, MRI) at the epilepsy clinic of NishiNiigata Chuo National Hospital from 2013 to 2018. This group ("Control" n = 56) had a mean age of 26.3 (13−38) years; 27 subjects were male, and 29 were female. No significant differences were observed in age or sex distributions between the MTLE-HS and Control groups (age: t (139) = 1.96, $p = 0.0515$; sex: Pearson's chi (1) = 0, $p > 0.99$).

2.4. MR image acquisition

This study exclusively utilized axial FLAIR images acquired using 1.5-T MRI scanners: a Magnex Epios 15 from 2000 until 2009 (Shimadzu Corporation), followed by a Signa HDxt 1.5 T starting in 2010 (GE Medical Systems). Scanning conditions for the Magnex Epios 15 were as follows: slice thickness, 4.0 mm; echo time, 110 ms; inversion time, 1700 ms; slice gap, 4.5 mm; matrix size, 256 × 256. The scanning conditions for the Signa HDxt were as follows: slice thickness, 4.0 mm; echo time, 140.8 ms; inversion time, 2250 ms; slice gap, 4.5 mm; matrix size, 512 × 512.

2.5. MR image processing

The diagnostic performances of the mCNN and epilepsy specialists were evaluated using 5-fold cross-validation. The MR images of the 85 patients in the MTLE-HS Group and the 56 participants in the Control Group were further divided into five specimen groups of approximately the same size. One specimen group was selected as the test dataset, while the remaining specimen groups were used as the training dataset. The training dataset was used to fine-tune the mCNN, and the test dataset was used to evaluate the diagnostic performances of the fine-tuned mCNN and epilepsy specialists. The abovementioned method was repeated so that each of the five specimen groups would form a test database once (Fold 1–5). Allocating the test datasets in this manner guarantees that a specific subject will not appear simultaneously in the training and test datasets.

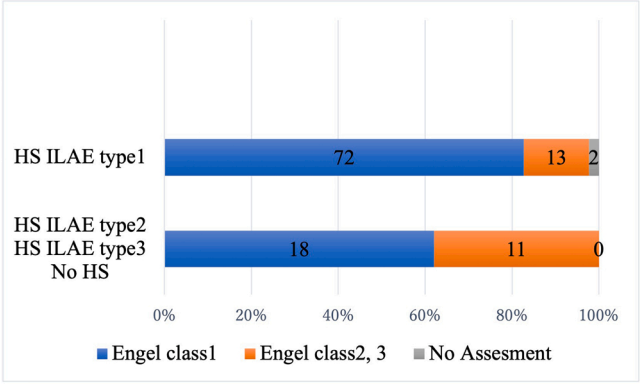

Fig. 2. Postoperative seizure outcomes of 114 patients one year after anterior temporal lobectomy (Engel Epilepsy Surgery Outcome Scale). HS, hippocampal sclerosis; ILAE, International League Against Epilepsy.

One 16-slice MR series of axial FLAIR sections that was taken parallel to the long axis of the hippocampus was examined per subject. Only three of these slices were selected for inclusion in the training dataset: the slice in which the hippocampus was the most visible, the slice immediately above it, and the slice immediately below it (Fig. 3). Only the single slice in which the hippocampus was the most visible was included in the test dataset. Originally in the DICOM format, each of the MR images selected was converted to the JPEG format using DICOM viewer software (OsiriX MD 11.0; www.osirix-viewer.com). After the FLAIR images were adjusted for size, 160 × 120-pixel sections were taken out: in the MTLE-HS group, sections were taken from the temporal lobe on the operated side, and in the Control group, sections were taken from the temporal lobes on both sides. Extracted sections were then magnified and white margin spaces filled in with a black background image; images were thus adjusted to 224 × 224 pixels. The pixel values of each FLAIR image were divided by 255 to normalize them, and data values were rescaled between 0 and 1. Next, CCN model training was carried out. This preprocessing was performed in Matlab2020a and Python using homemade scripts.

2.6. CNN classification model

Our model was developed by adapting and fine-tuning VGG-16—an existing CNN developed by the Visual Geometry Group—which was pretrained to classify inputs according to 1000 classes using the massive image database ImageNet (Lin et al., 2014). VGG-16 is a CNN composed of 16 layers, 13 convolutional layers, and two fully connected layers (Simonyan and Zisserman, 2015). This default architecture was modified by editing the fully connected layers to output only two classes: MTLE-HS and Control. As shown in Fig. 4, batch normalization (Ioffe and Szegedy, 2015) was applied to the first convolutional layer and additional dropout (Srivastava et al., 2014) before the fully connected layers. The resulting network is hereafter referred to as the mCNN. To prevent overfitting, training data were augmented using the ImageDataGenerator module (https://keras.io/ja/preprocessing/image/), which transformed the original images by shifting the height and width by 0–10%, shearing by 0–20%, and zooming by 0–20%, as well as flipping half of them horizontally. For fine-tuning, the mCNN was initialized with the pretrained (VGG-16) weights and then trained on the augmented training dataset, allowing all node weights to be updated in the learning process under the following hyperparameters: learning rate = 1×10^{-6}, batch size = 64, epoch number = 3000. RMSProp was utilized as the loss function (Geoffrey Hinton and Kevin, 2016).

