IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 11, 2022, accepted March 9, 2022, date of publication March 16, 2022, date of current version March 25, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3159923

Ensemble Convolutional Neural Networks With Support Vector Machine for Epilepsy Classification Based on Multi-Sequence of Magnetic Resonance Images

IRWAN BUDI SANTOSO[1,2], YUDHI ADRIANTO[3], ANGGRAINI DWI SENSUSIATI[4], DIAH PUSPITO WULANDARI[1,5], AND I. KETUT EDDY PURNAMA[1,5], (Member, IEEE)

[1]Department of Electrical Engineering, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia
[2]Department of Informatics Engineering, Universitas Islam Negeri Maulana Malik Ibrahim Malang, Malang 65144, Indonesia
[3]Department of Neurology, Universitas Airlangga, Surabaya 60115, Indonesia
[4]Department of Radiology, Universitas Airlangga, Surabaya 60115, Indonesia
[5]Department of Computer Engineering, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia

Corresponding author: I. Ketut Eddy Purnama (ketut@te.its.ac.id)

This work was supported in part by the Indonesian Endowment Fund for Education (LPDP) through the Scheme of Riset Inovatif Produktif (RISPRO) - Invitasi 2019 Grant, under Contract PRJ-41/LPDP/2019.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethical Committee of Airlangga University Hospital, under Application No. 142/KEP/2020.

ABSTRACT Classification of brain abnormalities as a pathological cue of epilepsy based on magnetic resonance (MR) images is essential for diagnosis. There are some types of brain structural abnormalities as a pathological cue of epilepsy. To identify it, a neurologist can involve some sequence of MR images at a time. Existing algorithms for abnormalities classification usually involve only one or two sequences of MR images. In this paper, we proposed ensemble convolutional neural networks with a support vector machine (SVM) scheme to classify brain abnormalities (epilepsy) vs. non-epilepsy based on the axial multi-sequence of MR images. The convolutional neural network (CNN) models on the proposed method are base-learner models with different architectures and have low parameters. The performance improvement on the proposed method is made by combining the output of the base-learner models and the combination of predictions from these models. The combination of predictions uses majority voting, weighted majority voting, and weighted average. Henceforth, the combined output becomes input in the meta-learning process with SVM for the final classification. The dataset for evaluation is the axial multi-sequences of MR images that include abnormal brain structures causing epilepsy and non-epilepsy with various subjects' histories. The experimental results show the proposed method can obtain an accuracy average and F_1-score of 86.37% and 90.75%, respectively, and an improvement of accuracy of 6.7%-18.19% against the CNN models on the base-learner and 2.54%-2.65% against the combination of predictions. With these results, the proposed architecture also provides better performance compared to the two existing CNN architectures.

INDEX TERMS Convolutional neural network, ensemble, epilepsy, magnetic resonance image, support vector machine.

I. INTRODUCTION

Epilepsy is a chronic disease of the brain characterized by repeated seizures and is an unconscious movement that involves part of the body or the whole body [1]. Efforts to detect the disease early will help determine the cause of epilepsy. EEG (electroencephalogram) is generally used to check whether a patient is having an epileptic seizure, determine the type of seizure, or even a trigger factor for epilepsy. However, this diagnosis has not been able to understand the etiology and has the low spatial resolution to detect the brain abnormality as the cause of epilepsy [2]. Magnetic resonance

The associate editor coordinating the review of this manuscript and approving it for publication was Xinyu Du.

imaging (MRI) can detect changes in the microstructure of the source of epilepsy because it has a relatively high spatial resolution. Therefore, the study in [3] recommended structural MRI as the standard of investigation in epilepsy patients. Identification involving several sequences of MR images will be advantageous in detecting the brain abnormalities as a source of epilepsy (e.g., hippocampal sclerosis, cortical dysgenesis, brain tumor, cerebral vascular, and others). The HARNESS-MRI protocol shows the advantage of each sequence of MR images in identifying the brain's structural abnormalities (microstructural changes)[4]. However, each sequence of MR images provides different benefits in identifying any structural brain abnormality, as reported in [5]. Therefore, increasing the performance of the automatic method in processing MR images will help improve the sensitivity in the epilepsy identification.

Several researchers have previously reported the detection/classification results of epilepsy based on brain structure abnormalities (e.g., temporal lobe epilepsy, focal cortical dysplasia). Most of the researches they do are for the detection or classification of only one abnormality type, e.g., detection or classification abnormalities in temporal lobe epilepsy shown in [2], [6]–[9], focal cortical dysplasia (FCD) is reported in [10]–[12]. The results of studies in[6] have shown the use of one sequence of MR images to classify microstructural abnormalities in temporal lobe epilepsy (TLE) against non-TLE. Visual assessment of two sequences T1 and T2, has also been used for the diagnosis of hippocampal sclerosis (HS) in patients with mesial temporal lobe epilepsy (MTLE) [7]. In the case of FCD lesion detection, studies in [10] and [11] have reported the use of T1-weighted sequence as input for detection. Meanwhile, the use of two sequences (T1-MPRAGE and T2-FLAIR) for FCD detection is also discussed in [12]. These two abnormalities constitute the most significant percentage of epilepsy patients, as reported by Wellmer *et al.* [13]. A diagnosis of other types of brain abnormalities also uses a specific sequence of MR images to get the best results. Therefore, specific imaging protocols are required to identify a structural abnormality [13]. The initial diagnosis of whether a person has structural abnormalities of the brain or not must involve several sequences of MR images. Involving these many sequences of MR images in manual diagnosis is a maximal effort, but it is complicated and time-consuming. Therefore, the need for automated detection or classification with reliable methods, in this case, is essential. However, the automated detection/classification of epilepsy involving multiple sequences of MR images and types of abnormalities as simultaneous has not been investigated. Fig. 1. shows most of the previous studies, only using one or two sequences of MR images for identification/detection /classification of one brain structural abnormality type. Consequently, the studies involving only one or two sequences of MR images and a type of abnormality have drawbacks such as: not being able to identify/detect/classify epilepsy caused by other types of abnormalities at initial diagnosis and can decrease sensitivity.

Based on the weaknesses of the previous studies and the diagnostic protocol for each type of brain abnormality in [4], [5], and [13], the initial diagnosis needs many sequences of MR images to see the various possible abnormalities in each of these sequences. Therefore, we propose the method for the two-class classification of brain structures (epilepsy, non-epilepsy) by involving several sequences (multi-sequence) in the training process. Fig. 1 illustrates the focus of our study using multi-sequence of MR images with some types of brain abnormalities that cause epilepsy in training.

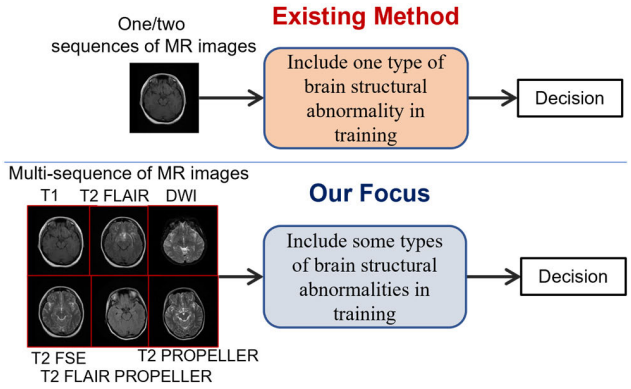

FIGURE 1. Most of the previous studies and our research focus on classifying brain structural abnormalities that cause epilepsy.

The multi-sequence of MR images impacts high data variability that it greatly affects the classifier's performance in identifying/classifying brain structural abnormalities. We use a convolutional neural network (CNN) as a classification method that has proven powerful for image data [14] and a CNN model ensemble technique to improve classification performance. The CNN model in this study is built by considering the low model parameters and the limited learning data, and maintaining the resulting performance. These CNN models serve as base-learner models in the ensemble technique. We use the ensemble technique to improve classification accuracy and reduce the variability of the results [15], [16]. The meta-learner stage using machine learning is beneficial in improving classification performance. Support vector machine (SVM) is one machine learning that has proven reliable in classifying brain abnormalities that cause epilepsy [2], [6]. Therefore, we propose an ensemble scheme for these CNN models using SVM at the meta-learner stage based on an axial multi-sequence of MR images (emsCNN-SVM) to improve classification performance. For that, we have conducted several experiments to evaluate the proposed emsCNN-SVM. The main contributions of this research are as follow:

- We propose axial multi-sequence of MR images approach to classify brain structural abnormalities causing epilepsy against non-epilepsy brain structures. Axial multi-sequence of MR images involved in the learning process contains some types of brain structural

abnormalities for epilepsy patients and some types of brain structures for non-epilepsy patients.

- We build the CNN model based on the multi-sequence of MR images as a base-learner model by considering the low parameter model and overfitting on the limited dataset to classify brain structural abnormalities that cause epilepsy vs. non-epilepsy brain structures.
- We propose a scheme CNN models ensemble on the base-learner with SVM on the meta-learner. It involves the output of the base-learner model and the predictions combination of these models, thus, it improves the performance and reduces variability in the classification of brain structural abnormalities that cause epilepsy vs. non-epilepsy brain structures.

The remainder of this paper is structured as follows: Section II discusses a survey of relevant previous research work on the classification of brain structural abnormalities that cause epilepsy. Section III describes the dataset of the experiment and the proposed method. The experimental scenarios and results are in Section IV. Section V discusses the experimental results. Finally, Section VI states the conclusions and suggestions for future research.

II. RELATED WORK

In this study, we classify brain structural abnormalities as cues that cause epilepsy vs. non-epilepsy subjects based on an axial sequence of MR images. Therefore, this section explores the relevant current research work in the literature from two prospective studies: first, the classification of brain structural abnormalities using machine learning, and second, the classification using CNN.

