Classifying Electrocardiograms through Empirical Mode Decomposition and Transfer Learning

Yi Da Kang, Jackleen Atallah, Yujing Duan, Siyu Zhang

***Abstract*—Early detection as well as efficient and speedy medical treatment of people suffering from cardiovascular problems necessitates the multiclass classification of cardiovascular diseases from Electrocardiogram (ECG) signals. In this project, the described task is solved by signal processing and classification. Signal processing involves two steps: raw ECG data segmentation and feature extraction. Empirical Mode Decomposition (EMD) is adopted to generate Intrinsic Mode Functions (IMFs) which are then used to construct feature encodings of the ECG signals. As for classification, ResNet which is a deep convolutional neural network (CNN) is employed as the pre-trained model for transfer learning.**

***Index Terms*—Preprocessing, EMD, CNN, VGG16, ResNet34 Transfer Learning**

I. Introduction

Cardiovascular diseases are one of the most fatal reasons that cause death. In recent years, cardiovascular issues have emerged as the leading cause of death for millions of people worldwide. Because of high prevalence and high mortality rates, early detection and active prevention the most important job people working on nowadays. Multiple cardiovascular disease classification from Electrocardiogram (ECG) signal enables early detection of abnormal heart signal which is crucial for medical prevention and treatment. Therefore, machine learning kicks in to facilitate and enhance the performance of such classification problem.

Instead of carrying out classification task on raw ECG data, pre-processing of raw ECG is the key to denoise the signals and improve the overall performance of the network. There are some popular techniques such as Kalman filtering method [1], Bayesian filtering [2], Fourier decomposition wavelet [3], Empirical Mode Decomposition (EMD) [4], and so on. Among these alternatives, EMD is an outstanding choice since it demonstrates remarkable ability to deal with nonlinear and nonstationary signals which have sharply varying dynamics. In contrast to Fourier Transforms and wavelet decomposition, EMD leverages Intrinsic Mode Functions (IMFs) to extract inherent features of a signal. With this method, a signal can be denoised in the EMD time domain by discarding lower-order IMF signals based on the assumption that noise and signal are unrelated from the perspective of frequency bands [5]. These properties enable EMD to be reasonable and suitable for dealing with the ECG signals.

Inspired by [6], this project has built a convolutional neural network (CNN) based on VGG16. The overarching strategy of this project can be divided into three stages. In the first stage, plain ECG data is going to be segmented into pre-determined lengths. Next is the feature extraction stage. Segmentations of ECG signals are processed using EMD. Selective Intrinsic Mode Functions obtained from EMD are utilized to create 2-D global feature encodings which are more distinct and representative comparing to original ECG data. The classification is carried out in the last stage. 2-D CNN model is adopted to classify the heartbeat condition based on global features captured from previous stage. Transfer learning is employed in this stage with VGG16 and ResNet34 for the model construction. The filtered output from previous stages would be forwarded to the CNN. Transfer learning is applied and the fully-connected layers right before the output are modified and adapted according to the input data. The database used in this project is MIT-BIH arrhythmia database which are raw ECG signals.

II. System Design

In this section, the steps for solving the multiclass classification of raw ECG signals are detailed, including the significant solutions tested in the experiment and methods for tuning the hyperparameters. As previously mentioned, the project is divided into three parts, with the first part focusing on ECG pre-processing. The signal is then passed through a Bandpass Filter, Notch Filter to reduce powerline noise(60 Hz).
The second part involves using Empirical Mode Decomposition (EMD) to extract intrinsic mode functions (IMFs), highlighting significant features inherent in the ECG data. In this project, only the first and second IMFs are used to generate the 2D feature map. The lower numbered IMF contain higher frequency component from the original signal. Therefore, IMF1 and IMF2 are the most informative and are able to form distinct and distinguishable feature maps.

Aside from feature map, the Hilbert-Huang transformation (HHT) is applied to IMFs as another alternative, forming a comprehensive time-frequency representation. To classify the feature images of ECG signals into four classes: N, LBBB (L), RBBB (R), and P, Convolutional Neural Networks, specifically VGG16 and ResNet34, are used for transfer learning to train the model using MIT-BIH arrhythmia database as the raw ECG data.