2.7. Evaluation of the trained model

The performance of the fine-tuned mCNN in classifying the test dataset was evaluated using receiver operating characteristic (ROC) analysis in terms of the area under the ROC curve (AUC). Images were classified as "MTLE-HS" or "Control" when the probability of the respective class estimated by the fine-tuned mCNN exceeded 50% ($p \geq 0.50$). Diagnostic performance was evaluated in terms of sensitivity, specificity, false negative rate, and false positive rate. In parallel, five epileptologists were sent the same test dataset and requested to judge whether each image belonged to an MTLE-HS or normal brain. Diagnostic performance was evaluated in terms of sensitivity, specificity, and associated 95% confidence intervals (95% CI). Differences in

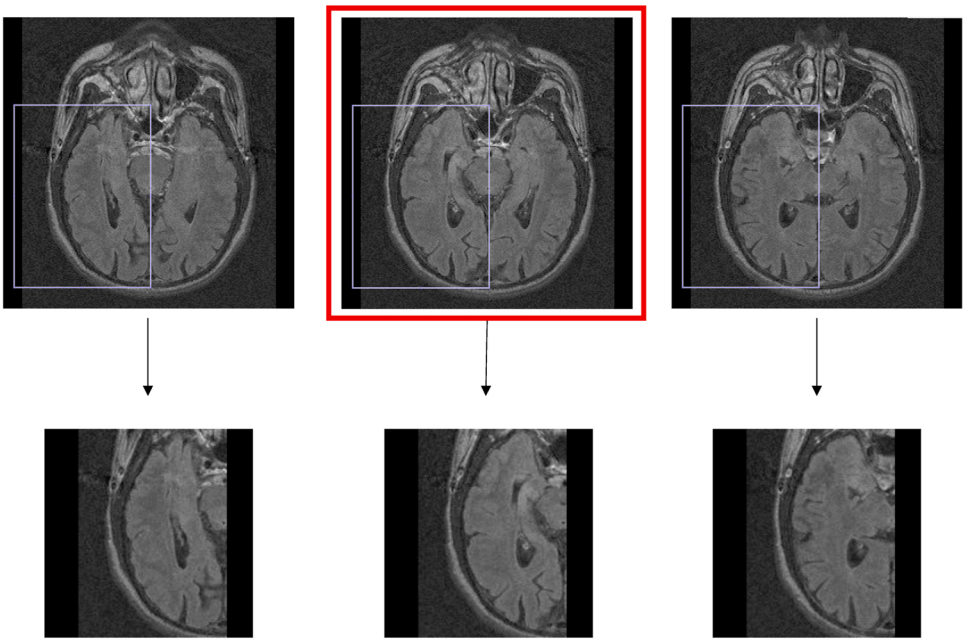

Fig. 3. Each subject's 16-slice FLAIR MRI series was reviewed to identify the slice, in which the hippocampus was most visible (denoted by the red rectangle). Only this slice and the slices immediately above and below were selected for inclusion in the training dataset (3 images/subject). Only the slice in which the hippocampus was most visible was included in the test dataset (1 slice per case). Regardless of group membership, each image was cropped to isolate the temporal lobe (selection areas denoted by blue rectangles) and subsequently resized to 224 × 224 pixels. FLAIR, fluid-attenuated inversion recovery; MRI, magnetic resonance imaging.

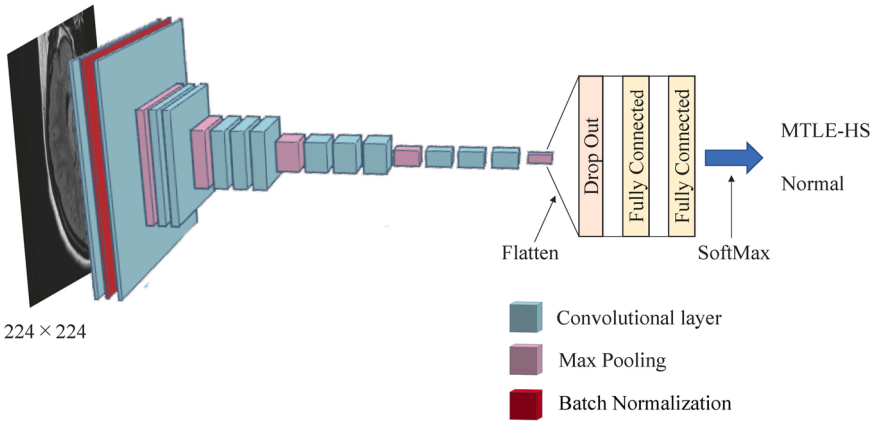

Fig. 4. Diagram of the mCNN architecture (adapted from VGG-16). HS, hippocampal sclerosis; mCNN, modified convolutional neural network; MTLE, mesial temporal lobe epilepsy.

the diagnostic performance between the evaluation by the fine-tuned mCNN and the visual interpretation performed by the epilepsy specialists was then compared. Finally, the image regions preferentially used by the fine-tuned mCNN when classifying the images were visualized as saliency heatmaps using Grad-CAM+ + (Chattopadhay et al., 2018).

2.8. Software and statistical analysis

The CNN was constructed using open-source libraries for TensorFlow 2.1.0 and Python 3.7.8 (Python Software Foundation) installed on an Ubuntu 16.04 operating system and trained on a PC equipped with an NVIDIA GeForce® 2080Ti GPU and Intel® Core™ i9–9820X CPU. Statistical analysis was conducted using the statistical programming language R (version 4.04) in an RStudio environment (version 1.4.1106) installed with the packages pROC and Epi.

3. Results

The training dataset was divided according to an 80/20 training/validation split to train the mCNN. Accuracy refers to the percentage of correct class predictions made by a model on the training data, whereas loss represents the aggregate error between the true class of images and the class predicted by the model.