Classification of brain structural abnormalities that cause epilepsy using machine learning is reported in [6], [10]–[12], and [17]. Del Gaizo *et al.* [6] used diffusion MRI sequence to classify temporal lobe epilepsy (TLE) vs. non-TLE. They determined scalar diffusion from diffusion kurtosis imaging (DKI). Then, they used the weighted average of support vector machines models to classify TLE vs. non-TLE based on the scalar diffusion input. Their method yielded an accuracy of 68% (fractional anisotropy), 51% (mean diffusivity), dan 82% (mean kurtosis). The use of SVM was also reported by Wang *et al.* [17] to detect mesial temporal sclerosis (MTS) based on T1-weighted sequence. The detection begins with the segmentation of tissue (grey matter, white matter, and cerebrospinal fluid (CSF)), and hippocampus, followed by feature extraction of volume, shape, and ratio of CSF. The experimental results showed that their proposed technique provides promising performance for MTS. Studies on the detection of abnormalities in TLE are also reported in [7]–[9], only not using machine learning in its detection. Another abnormality classification, FCD, was performed by Qu *et al.* [10] using a multiple classifier fusion and optimization (MCFO) feature-based voxel-based morphometry (VBM) on T1-weighted MRI sequence. Their proposed MCFO involved several classifiers and minimized false positives using F-scores. The testing results with this method showed a decrease in false positives. The same study was conducted by Jin *et al.* [11] using T1-weighted sequence produced by three different magnetic resonance imaging scanners. They determined the morphological and intensity features as inputs for the non-linear neural network classifier. Their experiments at a threshold of 0.9 obtained an optimal sensitivity of 73.7% and a specificity of 90% in FCD detection. Mo *et al.* [12] also performed FCD lesion detection by combining quantitative multimodal surface features with an artificial neural network (ANN) to assess its clinical value. The testing results showed that the method's accuracy, sensitivity, and specificity were 70.5%, 70%, and 69.9%, respectively, which outperformed the unimodal classifier.

For the classification of brain structural abnormalities (epilepsy) by applying deep learning, some of them are reported in [2], [18], and [19]. Huang *et al.* [2] identified epilepsy using the DKI image. They segmented the hippocampus and used transfer learning VGG16 to get DKI image features. This feature was an input support vector machine (SVM) to classify epilepsy (hippocampus) vs. normal control. Their proposed method obtained the best classification accuracy of 90.8%. Torres-Velazquez *et al.* [18] used multimodal MRI to classify TLE. They introduced the Multi-Channel Deep Neural Network (mDNN) for TLE classification. Their experiments showed the potential of the mDNN approach to combine multiple data sets for TLE classification. Another abnormality classification (juvenile myoclonic epilepsy/JME) was conducted by Si *et al.* [19] using CNN-based transfer learning. They used diffusion MRI sequence to detect subtle changes in white matter. Using three CNN models, the experimental results showed that inception_resnet_v2 based transfer learning is better than Inception_v3 and Inception_v4 in classifying JME, with a classification accuracy of 75.2%.

It is considering the results of previous studies that combined extracting MR images features and machine learning to classify brain structural abnormalities as epilepsy cues. Most of these studies proposed the method to obtain MR image features that represent or combine some features [10]–[12], [17]. The researchers usually focused on one or two sequences of MR images to get these features. Besides, they typically used one classifier [6], [11], [12] or several classifiers [10] to get the best performance in the classification. These efforts are reasonable, but the best classification performance is not necessarily obtained by using the features that are considered representative. This approach can be ineffective and time-consuming, especially in studies involving multiple sequences of MR images and some types of abnormalities. Therefore, a reliable classifier is needed to solve this problem, such as the CNN classifier [2], [19]. A study in [2] showed that CNN is a robust classifier with a convolution process that will optimally perform feature extraction based on the classification results' loss function. The main problem we often encounter is that the dataset of MR images for epilepsy cases is relatively limited, consequently many researchers

rarely use CNN because it will have an overfitting effect. Several techniques can be used to solve an overfitting, such as augmenting data [20], [21], architectural design with low parameters [16], and validation techniques in learning.

In a previous study in [22], we have reported the CNN model with low parameters for epilepsy classification based on EEG signals. To overcome the limitations of the dataset in training, we divided the EEG signal into many segments (multi-segment) and converted it into a spectrogram image. This study used a CNN model and decided on the final classification results using majority voting based on the model predictions in each segment. Although the method in this study yielded good performance, it did not necessarily obtain good performance for the epilepsy classification based on MR images. These results occurred because the signal pattern was different from MR images.

In this study, we included multi-sequence of MR images for the brain abnormalities classification (epilepsy) against non-epilepsy to increase the performance (accuracy, sensitivity) and to overcome the limitations of the dataset. Involving multi-sequence of MR images on CNN will have high variability in results [23] so that the ensemble technique of some CNN models is a solution to improve accuracy and can reduce variability [15], [16]. Therefore, in this study, we propose ensemble CNN that differs from the existing methods in some aspects: (i) involving multi-sequence MR images and some types of brain abnormalities causing epilepsy, (ii) using some CNN models with low parameters as base-learner models, (iii) involving the output of the base-learner models and combinations of predictions as input to the meta-learner.

III. MATERIALS AND METHODS

A. DATASET ACQUISITION

We investigated several T1 and T2 sequences of 37 epilepsy patients. The patients consisted of 17 males and 20 females, including 48.6% with an additional history of epilepsy and seizures and 51.4% with an additional history of stroke, tumor, traumatic, temporal lobe, left focal epilepsy, syncope, cerebral edema, syncope, and hemianopia. Dataset sequences of MR images were obtained from Universitas Airlangga Hospital (Rumah Sakit Universitas Airlangga-RSUA), Surabaya, Indonesia, using a 1.5 T MRI scanner from 2018 to 2020. We have obtained the ethical clearance to use this retrospective dataset for research from the hospital's ethics committee. For the non-epilepsy dataset, we used nine healthy subjects and free of neurological disease, seven tumor patients, six patients of stroke, and five meningioma patients.

In this study, MRI sequences were acquired from each subject for the axial plane, including T1, T2-FLAIR, T2-FSE, DWI, T2-FLAIR PROPELLER, T2 PROPELLER. All MRI sequences were obtained with 2D acquisition type, slice thickness 5 mm, matrix 512×512 except for DWI 256×256, flip angle 90 degrees except for T2-FLAIR PROPELLER and T2 PROPELLER 160 degrees. While the repetition time in taking each MRI sequence was different, including T1 with a repetition time of 500 ms, T2 FLAIR 8800 ms, T2FSE 4212 ms, DWI, T2 FLAIR PROPELLER 8000 ms, and T2-PROPELLER 4780 ms.

From each sequence and a slice of epilepsy and non-epilepsy subjects, it was then converted into an MR image. Each image (frame) was selected and collected in an image dataset for experimental purposes. The total MR images used for the experiment were 4231, including 2515 epilepsy MR images and 1716 non-epilepsy MR images, as shown in Table 1.

TABLE 1. Sample of subject and frame MRI for experiment.

Dataset	Additional of history	#Training		#Testing		Total
		Subj	Frm	Subj	Frm	Frm
Epilepsy	Epilepsy, seizure,	13	824	5	328	2515
	stroke, tumor, traumatic, temporal lobe, another*	12	898	7	465	
Non-Epilepsy	Healthy, Tumor, Stroke, Meningioma	20	1413	7	303	1716
	Total	45	3135	19	1096	4231

*Left focal epilepsy, cerebral edema, syncope, and hemianopia
Subj= Subject, Frm= Frame

B. DATA PRE-PROCESSING

The input image for the CNN model must be the same size. Therefore, resizing the image of each slice is an essential pre-processing step. In this work, we decided to use a fixed size of 512×512 pixels because most of the results in the acquisition of MRI scanners were 2D type with a size of 512×512 except for the DWI sequence 256×256. This effort was to avoid a negative impact on the performance of the classification model [24]. The DWI image sequence from 256×256 size was changed to a predetermined target of 512×512 using MicroDicom, as shown in Fig. 2. The following pre-processing, which is also essential, is the normalization of each image pixel. The normalization is done to maintain process stability and convergence in the network. In this study, we normalized each image by changing each image pixel value from the range [0,255] to [0,1]. The normalization value was obtained by multiplying each image pixel by a scale factor of 1/255.

C. CONVOLUTIONAL NEURAL NETWORKS (CNNs)

A convolutional neural network is a deep learning model often applied to visual images and is proven to have high accuracy [14], [25]. There are five CNN architectures proposed in this study, each of which has several layers, namely input layer, convolutional layer, activation layer, pooling layer, fully-connected layer, and output layer. We name the five CNN architectures as $msCNN_1$, $msCNN_2$, $msCNN_3$, $msCNN_4$ and $msCNN_5$, as shown in Fig. 3. The CNN architectures are built to classify brain structural abnormalities

causing epilepsy vs. non-epilepsy brain structures based on axial multi-sequence of MR images.

In this study, we built the architectures of CNN with different structures for the epilepsy classification. In 2D/3D, areas of structural abnormalities in the brain have different sizes between subjects (patients). In addition, the involvement of some brain abnormalities types in this study also causes higher variability in the shape and size of brain structural abnormalities. Therefore, we decided to build some CNN models with different structures to strengthen the classification of brain structural abnormalities that cause epilepsy.

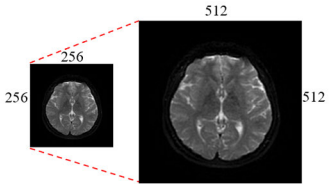

FIGURE 2. Example of converting a 256 × 256 image to 512 × 512 with MicroDicom.

1) INPUT LAYER

In this study, the input layer is the layer to enter the normalized sequence of MR images in the pre-processing stage into the convolution process. The input image size for each proposed CNN architecture is 512 × 512. These sizes are made equal to most of the original dimensions of each sequence of MR images to obtain complete feature information.

2) CONVOLUTIONAL LAYER

In this layer, the convolution process will be carried out on the input image of each MR sequence or input from the previous layer by shifting a filter. This process produces a feature map or image sequence pattern from a low to a high level [22]. Therefore, this convolution process will use many feature maps to obtain the characteristics of an image [26], [27].In this study, the convolution operation on the five proposed CNN models can be written as follows:

$$Z_i = f\left(W_i X + b_i\right), \quad i = 1, \ldots, 5 \tag{1}$$

where Z_i is the output of the convolution process of the msCNN $_i$ model, X is the input of the sequence of MR images, f (.) is the activation function, W_i is the weight of the convolution process of the msCNN$_i$ model, and b_i is the bias of the convolution process of the msCNN$_i$. These weights will undergo an update process to improve the classification results in the training process [28]. In this study, the number of filters used in each model is not the same. The architecture of msCNN$_1$ has five convolution layers, with the number of filters for each layer being 64, 128, 64, 32, and 16. The msCNN$_2$ has four convolution layers, with the number of filters in each layer being 32, 128, 64, and 32. The msCNN$_3$ has five convolution layers, with the number of filters for each layer being 32, 64, 32, 16, and 8. The msCNN$_4$ has five convolution layers, with each layer having 8, 16, 32, 16, and 8 filters. While the msCNN$_5$ also has five layers, with each layer having 16, 64, 32, 16, and 8 filters. The size of the filter in each convolution process is 3 × 3 with the same padding [29].