The **first dataset** comprises 2D feature maps that involves IMF1 and IMF2. The **second dataset** contains the images which are the transformation of ECG signals into 2D Hilbert Spectrum using HHT. In HHT, Hilbert transform is combined with EMD so that ECG signals are broken down into cleaner components which allows effective feature extraction from ECG data comparing to raw data. This method is also able to tackle the non-linear and non-stationary nature of ECG signals.

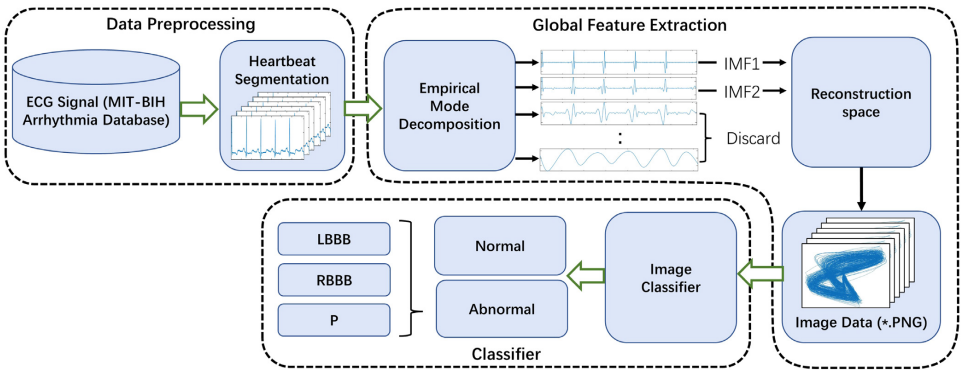


Fig. 1: Flow chart of the classification method proposed in the original study.

Both of two datasets are splited into three subsets: training, validation, and testing. Both validation and testing dataset account 20% of the total dataset. The flow chart in Fig.1 reveals the general steps of the proposed approach.

A. ECG pre-processing

ECG pre-processing was conducted using the MIT-BIH arrhythmia database to assess the efficacy of our suggested methodology. This database comprises ECG recordings from 48 participants, with each recording having a sampling rate of 360 Hz and a duration of roughly 30 minutes. It encompasses 15 varieties of cardiac rhythms. Normal heartbeats constituted approximately 70% of the database, while the remaining 30% consisted of abnormal heartbeats. The four types of heartbeats with the highest proportions were selected for study from the 48 samples in the database, namely Normal heartbeat (N), Left Bundle Branch Block heartbeat (L), Right Bundle Branch Block heartbeat (R), and Paced heartbeat (P).

B. Extracting global features

The Empirical Mode Decomposition (EMD) method, introduced by Huang, is widely utilized for examining data characterized by nonlinearity and non-continuity features. EMD can break down a signal into a series of Intrinsic Mode Functions (IMFs), with each IMF correlating to a distinct frequency scale in the signal. This method stands out from other techniques like the Fourier transform or wavelet decomposition, which rely on predefined basis functions because EMD adapts to the data without requiring any basis. It operates on the principle that signal and noise are not correlated within the frequency spectrum, allowing for noise removal within the EMD time domain by omitting IMFs of lower order. The EMD process involves extracting an IMF through what is known as the sifting process, which must meet two criteria: first, the number of local extrema and the number of zero-crossings should either be the same or differ by no more than one; and second, at every point, the mean of the envelope delineated by the local maxima and minima should be zero.

C. Feature extraction from 2-D ECGs

Global characterization of ECG data was accomplished using EMD to isolate high-frequency elements and preserve essential ECG signal information. After EMD analysis, it was observed that the initial two IMFs significantly mitigate noise and interference. These are then utilized to synthesize feature images by plotting two IMFs against each other, one on the x-axis and the other on the y-axis. These resulting feature images distinctively outline the unique forms corresponding to each data type. Upon visual analysis, these images simplify the task of distinguishing various abnormality classes compared to the unprocessed signals. Although phase-space representations can similarly illustrate quasi-periodicity, they are often challenging to interpret due to their complex, high-dimensional nature. The feature visualization technique we implemented bears a resemblance to phase-space reconstructions applied to IMF components, where our visual examination suggests the presence of recurrent trajectories within the phase space.