The fine-tuned mCNN performed exceptionally well in diagnosing MTLE-HS, achieving mean accuracy of 87.8% [83%−93%], mean sensitivity of 91.1% [85%−96%], and specificity of 83.5% [75%−91%]. The five epileptologists interpreting the same test data set diagnosed MTLE-HS with mean sensitivity 73.1% [65%, 82%]) and mean specificity 66.3% [50%, 82%]). Table 1 The ROC curve of the fine-tuned mCNN is shown in Fig. 5. The associated mean AUC value is 0.94 (95% CI: [0.90, 0.98]; DeLong's method). The figure also shows the 95% confidence intervals calculated for the mean sensitivity/specificity of the five epileptologists. Both intervals are located entirely under the ROC curve for the fine-tuned mCNN.

4. Discussion

The model tested in this study was a modified CNN fine-tuned on preoperative MR images acquired from patients known to have MTLE with HS, as well as from control individuals diagnosed as nonepileptic after an outpatient visit to our hospital's epilepsy center. Our fine-tuned mCNN was not only able to diagnose MTLE-HS with high sensitivity and

Table 1
Diagnostic performance of the fine-tuned mCNN and epileptologists on five-fold cross-validation.

	Fine-tuned mCNN			Epileptologists		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
Fold 1	95%	88.2%	100%	67.5% [63–72%]	69.4% [58–81%]	66.0% [51–81%]
Fold 2	82.5%	76.4%	87.0%	71.5% [61–82%]	69.4% [59–80%]	73.0% [49–97%]
Fold 3	84.6%	76.4%	90.9%	68.7% [61–76%]	80.0% [70–90%]	60.0% [43–77%]
Fold 4	92.3%	94.1%	90.9%	65.6% [59–73%]	70.6% [62–79%]	61.8% [47–77%]
Fold 5	84.6%	82.3%	86.3%	73.3% [70–77%]	76.5% [66–87%]	70.9% [58–84%]
Mean	87.8% [83–93%]	91.1% [85–96%]	83.5% [75–91%]	69.3% [64–75%]	73.1% [65–82%]	66.3% [50–82%]

Accuracy, sensitivity, and specificity were presented on each of the five folds of the split test dataset (Folds 1–5) and on mean (average). Parenthesis means 95% confidence intervals.
Abbreviations: Fine-tuned mCNN, Fine-tuned modified convolutional neural network.

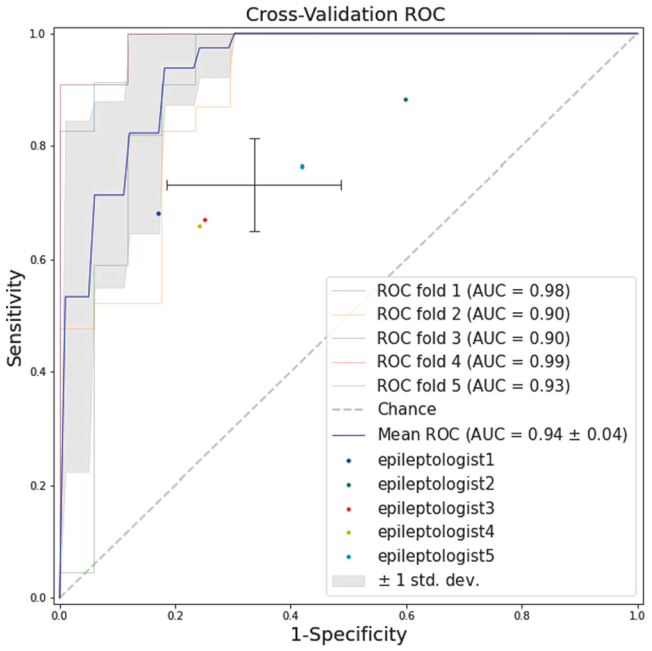

Fig. 5. The mean ROC curve of the fine-tuned mCNN is indicated by the blue line. The five colored dots (green, yellow, red, blue, aqua) correspond to the five epileptologists (epileptogist1–5); their positions indicate their sensitivity and 1-specificity values. The cross denotes the (aggregate) 95% confidence intervals for the sensitivity/1-specificity values of the expert interpretations. CI, confidence interval; mCNN, modified convolutional neural network; ROC, receiver operating characteristic.

specificity but also performed better than epilepsy specialists.

To the best of our knowledge, there are no reports that have used CNN to diagnose MTLE-HS from FLAIR images. We believe that our study is the first such report.

4.1. VGG-16 modifications and fine-tuning to minimize overfitting

In cases of small training datasets, classification performance can be improved through the use of a technique called *transfer learning*, whereby a CNN pretrained in one domain is applied to a different one (Tan et al., 2018). However, as the weights of the convolutional layers responsible for extracting image features are used 'as is' in transfer learning, the CNN classification models thus prepared may perform worse in situations where the images to be classified are weakly related to the images used in pretraining (Caruana, 1998). Because "MRI" is not represented among the 1000 classes of correct labels available in the massive image dataset used to pretrain the VGG-16 utilized in this work, we decided to adapt it to MR imaging diagnosis by means of *fine-tuning*: a technique whereby the weights of all nodes in a model are updated during retraining, including in the convolutional layers (Li and Hoiem, 2018). However, when a network is fine-tuned on a small dataset, it is prone to *overfitting*: a phenomenon that causes the models to classify training data with extraordinary accuracy but reduces their performance on unseen test data. Batch normalization can improve the stability and learning speed of a network's learning process, preventing gradient loss by normalizing the inputs of successive mini-batches (i.e., small datasets split from the training dataset) to have a mean of 0 and a standard deviation of 1 (Bjorck et al., 2018; Ioffe and Szegedy, 2015). Dropout prevents overfitting by probabilistically ignoring the output of certain nodes within a network so that learning proceeds independently of the output of any given node (Srivastava et al., 2014).