3) ACTIVATION LAYER

In this layer, an unsaturated activation function is applied to improve the nonlinearity of the decision function. In this study, the activation function used is the rectified linear unit (ReLU) [26], and for each model, it is presented in the following equation:

$$\hat{Z}_i(Z_i) = \begin{cases} Z_i, & Z_i \geq 0 \\ 0, & Z_i < 0, \end{cases} \qquad i = 1, \ldots, 5 \tag{2}$$

with $\hat{Z}_i$ is the ReLU process outputs of the msCNN$_i$ model.

4) POOLING LAYER

The pooling process at the layer aims to reduce the spatial size of the representation, reduce computations, and prevent overfitting. In this study, the pooling used is max-pooling [30], with the filter size of each proposed model being 2 × 2.

5) FULLY-CONNECTED LAYER

After the convolutional layer and max-pooling layer is the fully-connected layer. In this layer, the feedback process is carried out by refreshing the weights and biases against the previous layer and reducing the loss of feature information. The feature matrix of the prior layer process is converted into a feature vector (flatten) before the classification process. In this study, several proposed CNN architectures have different fully-connected layers. msCNN$_1$ and msCNN$_2$ have fully connected layers with all feature vectors (flatten) connected to the output layer, and 0.5 (50%) dropout is added. Meanwhile, for msCNN$_3$, msCNN$_4$ and msCNN$_5$ all have fully-connected layer 1 with dropout 0.5 process and fully-connected layer 2, which is fully connected with output layer. The number of neurons in the hidden layer for the msCNN$_3$ architecture is 32 with the ReLU activation function, while msCNN$_4$ and msCNN$_5$ have 64 neurons with the same activation function. In this study, the addition of a dropout process for fully-connected layer is proposed to prevent overfitting.

6) OUTPUT (CLASSIFICATION) LAYER

After the fully-connected layer, the results from this layer forward to the output (classification) layer to display the classification results, accuracy, and loss function. The loss function used in each proposed model is binary cross-entropy, while the activation function for classification is softmax. The softmax function of each proposed model can be written as in the following equation:

$$y_{ik}\left(\tilde{Z}_i\right) = \frac{\exp\left(\tilde{Z}_{ik}\right)}{\sum_{j=1}^{C} exp\left(\tilde{Z}_{ij}\right)}, \quad k = 1, \ldots, C; \quad i = 1, \ldots, 5 \tag{3}$$

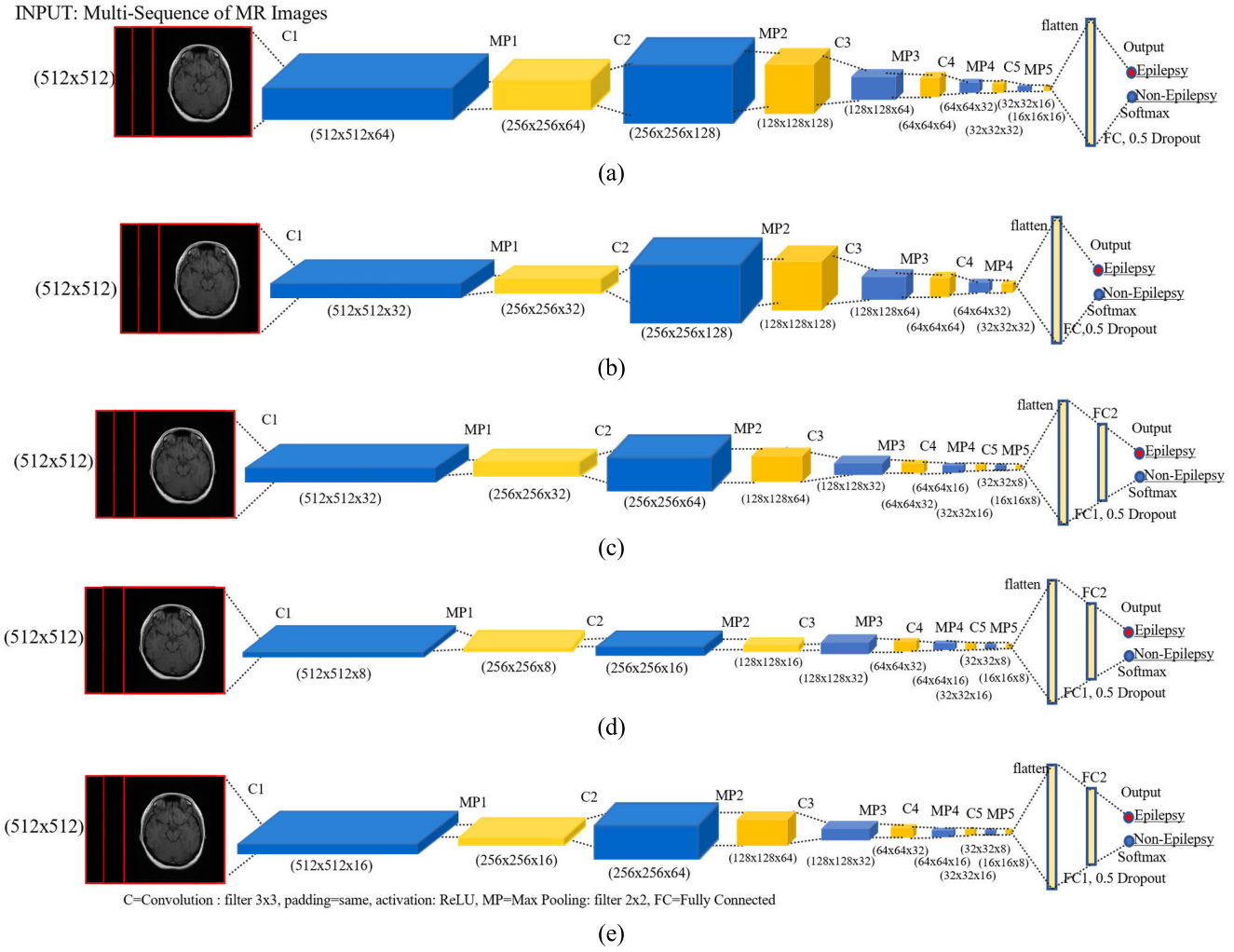

FIGURE 3. Proposed CNN architectures with multi-sequence of MR images input: (a) msCNN_1 (b) msCNN_2 (c) msCNN_3 (d) msCNN_4 (e) msCNN_5.

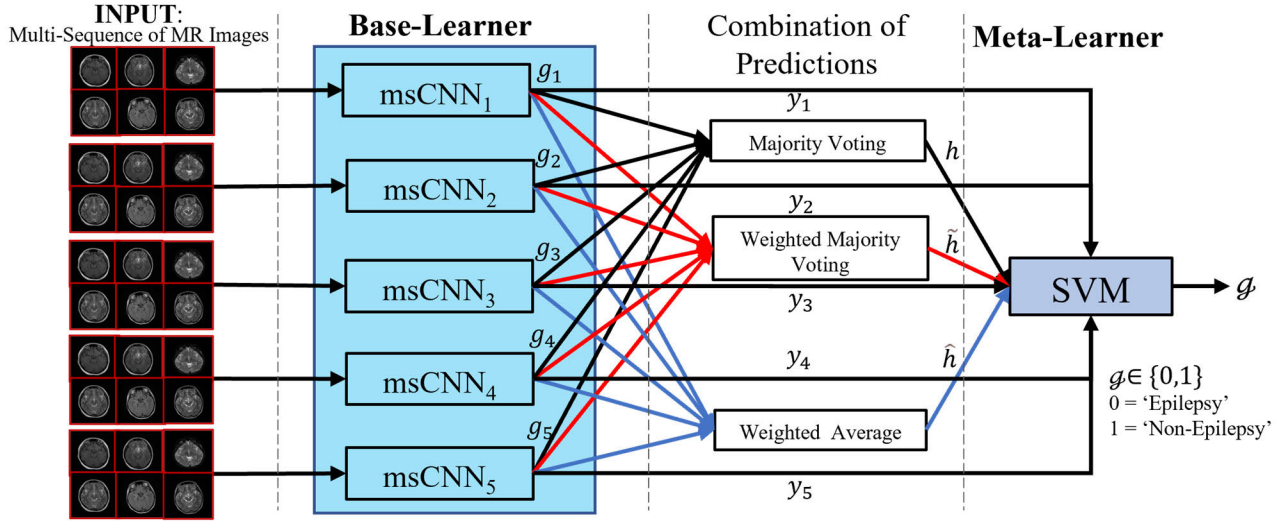

FIGURE 4. Proposed scheme: the ensemble of CNN models using SVM with the input of the CNN predictions, the softmax output, and the combination of predictions.

with y_{ik} is softmax outputs for the msCNN$_i$ model in k^{th} class. $\tilde{Z}_i$ is the process outputs at the fully-connected layer for the msCNN$_i$ models, and C is the number of classes (labels). In this study, the number of classes in training and testing is two (epilepsy, non-epilepsy).

In addition to the CNN architecture proposed in the scope of the study, we used three CNN architectures presented in the literature. The three architectures were CNN in [22], VGG16 [31], and ResNet50 [32], which we used as a comparison against the architectures proposed in this study. We transferred the architectures and trained these architectures with the dataset used in the study. CNN in [22] has a simple architecture and consists of three convolution layers with an output layer of 2 (epilepsy and non-epilepsy).The VGG16 model has 19 layers arranged sequentially, consisting of 16 convolutional layers and three fully-connected layers. The input image dimensions of the original VGG16 architecture are 224 × 224 × 3 with a fully connected output layer of 1000. While ResNet50 consists of 50 layers with five stages of the convolution process. The input and output layers dimensions of the architecture are the same as VGG16. In this study, we made some modifications to the two architectures. We made the image input of these architectures the same as the original architecture. In this context, we classified two classes (epilepsy and non-epilepsy), therefore, the size of the output layer was adjusted to two labels in both architectures. For the ResNet50 architecture, besides being modified in the output layer, a GlobalAveragePooling layer was also added before that layer.