D. Classifier Using Transfer Learning

A convolutional neural network (CNN) is widely used in various applications such as image and video recognition, image classification, and natural language processing. For image classification tasks, there are several robust models like LeNet [7], AlexNet [8], VGGNet [9], and ResNet [10]. A typical CNN functions similarly to a deep neural network and consists of an input layer, one or multiple hidden layers, and an output layer. The hidden layers comprise a combination of convolutional layers and different types of pooling layers such as max pooling and average pooling. Additionally, the kernel size in convolutional layers can vary, typically offering options like 3x3, 5x5, and 7x7. These layers collaborate to identify key features and reduce the spatial dimension of the image data, which potentially speeds up the learning process and improves the performance of the model. Typically, fully connected layers are positioned towards the end of the hidden layers. However, before the input reaches these fully connected layers, it undergoes a transformation where it is flattened. This means the multidimensional data is converted into a vector before being passed to the fully connected layers. Once in 1-D form, the data flows through the fully connected layers. These layers aim to learn the various patterns and features previously extracted from the data. By synthesizing this information, the fully connected layers play a crucial role in analyzing complex

Layer	input and out channel	Kernel/Stride/Padding
Conv2D + ReLu	3 / 64	3/ 1 / 1
Conv2D + ReLu	64 / 64	3 / 1 / 1
MaxPool2D	-	2 / 2 / 0
Conv2D + ReLu	64 / 128	3 / 1 / 1
Conv2D + ReLu	128 / 128	3 / 1 / 1
MaxPool2D	-	2 / 2 / 0
Conv2D + ReLu	128 / 256	3 / 1 / 1
Conv2D + ReLu	256 / 256	3 / 1 / 1
Conv2D + ReLu	256 / 256	3 / 1 / 1
MaxPool2D	-	2 / 2 / 0
Conv2D + ReLu	256 / 512	3 / 1 / 1
Conv2D + ReLu	512 / 512	3 / 1 / 1
Conv2D + ReLu	512 / 512	3 / 1 / 1
MaxPool2D	-	2 / 2 / 0
Conv2D + ReLu	512 / 512	3 / 1 / 1
Conv2D + ReLu	512 / 512	3 / 1 / 1
Conv2D + ReLu	512 / 512	3 / 1 / 1
MaxPool2D	-	2 / 2 / 0
AvgPool	-	7 / - /-
Full Connect	25088 / 4096	-
Full Connect	4096 / 4096	-
Full Connect	4096 / 4	-

TABLE I: Detailed parameters for VGG16 used in this project

patterns and producing accurate classification outcomes. This process is essential for transforming the high-level features learned by the network into final predictions, thereby effectively enabling the model to distinguish between different classes.

In our project, using the pertained model VGG16, ResNet34 Model both of them trained on the ImageNet dataset to classify 1000 classes we freeze the models except the last fully-connected layers to retain their learned features. In order to better adapt our problem which has four classes, we modified (fine-tuning):

- VGG16: replaced the original classifier with a simpler one. This new classifier consists of a flattening layer and a single linear layer designed to output 4 classes as shown in Table.I.
- ResNet34: The fully-connected layer of the model is replaced with a new sequence of layers: a flattening layer, a linear layer reducing features from 512 to 128, and another linear layer that maps these 128 features to 4 classes for our problem. The detailed layers implemented in this model are shown in Table.II.
  Residual Blocks are the fundamental components of ResNet architectures and consist of two main parts, a sequence of convolutional layers (each followed by Batch Normalization and ReLU activation function), and a distinctive skip connection that adds the input of the block directly to its output. This design helps in addressing the vanishing gradient problem, allowing for effective training of deep neural networks. Residual Blocks are designed to learn residual functions regarding the layer inputs, facilitating the training of deeper network architectures by providing pathways for gradients during back-propagation.

Layer Type	Input and out Channel	Kernel/Stride/Padding
Conv2D + BN + ReLU	3 / 64	7 / 2 / 3
MaxPool2D	-	3 / 2 / 1
Residual Block1 x3	64 / 64	3 / 1 / 1
Residual Block2 x4	64 / 128	3 / 2 / 1
Residual Block3 x6	128 / 256	3 / 2 / 1
Residual Block4 x3	256 / 512	3 / 2 / 1
AdaptiveAvgPool2D	-	-
Flatten	512	-
Fully Connected	512 / 128	-
Fully Connected	128 / 4	-

TABLE II: Detailed parameters for ResNet34 used in this project

E. Model Training

Since the CNN models for the project have been chosen, the primary approach involves utilizing the models:

1) First and second approach focuses on training VGG16 with the two datasets.
2) Third and fourth approach focuses on training ResNet34 with the two datasets.