We believe that our use of these techniques helped to train our CNN classification model to become capable of diagnosing MTLE-HS based on MR image data with high accuracy.

4.2. Diagnostic performance of the fine-tuned mCNN with respect to expert interpretation

The diagnostic performance of the fine-tuned mCNN was superior to the interpretation of MRI scans by practicing specialists in the field of epilepsy. Studies of MRI interpretation-based methods of diagnosing MTLE-HS have reported sensitivity and specificity values ranging from 86% to 93% and from 83% to 85%, respectively (Cheon et al., 1998; Kuzniecky et al., 1997). The epileptologists enlisted in this study performed somewhat worse, diagnosing MTLE-HS with lower sensitivity and specificity than did those documented in the research cited above. Our mCNN was fine-tuned on MTLE-HS training images cropped to contain only the affected temporal lobe because it was the only area in which this diagnosis (i.e., the true label) was confirmed by the pathological evaluation of the resected hippocampal tissue. The performance of the fine-tuned mCNN on the test data was similarly evaluated using only cropped images of the affected temporal lobe, and the experts interpreted the same MR images so that their performance could be compared under the same conditions. Our experts may have diagnosed the images with lower sensitivity and specificity than in the reports cited above because of these test conditions, which meant they were unable to compare anatomical features between the affected and unaffected temporal lobes during interpretation. Methods involving the estimation of hippocampal volume from MR scans are considered useful alternatives to expert interpretation when diagnosing MTLE-HS because this parameter correlates well with the pathological features of hippocampal neuron loss (Watson et al., 1996). Efforts to diagnose MTLE-HS using hippocampal volume estimates derived by manual tracing in MR images have achieved a sensitivity of 81–97% and specificity of 82–88%(Cheon et al., 1998; Kuzniecky et al., 1997). In comparison, 82% sensitivity and 73% specificity have been achieved on hippocampal volume data obtained automatically using FreeSurfer brain imaging software (Hakimi et al., 2019). Our fine-tuned mCNN performed as well or better than the methods cited above. It has been reported that machine learning can be used to classify the lateralization of temporal lobe epilepsy (TLE) focus

with high precision (Sone and Beheshti, 2021). Most of the methods used in machine learning involve support vector machine (SVM); deep learning with convolutional neural networks is rarely used (Focke et al., 2012) (Chen et al., 2020; Rudie et al., 2015). It has been reported that deep learning can acquire features suitable for classification from input data by learning and shows high performance in classifying brain tumors (Lotlikar et al., 2021). Similarly, in this study, we were able to demonstrate that deep learning was useful in the categorization of MTLE-HS from FLAIR images.

4.3. Feature saliency in the fine-tuned mCNN

We visualized the feature saliency of our CNN using Grad-CAM++: an improved method that is capable of extracting even minuscule features (Chattopadhay et al., 2018). Grad-CAM++ can use the gradient information in the last convolutional layer to depict, using a heatmap, the pixels that influenced the class categorization. The heatmap is normalized between 0 and 1; hot colors indicate regions that strongly affected class categorization, whereas cold colors indicate regions that weakly affected class categorization (Chattopadhay et al., 2018; Selvaraju et al., 2020). MRI features considered to provide support for diagnosing a patient with MTLE-HS include signal abnormalities and atrophy in the hippocampus and enlargement of the temporal horn (Jackson et al., 1993; Meiners et al., 1994). Hippocampal hyperintensities, atrophy, and temporal horn enlargement are clearly visible in the images of all three MTLE-HS patients shown in Fig. 6, all of whom were correctly identified by both our model and expert interpretation. The Grad-CAM++ heatmaps indicate high salience in the hippocampus and temporal horn, confirming the importance of these regions to the fine-tuned mCNN. The fact that the local saliency of the fine-tuned mCNN is concentrated in the same regions examined by the epileptologists when diagnosing MTLE-HS supports the reliability of our model in recognizing this pathology (Fig. 6).

4.4. Research limitations and future directions

Our training dataset contained fewer data than are typically used to train models in the field of artificial intelligence (Esteva et al., 2017; Litjens et al., 2016). Sampling the data used for CNN learning and testing from different distributions reportedly results in "domain shift": i.e., causes the classification accuracy of the trained CNN to drop (Pooch et al., 2020). Using MRI at multiple facilities may resolve the problem of domain shift.

The volume of the hippocampus of healthy individuals reportedly differs between the right and left sides (Honeycutt and Smith, 1995); hence, studies that employ hippocampal volume and Voxel-based Morphometry (VBM) use the same side of the hippocampus to compare its volume in patients with MTLE with that in controls (Cheon et al., 1998; Focke et al., 2012). In our study, to increase the number of data items used for learning, we employed a data expansion technique that can randomly expand or shrink or distort images. Since the size of the hippocampus is not believed to be taken into account in learning, we also used images with the left-right reversed, without comparing the temporal lobes of the MTLE-HS and control groups on both sides.