D. ENSEMBLE CONVOLUTIONAL NEURAL NETWORKS

In this study, we used ensemble learning on the classification results of each proposed CNN model to improve performance and reduce the variability of the classification results. One type of ensemble learning is stacking or stacked generalization, which includes two main parts, namely base-learner and meta-learner [15], [33]. In this study, the models of msCNN$_1$, msCNN$_2$, msCNN$_3$, msCNN$_4$ and msCNN$_5$ are the base-learner models. While the support vector machine (SVM) is the meta-learner model. In our proposed scheme, between the base-learner and meta-learner, there is an ensemble process of base-learner models with a combination of predictions. The process is carried out by combining the prediction results of the base-learner model using majority voting, weighted average, and weighted majority voting [33]. The proposed scheme involving the combination of predictions is shown in Fig. 4.

The process of our proposed scheme begins with training on each base-learner model to get the y_1, y_2, y_3, y_4, dan y_5 using (3). For classifying brain abnormalities causing epilepsy vs. non-epilepsy (binary classification), the output has two probability values. Meanwhile, to predict the classification results of each model in the base-learner based on the largest probability value and mathematically, it can be written as follows:

$$g_i = \underset{k}{\text{argmax}}\,(y_{ik}), \quad g_i \in \{0, 1\}, \quad i = 1, \ldots, 5; \quad k = 1, 2 \tag{4}$$

with g_i is the predicted result of the msCNN$_i$ model. In our proposed scheme, we combine the results of g_1, g_2, g_3, g_4, and g_5 with majority voting (MV), weighted majority voting (WMV), and weighted average (WA).

This study uses majority voting to get predictive results based on the majority vote. If the msCNN$_1$, msCNN$_2$, msCNN$_3$, msCNN$_4$ and msCNN$_5$ models are as neurologists (experts), the final decision will be based on the results of the majority with a vote exceeding 50%. For example, it is known that v_{ik} is the voting result of the prediction of the i^{th} model, k^{th} class, then the value of $v_{ik} = 1$ is taken if the evaluation result of g_i is equal to the k^{th} class and $v_{ik} = 0$ if it is not the same. Furthermore, from the voting, the total vote for each class is $V_k = \sum_{i=1}^{5} v_{ik}, k = 1, 2$, and the ensemble result is determined based on the largest total voting value, which can be written as follows:

$$h = \underset{k}{\text{argmax}}\,(V_k), \quad h \in \{0, 1\}, \quad k = 1, 2 \tag{5}$$

A combination of predictions with a weighted majority voting is obtained by multiplying each prediction result of the model with a certain weight. In this study, the weights are obtained based on validation accuracy's proportional value in each base-learner model's last epoch. If a_i is the validation accuracy of the i^{th} model in the last epoch, then the weight of the results of each model is $\beta_i = a_i / \left(\sum_{i=1}^{5} a_i\right)$. For the case of binary classification with five models in the base-learner, if $g_i = 0$ the weights used are $\delta_{1i} = \beta_i$ and $\delta_{2i} = 0$ else $\delta_{1i} = 0$ and $\delta_{2i} = \beta_i$. Furthermore, the ensemble with a weighted majority voting can be written as follows:

$$\tilde{h} = \underset{k}{\text{argmax}} \left(\sum\nolimits_{i=1}^{5} \delta_{ki}\right), \quad \tilde{h} \in \{0, 1\}, \quad k = 1, 2 \tag{6}$$

The combination of predictions with the weighted average is obtained by averaging the value of the softmax (y_{ik}). Prediction result is determined based on the largest softmax average value among the existing classes. Mathematically the prediction result is written as follows:

$$\hat{h} = \underset{k}{\text{argmax}} \left(\sum\nolimits_{i=1}^{5} y_{ik}/5\right), \quad \hat{h} \in \{0, 1\}, \quad k = 1, 2 \tag{7}$$

The outputs of the CNN models on the base-learner and the combination of predictions will be input to the training process in the meta-learner. We used SVM in the meta-learner stage for training and final classification. The classifier was chosen because it required few assumptions for input data and flexibility in using kernel functions [34], [35]. If it is known that $\tilde{X}$ is input data on SVM with $\tilde{X} = \{g, y, h, \tilde{h}, \hat{h}\}$ then SVM, for binary classification, uses a linear model as follows:

$$\mathcal{g}\left(\tilde{X}\right) = \text{sign}(\omega^T \tilde{X} + \alpha) \tag{8}$$

where ω dan α are parameters. About the use of kernel functions in the training process, a transformation of $\tilde{X}$ is carried out with a function $\varphi(\tilde{X})$, which is called feature-space mapping, so that the classification function becomes

$$g\left(\tilde{X}\right) = \text{sign}(\omega^T \varphi(\tilde{X}) + \alpha) \tag{9}$$

The minimum geometric distance point $\tilde{X}$ from the hyperplane in the training sample is indicated by $\left|\omega^T \varphi(\tilde{X}) + \alpha\right| / \|\omega\|$. Next, we want all data points to be correctly classified so that $t_n \left(\omega^T \varphi\left(\tilde{X}_n\right) + \alpha\right) > 0$, for all n and $t \in \{-1, 1\}$ is the target. Accordingly, the distance of point $\tilde{X}_n$ to the decision surface is given by $t_n \left(\omega^T \varphi\left(\tilde{X}_n\right) + \alpha\right) / \|\omega\|$. To maximize the minimum geometric distance, it is equivalent to finding the following function

$$\underset{\omega,\alpha}{\text{argmax}} \left\{ \frac{1}{\|\omega\|} \min_n \left(t_n \left(\omega^T \varphi(\tilde{X}_n) + \alpha \right) \right) \right\} \tag{10}$$

The optimization problem requires that we maximize $1/\|\omega\| = \|\omega\|^{-1}$, which is equivalent to minimizing $\|\omega\|^2$ and mathematically, it can be written as follows:

$$\begin{aligned} &\underset{\omega,\alpha}{\text{argmin}} \left\{ \frac{1}{2} \|\omega\|^2 \right\} \\ &\text{subject to} : t_n \left(\omega^T \varphi(\tilde{X}_n) + \alpha \right) \geq 1, \quad n = 1, \ldots, N \end{aligned} \tag{11}$$

E. CLASSIFICATION RESULT EVALUATION

To evaluate the classification results, we adopted several measurement indicators, accuracy (AC), precision (PR) sensitivity (SE), and F_1-score ($F1$) [36]. The measurement is determined based on the parameter values of true positive (TP), false positive (FP), true negative (TN), and false negative (FN). In this study, TP is the number of times an epilepsy patient is labeled as epilepsy by the classification results. FP is the number of times a non-epilepsy data person is labeled as an epilepsy patient in the same way. TN is the number of times a non-epielpsy data person is labeled as a non-epilepsy patient in the same way. On the other hand, FN is the number of times an epilepsy data patient is labeled as a non-epilepsy person data in the same way. AC, PR, SE, and $F1$ values calculated using these parameters are defined mathematically in (12)-(15).

$$AC = (TP + TN)/(TP + TN + FP + FN) \tag{12}$$

$$PR = TP/(TP + FP) \tag{13}$$

$$SE = TP/(TP + FN) \tag{14}$$

$$F1 = 2(PR)(SE)/(PR + SE) \tag{15}$$

IV. EXPERIMENTS AND RESULTS

A. EXPERIMENTS

This study's total subjects were 64 people (37 epilepsy subjects and 27 non-epilepsy subjects). We divided the subjects into 45 subjects (25 epilepsy and 20 non-epilepsy) for training and the remaining (12 epilepsy and seven non-epilepsy) for testing, as shown in Table 1. We used stratified 5-fold cross-validation [37] to evaluate each method in the classification of epilepsy with the number of frames for training, validation for each fold, and testing, as shown in Table 2. In this study, the success of the class label ''epilepsy'' classification is more precedence because of the urgency. Therefore, the number of frames (epilepsy) for training or testing is more than that of the non-epilepsy. Based on this consideration, the evaluation of each method was determined using (12)-(15). On the other hand, the evaluation results of each method are worth comparing, the training process uses the same index file. This study uses Google Collaboratory to implement all these evaluations in each experimental scenario.

Algorithm 1: Meta Learning Phase (emsCNN-SVM)

Input: G (label data), q (the number of fold), $y, g, h, \tilde{h}, \hat{h}$ (the softmax output of CNN model in base-learner, the class prediction of CNN model in base-learner, the class prediction by majority voting, the class prediction by weighted majority voting, the class prediction by weighted average), kernel

Output: $\mathcal{M}$ (trained SVM model)

```
X̃ ← {y, g, h, h̃, ĥ}
G ← []
for j = 1 to q do
    train ← load the index file of training data of j^th fold from X̃, G
    X̃_train, G_train ← Set data and labels as X̃[train], G[train]
    M_j ← training data the SVM model using X̃_train, G_train and kernel
end
return M
```

The main stages of the proposed method in training refer to the proposed scheme as shown in Fig. 4, while the process steps at the meta-learner refer to Algorithms 1. Training of the base-learner model is carried out on each CNN model with the same input axial multi-sequence of MR images. The input shape in each scenario for the base-learner model is $512 \times 512 \times 1$, as shown in Table 3. The training process for the base-learner model refers to the CNN architecture in Fig. 3. All training in each fold used the Adam optimizer because it is relatively consistent [38]. The default learning rate for training in each base-learner model is 0.001, while the batch size and epoch are 16 and {50, 100, 150}. Algorithm 1 shows the steps of training with SVM at the meta-learner stage. We used SVM to train the dataset on each fold with several kernel functions, including linear, RBF, and polynomial (several degrees). Furthermore, we selected the best results from these experiments. Besides the training using the proposed model, we also conducted a training using the existing CNN models, including CNN in [22], VGG16, and Resnet50. Training with this model was also carried out in

each scenario with an input shape of $224 \times 224 \times 3$, except for CNN in [22] with an input shape of $512 \times 512 \times 3$. In this case, we used the stochastic gradient descent (SGD) optimizer for these models in the training process, while the learning rate used was 0.0001 (VGG16 and Resnet50) and 0.001 (CNN in [22]). We found that the optimizer and learning rate were suitable for these models in the pre-testing.