For each training session, hyperparameters including batch size, learning rate, and epochs are carefully tuned to optimize performance. Batch sizes of 16, 32, and 64 were experimented with to balance computational speed and model effectiveness. Smaller batch sizes tend to improve generalization, while larger batches speed up computation, but may lead to suboptimal solutions. Regarding the learning rate, values of 0.001, 0.002, 0.005, and 0.01 were evaluated to determine the appropriate increments for efficient learning without overlooking important minima. Lower learning rates ensure more stable convergence but may slow down training, whereas higher rates accelerate the process but potentially overlook the best solutions. For epochs, a maximum of 40 was set, incorporating an early stopping mechanism with a patience parameter of 5. This approach aims to prevent overfitting by stopping training if the validation loss does not improve after five epochs. It makes sure the model doesn't react to temporary spikes from noise, keeping predictions clear and not overly complex.

III. Results

The feature maps for IMF1 vs IMF2 and also the HHT of four different heartbeat classes are displayed in TableIII. For normal heartbeats like record 101, Paced heartbeats record 102. Right Bundle Branch Block (RBBB) record 118 while Left Bundle Branch Block (LBBB) record 109.

The hyperparameter tuning are carried out based on the performance of each model with respect to validation set. The optimal learning rate and batch size pair is selected from the trial which returns the highest validation accuracy with low validation loss. The results of hyperparameters tuning for four approaches are shown in Fig.IV, Fig.V, Fig.VI, and Fig.VII. According to the performance on validation dataset, the optimal hyperparameter pair are chosen and they are highlighted in the each chart.

After figuring out the optimal hyperparameter pair for each approach, the testing is carried out to check their performance

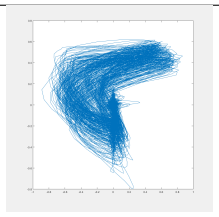
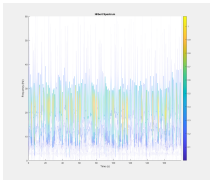
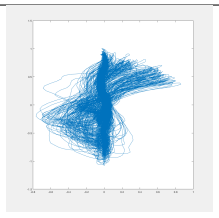
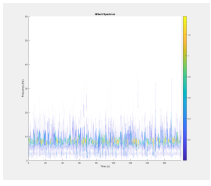
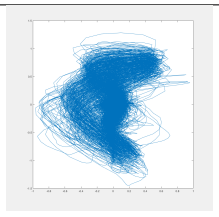
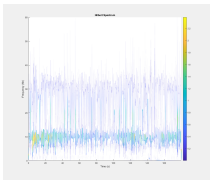
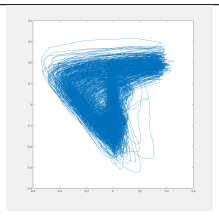
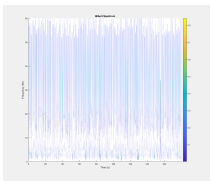
Dataset	IMFs	HHT
N		
L		
R		
P		

TABLE III: The feature maps for (IMF1,IMF2) and HHT of four different heartbeat classes

Learning Rate	Batch Size	Validation Loss	Validation Accuracy
0.001	16	0.2052	93.4426%
0.001	32	0.2308	90.1639%
0.001	64	0.3098	85.2459%
0.002	**16**	**0.2624**	**95.0820%**
0.002	32	0.2754	91.8033%
0.002	64	0.4834	83.6066%
0.005	16	0.3717	88.5246%
0.005	32	0.3042	95.0820%
0.005	64	0.8662	88.5246%
0.01	16	1.2630	93.4426%
0.01	32	0.9968	91.8033%
0.01	64	0.7146	90.1639%

TABLE IV: VGG16 with IMF1 and IMF2

Learning Rate	Batch Size	Validation Loss	Validation Accuracy
0.001	16	0.4246	85.4839%
0.001	32	0.4368	85.4839%
0.001	64	0.3766	85.4839%
0.002	16	0.4036	83.8710%
0.002	**32**	**0.3807**	**88.7097%**
0.002	64	0.3873	88.7097%
0.005	16	0.3583	87.0968%
0.005	32	0.4285	83.8710%
0.005	64	0.4193	80.6452%
0.01	16	0.6922	83.8710%
0.01	32	0.4955	87.0968%
0.01	64	1.3772	69.3548%