The FLAIR images used in this study included the cranium; however, in a study in which deep learning was used to identify the location of focal cortical dysplasia, learning was carried out with images from which the cranium had been removed (House et al., 2021). Removal of data from MRIs often requires manual adjustment or the use of FSL and/or SPM toolboxes. As for the 2D MRI images collected in this study, the aforementioned methods would have resulted in incomplete removal of bone data; thus, the extracted FLAIR images were used directly as input data. However, if we were able to use input data from which bone features had been removed, we may have been able to achieve even greater performance, and this needs further investigation in future research. The fine-tuned mCNN was trained and tested on images of left and right temporal lobes cropped from axial FLAIR MR images taken parallel to the long axis of the hippocampus and was not designed to make judgments by comparing the left and right temporal lobes of the same patient: an approach typically employed by physicians when interpreting MRI scans for suspected MTLE-HS. Furthermore, the images of the unaffected temporal lobe of patients with MTLE-HS were not used during model training or testing, as the true label(s) were unknowable in the absence of histopathological evidence. Whether CNNs trained on MR images of both the left and right temporal lobes of MTLE-HS patients can distinguish the pathology from a normal hippocampus is a crucial topic for further inquiry.

Furthermore, our model was trained on images that had been cropped to prioritize just a few brain structures—primarily the temporal lobe. However, brain damage and volume loss have been observed not only in the affected temporal lobes of MTLE patients but also in the cortical and subcortical structures in other cerebral lobes (Bonilha et al.,

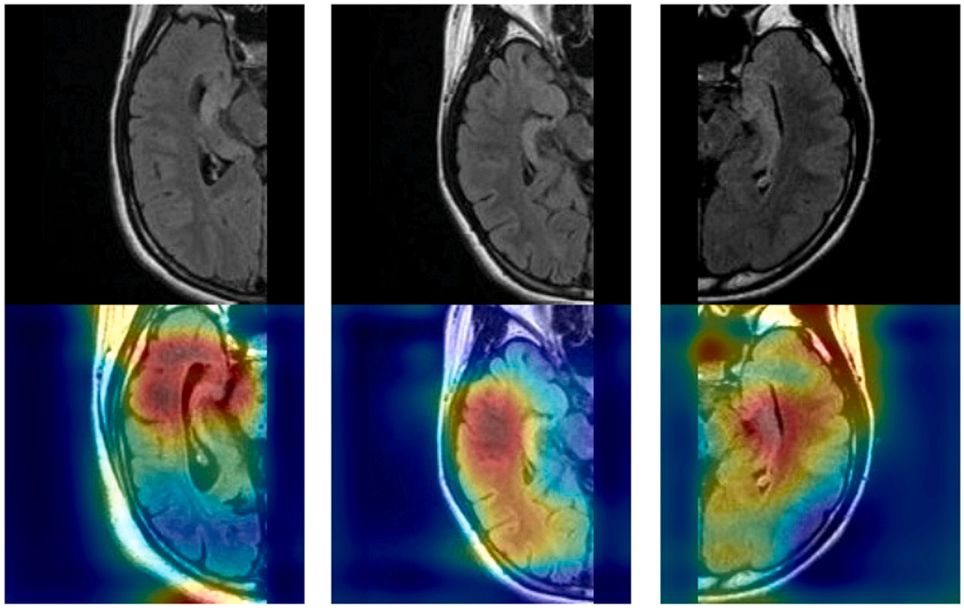

Fig. 6. Feature saliency in three cases correctly diagnosed as MTLE-HS by both mCNN and expert interpretation, visualized as Grad-CAM++ heatmaps. Hot colors indicate regions that strongly influenced determination of MTLE-HS classification. Cold colors indicate regions that weakly influenced determination of MTLE-HS classification. Each MR image is shown as used in the test dataset (upper row) and overlaid with the respective saliency heatmap generated by Grad-CAM++ (lower row). HS, hippocampal sclerosis; mCNN, modified convolutional neural network; MR magnetic resonance; MTLE, mesial temporal lobe epilepsy.

2005; Moran et al., 2001). Future research should investigate whether the present CNN's ability to accurately diagnose MTLE-HS could be improved by training it with whole-brain MR images.

We experienced the following problem: the more CNN layers we included, the higher the expressive power of the network, and the greater its recognition accuracy should have become, but at the same time, due to the vanishing gradient problem, recognition accuracy actually decreased. In InceptionV3, where micro-networks known as inception modules exist in an overlapping structure similar to that of a normal convolutional layer, or in ResNet50, which uses skip connections and residual networks, the vanishing gradient problem has been solved, giving these models high recognition accuracy (He et al., 2016; Szegedy et al., 2016). In the future, using these new deep learning models, we may be able to train a model with better recognition performance than the VGG16 model used in this study; therefore, further research is needed.

At least 30% of MTLE patients do not exhibit any visual, volumetric, or T2 relaxation time-related abnormalities (Carne et al., 2007; Connelly et al., 1998). In a recent systematic meta-analysis of such MRI-negative MTLE cases, evidence of HS in the resected temporal lobe was observed in only 13% of patients identified (n = 12/92) (Wang et al., 2016). Our model was trained exclusively on images of MTLE patients whose resected hippocampal tissue was graded as HS ILAE type 1. As the number of cases will increase in the future, it may become possible to train CNNs capable of identifying MTLE-HS of other histologic grades based on MRI data.

Beheshti et al. used DIR (double inversion recovery) images taken from MRI-negative TLE patients, and after applying SVM, they were able to achieve an accuracy of 84.1% when categorizing healthy individuals, left TLE, and right TLE from one another. They conclude that DIR images are beneficial biomarkers for the diagnosis of MRI-negative TLE (Beheshti et al., 2021). We believe that by combining DIR and deep learning in future research, we may well be able to train a model with even greater accuracy than that shown here.