Algorithm 2: Testing Phase (emsCNN-SVM)

Input: X_{test} (testing data), q (the number of fold), m (the number of,CNN model) β (the weight based on the validation accuracy in the last epoch), M (the model parameters of msCNN_1, msCNN_2, msCNN_3, msCNN_4, and msCNN_5 as the result of the training in the base-learner), $\mathfrak{M}$ (trained SVM model)

Output: $\mathcal{G}$ (the final prediction)

$\tilde{X}_{test} \leftarrow []$
for $i = 1$ *to* m **do**
  for $j = 1$ *to* q **do**
    $M_{ij} \leftarrow$ load the saved base-learner model parameters
    $y_{pred(ijk)} \leftarrow$ the probabilities of each class (k) and X_{test} using M_{ij}
    $g_{pred(ij)} \leftarrow \text{argmax}\,(y_{pred(ijk)})$
    $\varepsilon_{test(j)} \leftarrow \varepsilon_{test(j)} \cup g_{pred(ij)}$
    $V_{test(jk)} \leftarrow$ the total of voting of each class (k) and X_{test} based on $\varepsilon_{test(j)}$
    $h_{pred(j)} \leftarrow \text{argmax}\,(V_{test(jk)})$
    $\delta_{test(ijk)} \leftarrow$ the weight of voting of each class (k) and X_{test} based on β_{ij} and $g_{pred(ij)}$
    $\tilde{h}_{pred(j)} \leftarrow \text{argmax}\left(\sum_{i=1}^{m} \delta_{test(ijk)}\right)$
    $\hat{h}_{pred(j)} \leftarrow \text{argmax}\left(\left(\sum_{i=1}^{m} y_{pred(ijk)}\right)/m\right)$
  end
end
$\tilde{X}_{test} \leftarrow \{y_{pred}, g_{pred}, h_{pred}, \tilde{h}_{pred}, \hat{h}_{pred}$
for $j = 1$ *to* q **do**
  $\mathcal{G}_j \leftarrow$ the prediction of each $\tilde{X}_{test}$ using $\mathfrak{M}_j$
end
return $\mathcal{G}$

In this study, we saved each MR image in Portable Network Graphics (PNG) type with a resolution of 512 × 512 pixels. Images of type PNG have four channels (RGBA). To get the input shape of 512 × 512 × 3 (the three channels), we converted RGBA to RGB and RGB to grayscale to get the input shape of 512 × 512 × 1 (one channel). Meanwhile, to get the input shape with different resolution sizes (e.g. from (512 × 512 × 3) to (224 × 224 × 3)), we used the nearest interpolation method. We used this method to resize the resolution and applied it to each MR image and sequence of MR images for training and testing purposes.

We tested all methods on the test sample with the same dataset treatment in the testing phase. The test steps are shown in Algorithm 2. The training parameters for each fold were then used to classify all frames (images) on the same test dataset. Classification performance was obtained by determining the average value of the classification results of all folds. The testing was carried out to see the average performance of the proposed method against other methods.

TABLE 2. Brief description of the used dataset for experiment.

Class label	Fold	Number of samples (frames)			Sequence of MRI
		Tra	Val	Tes	
Epilepsy	1,2	1377	345	793	T1, T2-FLAIR, T2-FSE, DWI, T2-FLAIR PROPELLER, T2 PROPELLER
	3,4,5	1378	344		
Non-Epilepsy	1,2	1131	282	303	
	3,4,5	1130	283		

Tra=Training, Val= Validation, Tes=Testing

TABLE 3. The number of model parameters for experiment.

Model		Input shape	# Parameter
Base-Learner	msCNN_1	512x512x1	179,570
	msCNN_2	512x512x1	195,106
	msCNN_3	512x512x1	108,698
	msCNN_4	512x512x1	142,938
	msCNN_5	512x512x1	164,954
Base-learner + Meta-learner	Proposed emsCNN-SVM	512x512x1	791,608*
Existing	CNN in [22]	512x512x3	1,150,022
	VGG16	224x224x3	134,268,738
	ResNet50	224x224x3	23,591,810

*Total # parameter of msCNN_1, msCNN_2, msCNN_2, msCNN_4, msCNN_5 + #parameter of SVM

B. EXPERIMENTAL RESULTS

In this section, we report the experiment's results using our proposed method, including its constituent methods. The experimental results reported are the performance of the methods at the base-learner stage, the combination of predictions, and meta-learner. Therefore, all methods have been tested in each testing scenario, as shown in Table 4-8.

In the first scenario with epoch = 50, the CNN models on the base-learner yielded the classification accuracy average of 71.64%-77.43% with the standard deviation range of 2.1-5.94. The CNN model ensemble on the base-learner using the predictions combination (MV, WA, and WMV) obtained the classification accuracy average of 80.38%-80.53% with the standard deviation of 1.88 - 2.10. The combination of predictions in this scenario obtained better classification accuracy than all base-learner models and lowered classification accuracy variability. However, testing with the proposed emsCNN-SVM yielded the classification accuracy average still better than it was. SVM with kernel polynomial and degree of 50 on meta-learner provided an accuracy improvement of the CNN models on base-learner by 5.33%-11.11% and 2.23%-2.37% on the combination of predictions. Generally, the proposed emsCNN-SVM presented deviation of the classification accuracy of each fold a relatively smaller than others, as shown in Table 4. The proposed method also yielded an average of sensitivity and F_1-score

better than others even though the classification precision was lower than the combination of predictions.

In the scenario with epoch = 100, the base-learner model yielded an accuracy average of 68.18%-79.67% with a standard deviation of 2.92-7.28. The combination of predictions obtained an accuracy average of 83.72%-83.83% and a standard deviation of 1.94-2.10. These results showed that the combining predictions using MV, WA, WMV obtained better results than base-learner models, but testing with the proposed emsCNN-SVM yielded the best results. SVM with the polynomial kernel (degree = 25) on the proposed emsCNN-SVM provided an accuracy improvement of the base-learner models by 6.70%-18.19% and 2.54%-2.65% for the combination of predictions. Based on the standard deviation value for classification accuracy, the proposed emsCNN-SVM yielded relatively lower variability than others.

Based on the resulting classification sensitivity value, the ensemble using our proposed emsCNN-SVM obtained the best average of classification sensitivity. Meanwhile, the base-learner model ensemble using the combination of predictions yielded an average of classification sensitivity better than the base-learner model. The proposed emsCNN-SVM provided an average improvement of classification sensitivity of 9.68%-28.45% for base-learner models and 7.24%-7.41% for all combinations of predictions. In general, this method also yielded lower variability in classification sensitivity than the others.

From the precision value in the epilepsy classification yielded in this scenario, emsCNN-SVM with the polynomial kernel (degree = 25) obtained a lower precision than the combination of prediction (MV, WA, and WVM). MV, WA, and WVM yielded the highest average value for classification precision with the lowest level of variability. However, in general, the proposed emsCNN-SVM yielded better average classification precision than the base-learner model with lower variability than those models. This method also obtained the highest F_1-score and provided an average improvement of classification F_1-score of 5.35%-16.75% for the base-learner models and 2.35%-2.45% for the combination of predictions.

In the experimental scenario with epoch = 150, the proposed emsCNN-SVM in general still presented a better average performance in the classification than the CNN model on the base learner and the combination of predictions. Even though at epoch = 100, the average classification performance of the proposed emsCNN-SVM was still better than epoch = 150, but at epoch = 150, it produced a lower level of variability than all scenarios. In this scenario, the CNN model on the base-learner provided a better level of variability in classification accuracy than the CNN model in other scenarios with a standard deviation of 1.63-4.20. The same results are also shown for sensitivity and F_1-score.

TABLE 4. Accuracy for proposed emsCNN-SVM, base-learner models, the combination of predictions with stratified 5-fold cross-validation.

Model		Epoch=50			Epoch=100			Epoch=150		
		$\overline{AC}$(%)	σ	Δ	$\overline{AC}$(%)	σ	Δ	$\overline{AC}$(%)	σ	Δ
Base-learner	$msCNN_1$	72.76	5.94	-10.00	79.03	3.82	-7.34	76.08	1.78	-8.76
	$msCNN_2$	71.64	4.48	-11.11	68.18	7.28	-18.19	71.86	3.78	-12.97
	$msCNN_3$	77.43	3.12	-5.33	79.67	2.92	-6.70	79.62	2.34	-5.22
	$msCNN_4$	76.30	2.10	-6.46	79.23	3.50	-7.14	79.07	4.20	-5.77
	$msCNN_5$	74.31	3.05	-8.45	79.58	3.04	-6.79	77.74	1.63	-7.10
Combination of predictions	MV	80.38	2.10	-2.37	83.72	2.10	-2.65	82.96	1.92	-1.88
	WA	80.53	1.88	-2.23	83.83	1.94	-2.54	82.90	2.05	-1.93
	WMV	80.38	2.10	-2.37	83.72	2.10	-2.65	82.96	1.92	-1.88
Proposed emsCNN-SVM	Meta-learner: SVM kernel = polynomial	**82.76**	d=50 1.30	-	**86.37**	d=25 1.98	-	**84.84**	d =10 0.58	-

$\overline{AC}$ = average of AC , σ =standard deviation, Δ = $\overline{AC}$- $\overline{AC}$ of proposed emsCNN-SVM, d=degree

TABLE 5. Precision for proposed emsCNN-SVM, base-learner models, the combination of predictions with stratified 5-fold cross-validation.

Model		Epoch=50			Epoch=100			Epoch=150		
		$\overline{PR}$(%)	σ	Δ	$\overline{PR}$(%)	σ	Δ	$\overline{PR}$(%)	σ	Δ
Base-learner	$msCNN_1$	86.40	3.45	-1.47	87.48	2.52	-1.56	87.88	1.58	0.13
	$msCNN_2$	85.99	3.09	-1.88	88.76	1.66	-0.28	86.99	2.20	-0.76
	$msCNN_3$	87.74	1.30	-0.14	89.14	1.80	0.10	87.19	1.71	-0.56
	$msCNN_4$	88.96	1.95	1.08	88.90	1.85	-0.14	88.65	2.82	0.90
	$msCNN_5$	87.04	1.83	-0.84	88.53	2.64	-0.51	86.74	2.24	-1.01
Combination of predictions	MV	90.48	0.67	2.61	**91.76**	0.82	2.72	**90.53**	0.47	2.78
	WA	**90.62**	0.49	2.75	**91.76**	0.81	2.71	90.52	0.38	2.77
	WMV	90.48	0.67	2.61	**91.76**	0.82	2.72	**90.53**	0.47	2.78
Proposed emsCNN-SVM	Meta-learner: SVM kernel = polynomial	87.87	d=50 0.89	-	89.04	d=25 1.38	-	87.75	d=10 1.56	-

$\overline{PR}$ = average of PR , σ =standard deviation, Δ = $\overline{PR}$- $\overline{PR}$ of proposed emsCNN-SVM, d=degree

TABLE 6. Sensitivity for proposed emsCNN-SVM, base-learner models, the combination of predictions with stratified 5-fold cross-validation.