TABLE V: VGG16 with HHT

Learning Rate	Batch Size	Validation Loss	Validation Accuracy
0.001	16	0.2641	90.1639%
0.001	32	0.2896	86.8852%
0.001	64	0.2821	86.8852%
0.002	**16**	**0.2593**	**91.8033%**
0.002	32	0.2489	86.8852%
0.002	64	0.3016	86.8852%
0.005	16	0.4329	85.2459%
0.005	32	0.2671	90.1639%
0.005	64	0.5597	73.7705%
0.01	16	0.2692	88.5246%
0.01	32	0.3206	85.2459%
0.01	64	0.2831	86.8852%

TABLE VI: ResNet with IMF1 and IMF2

Learning Rate	Batch Size	Validation Loss	Validation Accuracy
0.001	16	0.4496	87.0968%
0.001	32	0.3954	85.4839%
0.001	64	0.3927	83.8710%
0.002	16	0.5029	87.0968%
0.002	32	0.4088	85.4839%
0.002	**64**	**0.4073**	**88.7097%**
0.005	16	0.4628	87.0968%
0.005	32	0.4652	87.0968%
0.005	64	0.4118	80.6452%
0.01	16	0.4945	85.4839%
0.01	32	0.4519	85.4839%
0.01	64	0.4443	87.0968%

TABLE VII: ResNet with HHT

on test dataset. The results of each approach are gathered in Fig.VIII. The method which involves ResNet34 with 2D feature maps of IMF1 and IMF2 generated using EMD demonstrate outstanding testing accuracy comparing to other alternatives, which achieves about 96%. During the testing, it is observed that although modified VGG16 model is quite effective at identifying N heartbeats, this model demonstrates a reduced capacity to differentiate between types of abnormal heartbeats. The model particularly struggles with LBBB heartbeats. This is a sign that the model probably overfits the normal heart signals. The reason for this phenomenon might be the imbalance datapoints in the training dataset, or the feature representation is insufficient for VGG16. As for data preprocessing, feature maps generated using IMF1 and IMF2 remarkably outperforms HHT. Even though HHT involves EMD and is also designed for nonlinear and nonstationary signal, the frequency-time distribution of signal amplitude can not capture the uniqueness of every ECG signal as illustrated in TableIII. Instead of just using EMD to perform global characterization and extract high-frequency components, 2D feature images which combines the first two IMFs can show the unique shape for each ECG signal. From visualization in TableIII, it is easier to identify different classes of abnormalities comparing to the frequency-time domain HHT images.

Model	hyperparameters	Test Loss	Test Accuracy
VGG16 + IMF1,2	0.002, 16	0.4771	89.4737%
VGG16 + HHT	0.002, 32	0.4685	83.1169%
ResNet + IMF1,2	**0.002, 16**	**0.2468**	**96.0526%**
ResNet + HHT	0.002, 64	0.4684	84.4156%

TABLE VIII: Comparison between four different approaches

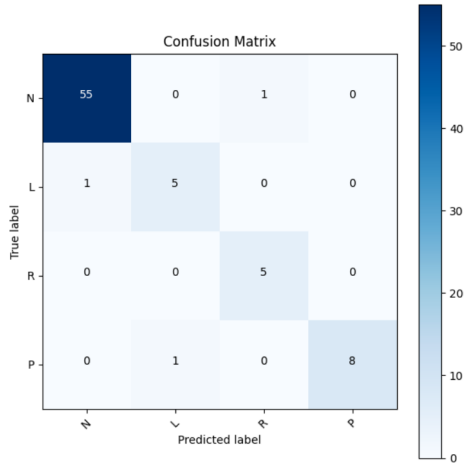


Fig. 2: Confusion Matrix for ResNet for feature map IMF1 vs IMF2

		Predicted				
		N(True)	N(False)		L(True)	L(False)
True Label	N(True)	55	1	L(True)	5	1
	N(False)	1	19	L(False)	1	69
		R(True)	R(False)		P(True)	P(False)
	R(True)	5	0	P(True)	8	1
	R(False)	1	70	P(False)	0	67

TABLE IX: Transformed confusion matrices for four types of heart signal separately

The confusion matrix for the best approach, ResNet with IMF1 and IMF2, is presented in Fig.2 to better visualize and summarize the performance of this approach. The true cases (true N, true L, true R, and true P) along the diagonal of the matrix outnumber other cases. There are just three misclassifications in total. Therefore, The overall classification accuracy on four types of heart signals is 96.1%.