5. Conclusions

The deep learning-based diagnosis of MTLE-HS using preoperative MR images performed superior to expert interpretation by epileptologists. Our mCNN model could serve as a useful preoperative diagnostic tool for identifying HS in patients with MTLE.

Funding

This study received no funding from external sources.

Declaration of Competing Interest

None.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.eplepsyres.2021.106815.

References

Beheshti, I., Sone, D., Maikusa, N., Kimura, Y., Shigemoto, Y., Sato, N., Matsuda, H., 2021. Accurate lateralization and classification of MRI-negative 18F-FDG-PET-positive temporal lobe epilepsy using double inversion recovery and machine-learning. Comput. Biol. Med. 137, 104805.

Berkovic, S.F., Andermann, F., Olivier, A., Ethier, R., Melanson, D., Robitaille, Y., Kuzniecky, R., Peters, T., Feindel, W., 1991a. Hippocampal sclerosis in temporal lobe epilepsy demonstrated by magnetic resonance imaging. Ann. Neurol. 29, 175–182.

Berg, A.T., Berkovic, S.F., Brodie, M.J., Buchhalter, J., Cross, J.H., van Emde Boas, W., Engel, J., French, J., Glauser, T.A., Mathern, G.W., Moshe, S.L., Nordli, D., Plouin, P., Scheffer, I.E., 2010. Revised terminology and concepts for organization of seizures and epilepsies: report of the ILAE Commission on Classification and Terminology, 2005-2009. Epilepsia 51, 676–685.

K Riney, A. Bogacz, E. Somerville, E. Hirsch, R. Nabbout, IE Scheffer, SM Zuberi, T. Alsaadi, S. Jain, J. French, N. Specchio, E. Trinka, S. Wiebe, S. Auvin, EC Wirrell, P. Tinuper, 2021a, ILAE Classification and Definition of Epilepsy syndromes with Onset at Variable Age: Position Statement by the ILAE Task Force on Nosology and Definitions, pp. 1–75, ⟨https://www.ilae.org/files/dmfile/Riney_VariableAges_7Apr21.pdf⟩.

Berkovic, S.F., Andermann, F., Olivier, A., Ethier, R., Melanson, D., Robitaille, Y., Kuzniecky, R., Peters, T., Feindel, W., 1991b. Hippocampal sclerosis in temporal lobe epilepsy demonstrated by magnetic resonance imaging. Ann. Neurol. 29, 175–182.

Bjorck, J., Gomes, C., Selman, B., Weinberger, K.Q., 2018. Understanding Batch Normalization, p. arXiv:1806.02375.

Blumcke, I., Thom, M., Aronica, E., Armstrong, D.D., Bartolomei, F., Bernasconi, A., Bernasconi, N., Bien, C.G., Cendes, F., Coras, R., Cross, J.H., Jacques, T.S., Kahane, P., Mathern, G.W., Miyata, H., Moshe, S.L., Oz, B., Ozkara, C., Perucca, E., Sisodiya, S., Wiebe, S., Spreafico, R., 2013. International consensus classification of hippocampal sclerosis in temporal lobe epilepsy: a Task Force report from the ILAE Commission on Diagnostic Methods. Epilepsia 54, 1315–1329.

Bonilha, L., Rorden, C., Castellano, G., Cendes, F., Li, L.M., 2005. Voxel-based morphometry of the thalamus in patients with refractory medial temporal lobe epilepsy. Neuroimage 25, 1016–1021.

Carne, R.P., O'Brien, T.J., Kilpatrick, C.J., Macgregor, L.R., Litewka, L., Hicks, R.J., Cook, M.J., 2007. 'MRI-negative PET-positive' temporal lobe epilepsy (TLE) and mesial TLE differ with quantitative MRI and PET: a case control study. BMC Neurol. 7, 16.

Caruana, R., 1998. Multitask Learning, Learning to Learn. Kluwer Academic Publishers,, pp. 95–133.

Cersósimo, R., Flesler, S., Bartuluchi, M., Soprano, A.M., Pomata, H., Caraballo, R., 2011. Mesial temporal lobe epilepsy with hippocampal sclerosis: study of 42 children. Seizure 20, 131–137.

Chattopadhay, A., Sarkar, A., Howlader, P., Balasubramanian, V.N., 2018. Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks, 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 839–847.

Chen, S., Zhang, J., Ruan, X., Deng, K., Zhang, J., Zou, D., He, X., Li, F., Bin, G., Zeng, H., Huang, B., 2020. Voxel-based morphometry analysis and machine learning based classification in pediatric mesial temporal lobe epilepsy with hippocampal sclerosis. Brain Imaging Behav. 14, 1945–1954.

Cheon, J.E., Chang, K.H., Kim, H.D., Han, M.H., Hong, S.H., Seong, S.O., Kim, I.O., Lee, S. G., Hwang, Y.S., Kim, H.J., 1998. MR of hippocampal sclerosis: comparison of qualitative and quantitative assessments. AJNR Am. J. Neuroradiol. 19, 465–468.

Coan, A.C., Kubota, B., Bergo, F.P.G., Campos, B.M., Cendes, F., 2014. 3T MRI quantification of hippocampal volume and signal in mesial temporal lobe epilepsy improves detection of hippocampal sclerosis. Am. J. Neuroradiol. 35, 77–83.

Connelly, A., Van Paesschen, W., Porter, D.A., Johnson, C.L., Duncan, J.S., Gadian, D.G., 1998. Proton magnetic resonance spectroscopy in MRI-negative temporal lobe epilepsy. Neurology 51, 61–66.