Model		Epoch=50			Epoch=100			Epoch=150		
		$\overline{SE}$(%)	σ	Δ	$\overline{SE}$(%)	σ	Δ	$\overline{SE}$(%)	σ	Δ
Base-learner	$msCNN_1$	74.50	12.37	-13.87	82.90	4.32	-9.68	77.68	2.42	-14.25
	$msCNN_2$	72.66	5.08	-15.71	64.14	11.21	-28.45	71.93	6.04	-20.00
	$msCNN_3$	80.03	5.34	-8.35	81.97	5.20	-10.62	84.26	3.73	-7.67
	$msCNN_4$	76.82	3.04	-11.55	81.61	6.88	-10.97	81.49	3.51	-10.44
	$msCNN_5$	75.84	4.95	-12.53	82.50	2.03	-10.09	81.84	2.95	-10.09
Combination of predictions	MV	81.46	3.09	-6.91	85.17	3.63	-7.41	85.37	2.69	-6.56
	WA	81.54	3.07	-6.83	85.35	3.49	-7.24	85.30	2.85	-6.63
	WMV	81.46	3.09	-6.91	85.17	3.63	-7.41	85.37	2.69	-6.56
Proposed emsCNN-SVM	Meta-learner: SVM kernel = polynomial		d=50			d=25			d=10	
		88.37	1.87	-	**92.59**	3.17	-	**91.93**	1.89	-

$\overline{SE}$ = average of SE , σ =standard deviation, $\Delta = \overline{SE}$ - $\overline{SE}$ of proposed emsCNN-SVM, d=degree

TABLE 7. F_1-score for proposed emsCNN-SVM, base-learner models, the combination of predictions with stratified 5-fold cross-validation.

Model		Epoch=50			Epoch=100			Epoch=150		
		$\overline{F1}$(%)	σ	Δ	$\overline{F1}$(%)	σ	Δ	$\overline{F1}$(%)	σ	Δ
Base-learner	$msCNN_1$	79.43	5.93	-8.68	85.09	2.92	-5.66	82.44	1.45	-7.32
	$msCNN_2$	78.70	3.77	-9.41	74.00	8.15	-16.75	78.62	3.55	-11.15
	$msCNN_3$	83.62	2.74	-4.49	85.30	2.60	-5.44	85.65	1.83	-4.11
	$msCNN_4$	82.40	1.76	-5.71	84.93	3.18	-5.82	84.91	3.11	-4.85
	$msCNN_5$	80.96	2.85	-7.15	85.40	2.12	-5.35	84.17	1.27	-5.60
Combination of predictions	MV	85.71	1.75	-2.40	88.30	1.76	-2.45	87.86	1.53	-1.91
	WA	85.81	1.62	-2.30	88.40	1.64	-2.35	87.81	1.64	-1.95
	WMV	85.71	1.75	-2.40	88.30	1.76	-2.45	87.86	1.53	-1.91
Proposed emsCNN-SVM	Meta-learner: SVM kernel = polynomial		d=50			d=25			d=10	
		88.11	0.98	-	**90.75**	1.46	-	**89.77**	0.39	-

$\overline{F}1$ = average of $F1$, σ =standard deviation, $\Delta = \overline{F}1$- $\overline{F}1$ of proposed emsCNN-SVM, d=degree

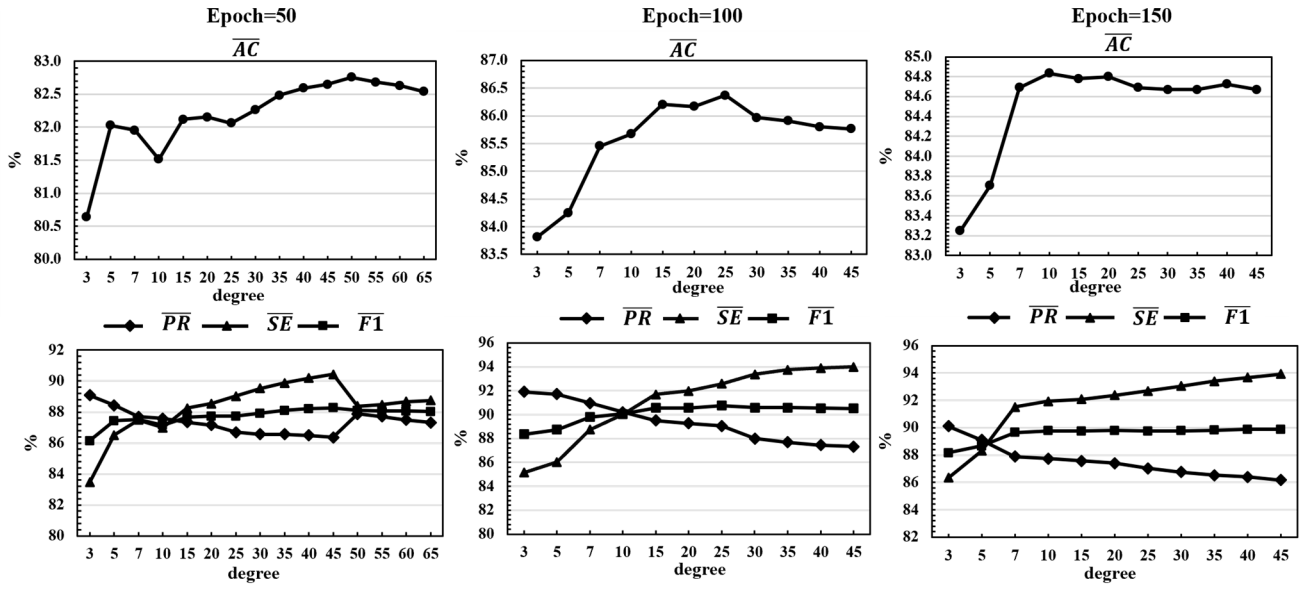

FIGURE 5. Classification performance of proposed emsCNN-SVM with kernel 'polynomial' and different degrees.

The testing results with CNN in [22], VGG16, and ResNet50 for each scenario can be seen in Table 8. The results of testing at epoch = 50, 100, 150 with split evaluation 5-fold cross-validation showed that VGG16 obtained an average of accuracy and precision better than ResNet50, but still lower than CNN in [22]. Whereas our proposed emsCNN-SVM and emsCNN-SVM* yielded an accuracy average better than the others. At epoch = 50, emsCNN-SVM provided an average improvement of classification accuracy of 7.67% for CNN in [22], 10.61% for VGG16, and 14.48% for ResNet50.

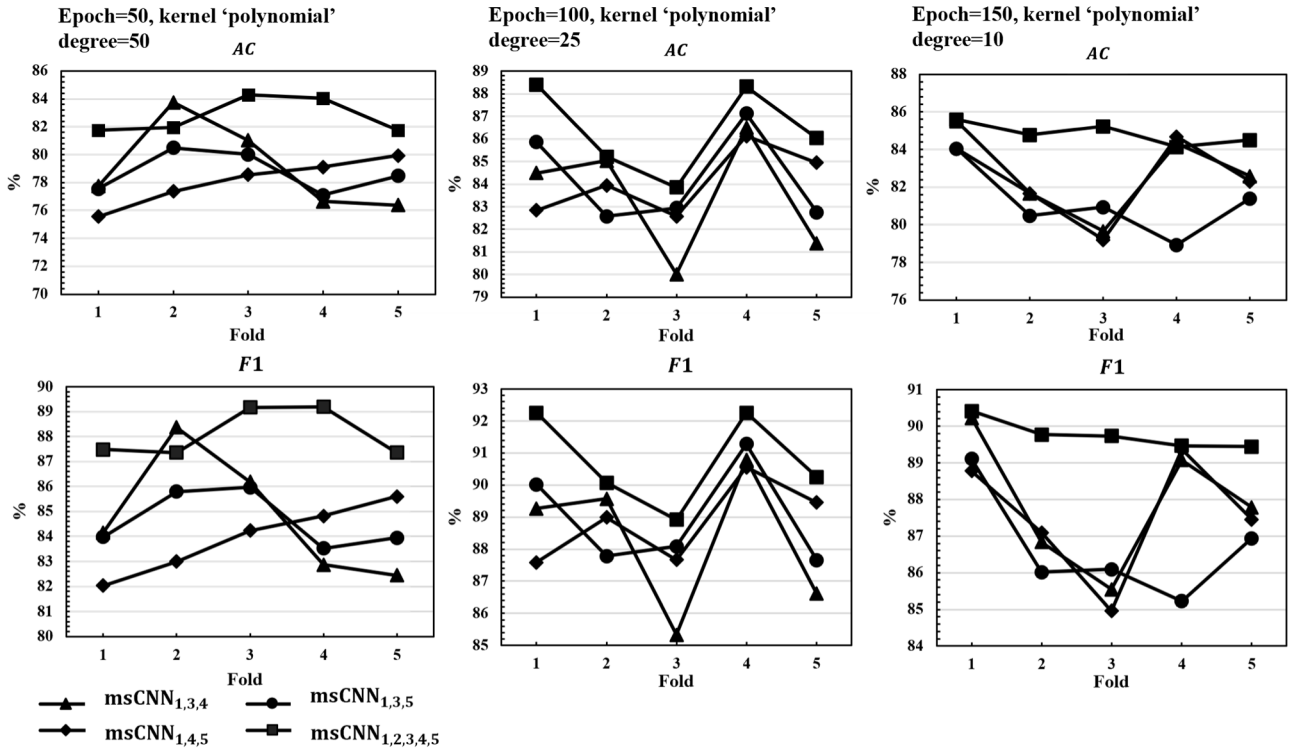

FIGURE 6. Accuracy and F_1-score of proposed emsCNN-SVM by including three and five based-learner models with kernel 'polynomial' and degree = 25.

While, at epoch = (100,150), our proposed emsCNN-SVM presented an accuracy improvement of (12.41%, 8.82%) for CNN in [22], (12.66%, 10.52%) for VGG16 and (16.97%, 14.66%) for ResNet50. For an average of sensitivity and F_1-score in classification, our proposed emsCNN-SVM also obtained the best results.