The overall confusion matrix can be transformed into four 2 by 2 matrices, each focusing on the performance of classifier on signals N, L, R, and P, respectively as shown in TableIX. Through this method, the functionality of classifier can be further evaluated more detailedly. True Positive (TP) means the number of positive examples correctly predicted. False Negative (FN) is the number of positive example wrongly predicted as negative. False Positive (FP) is the number of negative examples wrongly predicted as positive. True Negative (TN) is the number of negative examples correctly predicted. Sensitivity is calculated from TP/(TP+FN). Precision is calculated using TP/(TP+FP). Accuracy is obtained from (TP+TN)/(TP+FN+FP+TN). Lastly, F1 score is calculated using 2TP/(2TP+FP+FN). The evaluation of the classifier performance on different types of signal based on the metrics mentioned above is shown in TableX.

Metric	N	L	R	P
Sensitivity	98.21%	83.33%	100%	88.99%
Precision	98.21%	83.33%	83.33%	100%
Accuracy	97.37%	97.37%	98.68%	98.68%
F1 Score	98.21%	83.33%	90.91%	94.12%

TABLE X: Metrics comparison between the performance of classifier on 4 types of heart signals

IV. Concluding Remarks

Most of conventional classification schemes carry out feature learning based on the raw ECG signal. In this project, data processing is adopted before feeding signals to the neural network. Two data processing methods are used in this project. The first one is Hilbert Huang transform. The second method is converting one dimensional ECG signal into two dimensional feature images using the first two IMFs generated through EMD. On the other hand, two models are used for transfer learning: VGG16 and ResNet34. From the experiment results, ResNet34 with IMFs offers the best test accuracy which is around 96%. It is obvious that with proper data preprocessing for denoising and feature extraction, the performance of the network is largely improved. For future work, data augmentation would help us to address the issue of imbalance data. With generating more varied samples for(L, R, P) classes, it could help to balance the dataset and potentially improve the model's ability of generalization. Thus, the testing performance have the space to be further improved.

References

[1] O. Sayadi and M. B. Shamsollahi, "Ecg denoising and compression using a modified extended kalman filter structure," *IEEE transactions on biomedical engineering*, vol. 55, no. 9, pp. 2240–2248, 2008.

[2] R. Sameni, M. B. Shamsollahi, C. Jutten, and G. D. Clifford, "A nonlinear bayesian filtering framework for ecg denoising," *IEEE transactions on Biomedical Engineering*, vol. 54, no. 12, pp. 2172–2185, 2007.

[3] A. Singhal, P. Singh, B. Fatimah, and R. B. Pachori, "An efficient removal of power-line interference and baseline wander from ecg signals by employing fourier decomposition technique," *Biomedical Signal Processing and Control*, vol. 57, p. 101741, 2020.

[4] N. I. Hasan and A. Bhattacharjee, "Deep learning approach to cardiovascular disease classification employing modified ecg signal from empirical mode decomposition," *Biomedical signal processing and control*, vol. 52, pp. 128–140, 2019.

[5] B. Weng, M. Blanco-Velasco, and K. E. Barner, "Ecg denoising based on the empirical mode decomposition," in *2006 international conference of the IEEE engineering in medicine and biology society*. IEEE, 2006, pp. 1–4.

[6] Y. Li, J.-h. Luo, Q.-y. Dai, J. K. Eshraghian, B. W.-K. Ling, C.-y. Zheng, and X.-l. Wang, "A deep learning approach to cardiovascular disease classification using empirical mode decomposition for ecg feature extraction," *Biomedical Signal Processing and Control*, vol. 79, p. 104188, 2023.

[7] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, ser. NIPS'12. Red Hook, NY, USA: Curran Associates Inc., 2012, p. 1097–1105.

[9] S. Liu and W. Deng, "Very deep convolutional neural network based image classification using small training sample size," in *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, 2015, pp. 730–734.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.