Ehteshami Bejnordi, B., Veta, M., Johannes van Diest, P., van Ginneken, B., Karssemeijer, N., Litjens, G., van der Laak, J., the, C.C., Hermsen, M., Manson, Q.F., Balkenhol, M., Geessink, O., Stathonikos, N., van Dijk, M.C., Bult, P., Beca, F., Beck, A.H., Wang, D., Khosla, A., Gargeya, R., Irshad, H., Zhong, A., Dou, Q., Li, Q., Chen, H., Lin, H.J., Heng, P.A., Hass, C., Bruni, E., Wong, Q., Halici, U., Oner, M.U., Cetin-Atalay, R., Berseth, M., Khvatkov, V., Vylegzhanin, A., Kraus, O., Shaban, M., Rajpoot, N., Awan, R., Sirinukunwattana, K., Qaiser, T., Tsang, Y.W., Tellez, D., Annuscheit, J., Hufnagl, P., Valkonen, M., Kartasalo, K., Latonen, L., Ruusuvuori, P., Liimatainen, K., Albarqouni, S., Mungal, B., George, A., Demirci, S., Navab, N., Watanabe, S., Seno, S., Takenaka, Y., Matsuda, H., Ahmady Phoulady, H., Kovalev, V., Kalinovsky, A., Liauchuk, V., Bueno, G., Fernandez-Carrobles, M.M., Serrano, I., Deniz, O., Racoceanu, D., Venancio, R., 2017. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. In: JAMA, 318, pp. 2199–2210.

Engel Jr., J., 1993. Outcome with respect to epileptic seizures. Surg. Treat. Epilepsies 609–621.

Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., Thrun, S., 2017. Dermatologist-level classification of skin cancer with deep neural networks. Nature 542, 115–118.

Focke, N.K., Yogarajah, M., Symms, M.R., Gruber, O., Paulus, W., Duncan, J.S., 2012. Automated MR image classification in temporal lobe epilepsy. Neuroimage 59, 356–362.

Geoffrey Hinton, N.S., Kevin Swersky, 2016. Lecture 6a Overview of mini-batch gradient descent.

Gulshan, V., Peng, L., Coram, M., Stumpe, M.C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., Kim, R., Raman, R., Nelson, P. C., Mega, J.L., Webster, D.R., 2016. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 316, 2402–2410.

Hakimi, M., Ardekani, B.A., Pressl, C., Blackmon, K., Thesen, T., Devinsky, O., Kuzniecky, R.I., Pardoe, H.R., 2019. Hippocampal volumetric integrity in mesial temporal lobe epilepsy: a fast novel method for analysis of structural MRI. Epilepsy Res 154, 157–162.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.

Honeycutt, N.A., Smith, C.D., 1995. Hippocampal volume measurements using magnetic resonance imaging in normal young adults. J. Neuroimaging 5, 95–100.

House, P.M., Kopelyan, M., Braniewska, N., Silski, B., Chudzinska, A., Holst, B., Sauvigny, T., Martens, T., Stodieck, S., Pelzl, S., 2021. Automated detection and segmentation of focal cortical dysplasias (FCDs) with artificial intelligence: presentation of a novel convolutional neural network and its prospective clinical validation. Epilepsy Res 172, 106594.

Ioffe, S., Szegedy, C., 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, in: Francis, B., David, B. (Eds.), Proceedings of the 32nd International Conference on Machine Learning. PMLR, Proceedings of Machine Learning Research, pp. 448–456.

Jackson, G.D., Berkovic, S.F., Duncan, J.S., Connelly, A., 1993. Optimizing the diagnosis of hippocampal sclerosis using MR imaging. AJNR Am. J. Neuroradiol. 14, 753–762.

Jackson, G.D., Berkovic, S.F., Tress, B.M., Kalnins, R.M., Fabinyi, G.C., Bladin, P.F., 1990. Hippocampal sclerosis can be reliably detected by magnetic resonance imaging. Neurology 40, 1869–1875.

Jutila, L., Immonen, A., Mervaala, E., Partanen, J., Partanen, K., Puranen, M., Kalviainen, R., Alafuzoff, I., Hurskainen, H., Vapalahti, M., Ylinen, A., 2002. Long term outcome of temporal lobe epilepsy surgery: analyses of 140 consecutive patients. J. Neurol. Neurosurg. Psychiatry 73, 486–494.

K Riney, A.B., Somerville, E., Hirsch, E., Nabbout, R., Scheffer, I.E., Zuberi, S.M., Alsaadi, T., Jain, S., French, J., Specchio, N., Trinka, E., Wiebe, S., Auvin, S., Wirrell, E.C., Tinuper, P., 2021b. ILAE Classification and Definition of Epilepsy syndromes with Onset at Variable Age: Position Statement by the ILAE Task Force on Nosology and Definitions, pp. 1–75.

Kuzniecky, R.I., Bilir, E., Gilliam, F., Faught, E., Palmer, C., Morawetz, R., Jackson, G., 1997. Multimodality MRI in mesial temporal sclerosis: relative sensitivity and specificity. Neurology 49, 774–778.

Labate, A., Gambardella, A., Aguglia, U., Condino, F., Ventura, P., Lanza, P., Quattrone, A., 2010. Temporal lobe abnormalities on brain MRI in healthy volunteers: a prospective case-control study. Neurology 74, 553–557.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.

Li, Z., Hoiem, D., 2018. Learning without forgetting. IEEE Trans. Pattern Anal. Mach. Intell. 40, 2935–2947.

Lin, M., Chen, Q., Yan, S., 2014. Network In Network. CoRR abs/1312.4400.