V. DISCUSSION

In this section, we investigated the performance of CNN models on base-learner, the ensemble of the base-learner model with a combination of predictions (MV, WA, and WMV) and meta-learner. At the meta-learner stage, we investigated the ensemble of the CNN models on base-learner by meta-training using SVM to classify the dichotomous axial sequence of MR images of the brain as epilepsy vs. non-epilepsy. On the other hand, we also investigated some existing CNN models compared to our proposed emsCNN-SVM.

The results of testing showed that the CNN model on the base-learner obtained classification performance with high variation. The CNN model on the base-learner yielded a classification accuracy average in testing in the range of 68.18%-79.67% with a standard deviation of 1.63-7.28. When viewed from the many parameters in the base-learner model, $msCNN_2$ was more than the other models, as shown in Table 3. However, the large number of model parameters does not guarantee that it is proportional to the classification performance produced, especially in the axial multi-sequence of MR images. The classification accuracy average of each

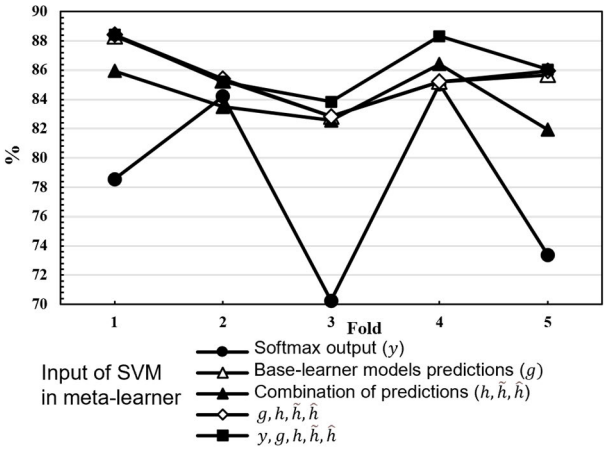

FIGURE 7. Accuracy of proposed emsCNN-SVM with epoch = 100, kernel 'polynomial' degree = 25, and different input in meta-learner.

CNN model on the base-learner is still below 80%. The training and testing data variability level are relatively high because it involves a multi-sequence of MR images, which affects the performance.

When likened to CNN models on the base-learner is a neurologist who reads axial multi-sequence of MR images, then the reading of each neurologist may give different results. Using the combination of predictions with majority voting, weighted majority voting, and weighted average can increase the accuracy of epilepsy classification and reduce

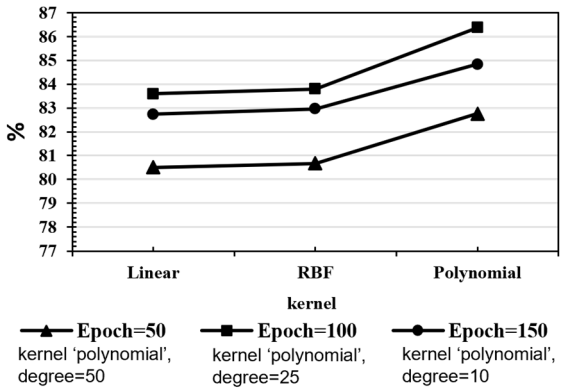

FIGURE 8. Accuracy average of the proposed emsCNN-SVM with different kernels and epochs.

TABLE 8. Performance comparison of the proposed emsCNN-SVM with the existing CNN.

Epoch	Model	Performance (%)			
		$\overline{AC}$	$\overline{PR}$	$\overline{SE}$	$\overline{F1}$
50	ResNet50	68.28	76.11	81.94	78.84
	VGG16	72.15	85.06	74.80	79.35
	CNN in [22]	75.09	89.59	74.20	81.15
	emsCNN-SVM*	79.25	**90.05**	80.18	84.81
	emsCNN-SVM	**82.76**	87.87	**88.37**	**88.11**
100	ResNet50	69.40	77.99	80.68	78.86
	VGG16	73.71	85.84	76.25	80.76
	CNN in [22]	73.96	**89.97**	72.06	79.99
	emsCNN-SVM*	81.88	89.12	85.45	87.19
	emsCNN-SVM	**86.37**	89.04	**92.59**	**90.75**
150	ResNet50	70.18	75.62	86.76	80.76
	VGG16	74.32	85.94	77.12	81.29
	CNN in [22]	76.02	89.38	75.89	82.08
	emsCNN-SVM*	83.70	**89.50**	87.79	88.62
	emsCNN-SVM	**84.84**	87.75	**91.93**	**89.77**

$\overline{AC}$ = average of AC, $\overline{PR}$ = average of PR, , $\overline{SE}$ = average of SE, , $\overline{F1}$ = average of $F1$,*) input shape : 224 x 224 x 3, kernel 'polynomial' degree=7

the variability of classification results. However, the increase stops at a certain level (saturation) and is difficult to increase because it depends entirely on the predictions of the classification of models on the base-learner. A meta-learner stage in the proposed emsCNN-SVM has become one of the solutions to improve classification accuracy and better than the combination of predictions. The improving classification accuracy can be found because, at the meta-learner stage, it depends not only on the results of the base-learner model but there is also meta-learning using SVM. The learning not only involves the prediction results of the base-learner models but also the results of the combination of predictions to improve classification performance. The proposed emsCNN-SVM accommodated the output of the CNN model on the base-learner and the combination of predictions (MV, WA, and WMV), accordingly, it yielded better and more stable performance for each scenario.

We realize that the best results in our proposed scheme involving SVM, in this case, do not apply to all kernels in training. At epoch = 50, 100, 150 kernel functions that give better results than others (e.g., RBF and linear) are polynomials with degree (d) = 50, 25, 10, as shown in Fig. 8. In this study, the criteria for determining the degree of the polynomial function are based on the best classification accuracy average, as shown in Fig. 5. In general, the greater the degree of the polynomial function, the higher the sensitivity values, but the impact on the precision decreases. The selection of polynomial kernel degrees based on the maximum sensitivity value will impact the low precision values. Therefore, the best choice is selecting polynomial kernel degrees in the proposed emsCNN-SVM based on the highest accuracy value. The option indirectly considers the value of precision, sensitivity, and F_1-score.

The number of models in the ensemble also influences the performance of the proposed emsCNN-SVM in classifying epilepsy against non-epilepsy. By involving five CNN models on the base-learner, it gives a better classification performance than applying only three CNN models. Fig. 6 shows the accuracy value and F_1-score for each fold involving five models giving better results than involving only three base-learner models. The involvement of inputs in meta-learning also affects classification performance. The proposed emsCNN-SVM involving three kinds of input: the base-learner model's predictions, the combination of predictions, and the softmax output of the base-learner models provides better classification accuracy than involving only two types of input and one kind of input, as shown in Fig. 7.

To know the performance or stability of our proposed method, we also compared the results with the existing models: CNN in [22], VGG16, and ResNet50. The results of testing with the same dataset treatment appeared that our proposed method improved all performances in the classification, as shown in Table 8. We realize that there are differences in the input image dimensions in these testing, which will affect the performance [39]. The proposed scheme has an input image dimension of $512 \times 512 \times 1$, while VGG16 and ResNet50 are $224 \times 224 \times 3$, respectively [31],[40]. We consider the comparison of these methods to be fair, even though our proposed method has a different input shape. In this case, we try to keep the original architecture of VGG16/Resnet50. However, the comparison results are fairer, we added the testing with an input resolution of $224 \times 224 \times 3$ for each proposed CNN model. In this study, we adjusted to the existing architecture in the model. Although the conditions in the comparison are still far from ideal, at least our proposed emsCNN-SVM is feasible to compare with these models, especially in the classification of brain structural abnormalities that cause epilepsy vs. non-epilepsy.

Our study has several limitations, including the relatively small samples of sequence of MR images used in training and testing. At the clinical level, validation must be carried out on more data involving many institutions. On the other hand, studies involving multi-sequence of MR images and different types of brain abnormalities within a class of

epilepsy certainly have the potential to reduce the classifier's performance. In addition, using only axial planes can also obtain lower performance than involving all other planes: sagittal and coronal.

This study only uses five CNN models on the base-learner. We understand that more CNN models in the base-learner will enrich the decisions and strengthen the results for the combination of predictions and processes on the meta-learner. However, more models in the base-learner will impact the number of model parameters used. Therefore, we decided to use five models of the base-learner for the ensemble process with better results than the three models of the base-learner.

VI. CONCLUSION

In this study, a method has been proposed to improve performance in the classification of epilepsy based on axial multi-sequence of MR images with an ensemble of several CNN models. The ensemble model is carried out by applying the principle of stacked generalization. The output of the CNN models of the base-learner and combination of predictions (majority voting, weighted average, and weighted majority voting) forwarded to SVM in the meta-learner stage. The proposed scheme can generally improve performance in classifying brain structural abnormalities causing epilepsy vs. non-epilepsy. The testing results show that the proposed scheme has a high potential to assist neurologists (clinicians) in identifying epilepsy patients based on multi-sequences of MR images.

For clinical purposes, in the future, there is still potential to improve the performance of epilepsy classification based on multi-sequence of MR images by increasing the amount of training or testing data and involving all planes of MR images.

REFERENCES

[1] World Health Organization. (2019). *Epilepsy*. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/epilepsy

[2] J. Huang, J. Xu, L. Kang, and T. Zhang, "Identifying epilepsy based on deep learning using DKI images," *Frontiers Human Neurosci.*, vol. 14, p. 465, Nov. 2020.

[3] W. D. Gaillard, C. Chiron, J. Helen Cross, A. Simon Harvey, R. Kuzniecky, L. Hertz-Pannier, and L. Gilbert Vezina, "Guidelines for imaging infants and children with recent-onset epilepsy," *Epilepsia*, vol. 50, no. 9, pp. 2147–2153, Sep. 2009.

[4] A. Bernasconi, F. Cendes, W. H. Theodore, R. S. Gill, M. J. Koepp, R. E. Hogan, G. D. Jackson, P. Federico, A. Labate, A. E. Vaudano, I. Blümcke, P. Ryvlin, and N. Bernasconi, "Recommendations for the use of structural magnetic resonance imaging in the care of patients with epilepsy: A consensus report from the international league against epilepsy neuroimaging task force," *Epilepsia*, vol. 60, no. 6, pp. 1054–1068, May 2019.

[5] I. Wang, A. Bernasconi, B. Bernhardt, H. Blumenfeld, F. Cendes, Y. Chinvarun, G. Jackson, V. Morgan, S. Rampp, A. E. Vaudano, and P. Federico, "MRI essentials in epileptology: A review from the ILAE imaging taskforce," *Epileptic Disorders*, vol. 22, no. 4, pp. 421–437, Aug. 2020.

[6] J. Del Gaizo, N. Mofrad, J. H. Jensen, D. Clark, R. Glenn, J. Helpern, and L. Bonilha, "Using machine learning to classify temporal lobe epilepsy based on diffusion MRI," *Brain Behav.*, vol. 7, no. 10, Oct. 2017, Art. no. e00801.

[7] C. Liao, K. Wang, X. Cao, Y. Li, D. Wu, H. Ye, Q. Ding, H. He, and J. Zhong, "Detection of lesions in mesial temporal lobe epilepsy by using MR fingerprinting," *Radiology*, vol. 288, no. 3, pp. 804–812, Sep. 2018.

[8] I. Beheshti, D. Sone, F. Farokhian, N. Maikusa, and H. Matsuda, "Gray matter and white matter abnormalities in temporal lobe epilepsy patients with and without hippocampal sclerosis," *Frontiers Neurol.*, vol. 9, p. 107, Mar. 2018.

[9] X. Feng, M. J. Hamberger, H. C. Sigmon, J. Guo, S. A. Small, and F. A. Provenzano, "Temporal lobe epilepsy lateralization using retrospective cerebral blood volume MRI," *NeuroImage: Clin.*, vol. 19, pp. 911–917, Mar. 2018.

[10] X. Qu, K. Deblaere, W. Philips, J. Yang, L. Platis, A. Kumcu, D. Ai, B. Goossens, T. Bai, Y. Wang, and J. Sui, "Multiple classifier fusion and optimization for automatic focal cortical dysplasia detection on magnetic resonance images," *IEEE Access*, vol. 6, pp. 73786–73801, 2018.

[11] B. Jin, B. Krishnan, S. Adler, K. Wagstyl, and W. Hu, "Automated detection of focal cortical dysplasia type II with surface-based magnetic resonance imaging postprocessing and machine learning," *Epilepsia*, vol. 59, no. 5, pp. 982–992, May 2018.

[12] J.-J. Mo, J.-G. Zhang, W.-L. Li, C. Chen, N.-J. Zhou, W.-H. Hu, C. Zhang, Y. Wang, X. Wang, C. Liu, B.-T. Zhao, J.-J. Zhou, and K. Zhang, "Clinical value of machine learning in the automated detection of focal cortical dysplasia using quantitative multimodal surface-based features," *Frontiers Neurosci.*, vol. 12, pp. 1–11, Jan. 2019.

[13] J. Wellmer, C. M. Quesada, L. Rothe, C. E. Elger, C. G. Bien, and H. Urbach, "Proposal for a magnetic resonance imaging protocol for the detection of epileptogenic lesions at early outpatient stages," *Epilepsia*, vol. 54, no. 11, pp. 1977–1987, Nov. 2013.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[15] A. I. Naimi and L. B. Balzer, "Stacked generalization: An introduction to super learning," *Eur. J. Epidemiol.*, vol. 33, no. 5, pp. 459–464, May 2018.

[16] A. Yazdizadeh, Z. Patterson, and B. Farooq, "Ensemble convolutional neural networks for mode inference in smartphone travel survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 6, pp. 2232–2239, Jun. 2020.

[17] H. Wang, S. N. Ahmed, and M. Mandai, "Efficient detection of mesial temporal sclerosis using hippocampus and CSF features in MRI images," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Inform. (BHI)*, Mar. 2018, pp. 178–181.

[18] M. Torres-Velazquez, G. Hwang, C. J. Cook, B. Hermann, V. Prabhakaran, M. E. Meyerand, and A. B. Mcmillan, "Multi-channel deep neural network for temporal lobe epilepsy classification using multimodal MRI data," in *Proc. IEEE 17th Int. Symp. Biomed. Imag. Workshops (ISBI Workshops)*, Apr. 2020, pp. 1–4.

[19] X. Si, X. Zhang, Y. Zhou, Y. Sun, W. Jin, S. Yin, X. Zhao, Q. Li, and D. Ming, "Automated detection of juvenile myoclonic epilepsy using CNN based transfer learning in diffusion MRI," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2020, pp. 1679–1682.

[20] J. Nalepa, M. Marcinkiewicz, and M. Kawulok, "Data augmentation for brain-tumor segmentation: A review," *Frontiers Comput. Neurosci.*, vol. 13, pp. 1–18, Dec. 2019.

[21] S. Sakib, T. Tazrin, M. M. Fouda, Z. M. Fadlullah, and M. Guizani, "DL-CRC: Deep learning-based chest radiograph classification for COVID-19 detection: A novel approach," *IEEE Access*, vol. 8, pp. 171575–171589, 2020.

[22] I. Santoso, Y. Adrianto, A. Sensusiati, D. Wulandari, and I. Purnama, "Epileptic EEG signal classification using convolutional neural network based on multi-segment of EEG signal," *Int. J. Intell. Eng. Syst.*, vol. 14, no. 3, pp. 160–176, Jun. 2021.

[23] F. Renard, S. Guedria, N. D. Palma, and N. Vuillerme, "Variability and reproducibility in deep learning for medical image segmentation," *Sci. Rep.*, vol. 10, no. 1, p. 13724, Dec. 2020.

[24] H. Alhichri, "CNN ensemble approach to detect COVID-19 from computed tomography chest images," *Comput., Mater. Continua*, vol. 67, no. 3, pp. 3581–3599, 2021.

[25] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[26] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1026–1034.

[27] Y. Yuan, G. Xun, K. Jia, and A. Zhang, "A multi-view deep learning framework for EEG seizure detection," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 1, pp. 83–94, Sep. 2019.

[28] X. Glorot, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, vol. 9, 2010, pp. 249–256.
[29] V. Dumoulin and F. Visin, "A guide to convolution arithmetic for deep learning," 2016, *arXiv:1603.07285*.
[30] J. Nagi, F. Ducatelle, G. A. Di Caro, D. Ciresan, U. Meier, A. Giusti, F. Nagi, J. Schmidhuber, and L. M. Gambardella, "Max-pooling convolutional neural networks for vision-based hand gesture recognition," in *Proc. IEEE Int. Conf. Signal Image Process. Appl. (ICSIPA)*, Nov. 2011, pp. 342–347.
[31] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
[32] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[33] L. Rokach, "Ensemble-based classifiers," *Artif. Intell. Rev.*, vol. 33, nos. 1–2, pp. 1–39, 2010.
[34] D. Decoste and B. Schölkopf, "Training invariant support vector machines," *Mach. Learn.*, vol. 46, nos. 1–3, pp. 161–190, 2002.
[35] C.-S. Lo and C.-M. Wang, "Support vector machine for breast MR image classification," *Comput. Math. Appl.*, vol. 64, no. 5, pp. 1153–1162, Sep. 2012.
[36] Y. Jiao and P. Du, "Performance measures in evaluating machine learning based bioinformatics predictors for classifications," *Quant. Biol.*, vol. 4, no. 4, pp. 320–330, Dec. 2016.
[37] J. D. Rodriguez, A. Perez, and J. A. Lozano, "Sensitivity analysis of K-fold cross validation in prediction error estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 569–575, Mar. 2010.
[38] S. R. Dubey, S. Chakraborty, S. K. Roy, S. Mukherjee, S. K. Singh, and B. B. Chaudhuri, "DiffGrad: An optimization method for convolutional neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 11, pp. 4500–4511, Nov. 2020.
[39] C. F. Sabottke and B. M. Spieler, "The effect of image resolution on deep learning in radiography," *Radiol., Artif. Intell.*, vol. 2, no. 1, Jan. 2020, Art. no. e190015.
[40] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.

IRWAN BUDI SANTOSO received the bachelor's degree from the Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia, in 2000, and the master's degree from the Department of Informatics Engineering, Institut Teknologi Sepuluh Nopember, in 2007. He is currently pursuing the Ph.D. degree with the Department of Electrical Engineering, Institut Teknologi Sepuluh Nopember. He is also a Lecturer with the Department of Informatics Engineering, Universitas Islam Negeri Maulana Malik Ibrahim Malang. His research interests include artificial intelligence, machine learning, and computer vision.

YUDHI ADRIANTO received the bachelor's degree from the Faculty of Medicine, Universitas Airlangga, Surabaya, Indonesia, in 2006, and the master's/specialty degree from the Universitas Airlangga, in 2013. He is currently a Lecturer with the Department of Neurology, Medical Faculty, Airlangga University (Universitas Airlangga). His research interests include neurointervention and neuroimaging. He is a member of the Division of Neurointervention & Neuroimaging Rumah Sakit Universitas Airlangga (Airlangga University Hospital), a member of the Korean Society of Radiology (KSR), a member of the ASEAN Neuro-Intervention Association (ANIA), and a fellow of the Interventional Neurology ASEAN (FINA).

ANGGRAINI DWI SENSUSIATI received the bachelor's degree from the Department of Medicine, Universitas Airlangga, Surabaya, Indonesia, in 1987, and the master's/specialty and Ph.D. degrees from the Universitas Airlangga, in 1996 and 2013, respectively. She is currently a Lecturer with the Department of Radiology, Medical Faculty, Airlangga University (Universitas Airlangga). She is also a Radiology Specialist at Husada Utama Hospital, Siloam Hospital, and Airlangga University Hospital, Surabaya. Her research interests include radiology (MRI of the brain) and neuroimaging. She is a Founding Member of the Association Cardiac Imaging and a member of AONHCR.

DIAH PUSPITO WULANDARI received the bachelor's degree from the Department of Electrical Engineering, Institut Teknologi Bandung, Indonesia, in 2004, the master's degree from the School of Informatics, The University of Edinburgh, U.K., in 2006, and the Ph.D. degree from the Department of Electrical Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia, in 2016. Since 2005, she has been a Lecturer with the Department of Computer Engineering, Institut Teknologi Sepuluh Nopember. Her research interests include signal processing and artificial intelligence.

I. KETUT EDDY PURNAMA (Member, IEEE) received the bachelor's degree from the Department of Electrical Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia, in 1994, the master's degree from the Department of Informatics Engineering, Institut Teknologi Bandung, Indonesia, in 1999, and the Ph.D. degree from the Department of Biomedical Engineering, University of Groningen, The Netherlands, in 2007. He is currently a Lecturer with the Department of Electrical Engineering and the Department of Computer Engineering, Institut Teknologi Sepuluh Nopember. His research interests include medical image processing and analysis, artificial intelligence in medicine, medical informatics, and medical data visualization.