Litjens, G., Sanchez, C.I., Timofeeva, N., Hermsen, M., Nagtegaal, I., Kovacs, I., Hulsbergen-van de Kaa, C., Bult, P., van Ginneken, B., van der Laak, J., 2016. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. Sci. Rep. 6, 26286.

Lotlikar, V.S., Satpute, N., Gupta, A., 2021. Brain tumor detection using machine learning and deep learning: a review. Curr. Med Imaging.

Meiners, L.C., van Gils, A., Jansen, G.H., de Kort, G., Witkamp, T.D., Ramos, L.M., Valk, J., Debets, R.M., van Huffelen, A.C., van Veelen, C.W., et al., 1994. Temporal lobe epilepsy: the various MR appearances of histologically proven mesial temporal sclerosis. AJNR Am. J. Neuroradiol. 15, 1547–1555.

Menzler, K., Iwinska-Zelder, J., Shiratori, K., Jaeger, R.K., Oertel, W.H., Hamer, H.M., Rosenow, F., Knake, S., 2010. Evaluation of MRI criteria (1.5 T) for the diagnosis of hippocampal sclerosis in healthy subjects. Epilepsy Res. 89, 349–354.

Moran, N.F., Lemieux, L., Kitchen, N.D., Fish, D.R., Shorvon, S.D., 2001. Extrahippocampal temporal lobe atrophy in temporal lobe epilepsy and mesial temporal sclerosis. Brain 124, 167–175.

Mueller, C.A., Scorzin, J., von Lehe, M., Fimmers, R., Helmstaedter, C., Zentner, J., Lehmann, T.N., Meencke, H.J., Schulze-Bonhage, A., Schramm, J., 2012. Seizure outcome 1 year after temporal lobe epilepsy: an analysis of MR volumetric and clinical parameters. Acta Neurochir. (Wien) 154, 1327–1336.

Pooch, E., Ballester, P.L., Barros, R.C., 2020. Can we trust deep learning models diagnosis? The impact of domain shift in chest radiograph classification. ArXiv abs/1909.01940.

Riederer, F., Seiger, R., Lanzenberger, R., Pataraia, E., Kasprian, G., Michels, L., Beiersdorf, J., Kollias, S., Czech, T., Hainfellner, J., Baumgartner, C., 2020. Voxel-based morphometry-from hype to hope. A study on hippocampal atrophy in mesial temporal lobe epilepsy. AJNR Am. J. Neuroradiol. 41, 987–993.

Rudie, J.D., Colby, J.B., Salamon, N., 2015. Machine learning classification of mesial temporal sclerosis in epilepsy patients. Epilepsy Res. 117, 63–69.

Scheffer, I.E., Berkovic, S., Capovilla, G., Connolly, M.B., French, J., Guilhoto, L., Hirsch, E., Jain, S., Mathern, G.W., Moshe, S.L., Nordli, D.R., Perucca, E., Tomson, T., Wiebe, S., Zhang, Y.H., Zuberi, S.M., 2017. ILAE classification of the epilepsies: position paper of the ILAE commission for classification and terminology. Epilepsia 58, 512–521.

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2020. Grad-CAM: visual explanations from deep networks via gradient-based localization. Int. J. Comput. Vis. 128, 336–359.

Semah, F., Picot, M.C., Adam, C., Broglin, D., Arzimanoglou, A., Bazin, B., Cavalcanti, D., Baulac, M., 1998. Is the underlying cause of epilepsy a major prognostic factor for recurrence? Neurology 51, 1256–1262.

Simonyan, K., Zisserman, A., 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. CoRR abs/1409.1556.

Sone, D., Beheshti, I., 2021. Clinical application of machine learning models for brain imaging in epilepsy: a review. Front. Neurosci. 15.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15, 1929–1958.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the inception architecture for computer vision. Proc. IEEE Conf. Comput. Vis. Pattern Recognit. 2818–2826.

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., Liu, C., 2018. A Survey on Deep Transfer Learning. Springer International Publishing, Cham, pp. 270–279.

Thom, M., Mathern, G.W., Cross, J.H., Bertram, E.H., 2010. Mesial temporal lobe epilepsy: how do we improve surgical outcome? Ann. Neurol. 68, 424–434.

Von Oertzen, J., Urbach, H., Jungbluth, S., Kurthen, M., Reuber, M., Fernandez, G., Elger, C.E., 2002. Standard magnetic resonance imaging is inadequate for patients with refractory focal epilepsy. J. Neurol. Neurosurg. Psychiatry 73, 643–647.

Wang, X., Zhang, C., Wang, Y., Hu, W., Shao, X., Zhang, J.G., Zhang, K., 2016. Prognostic factors for seizure outcome in patients with MRI-negative temporal lobe epilepsy: a meta-analysis and systematic review. Seizure 38, 54–62.

Watson, C., Nielsen, S.L., Cobb, C., Burgerman, R., Williamson, B., 1996. Medial temporal lobe heterotopia as a cause of increased hippocampal and amygdaloid MRI volumes. J. Neuroimaging 6, 231–234.

Wiebe, S., Blume, W.T., Girvin, J.P., Eliasziw, M., Effectiveness, Efficiency of Surgery for Temporal Lobe Epilepsy Study, G., 2001. A randomized, controlled trial of surgery for temporal-lobe epilepsy. N. Engl. J. Med. 345, 311–318.

Wieser, H.G., Epilepsy, I.Co.No, 2004. ILAE commission report. Mesial temporal lobe epilepsy with hippocampal sclerosis. Epilepsia 45, 695–714.