Multi-Class Heart Abnormalities Detection Based on ECG Graph Using Transfer Learning Method

Sugondo Hadiyoso, Suci Aulia, and Indrarini Dyah Irawati
School of Applied Science, Telkom University
Jl. Telekomunikasi No. 1, Terusan Buah Batu, Bandung, 40257
e-mail: sugondo@telkomuniversity.ac.id

Abstract—The heart is one of the vital organs in the circulatory system. Regular checkups are very important to prevent heart disease. The most basic examination is blood pressure then further examination is related to the evaluation of the electrical activity of the heart using an electrocardiogram (ECG). The ECG carries important information regarding various abnormalities of heart function. Several automated classification techniques have been proposed to facilitate diagnosis. However, not all digital ECG devices provide raw data for analysis. ECG classification method based on images can be an alternative in classification. Therefore, in this study, it is proposed to classify ECG based on signal images. The proposed classification method uses transfer learning with VGG, AlexNet, and DenseNet architectures. The method used for the classification of multi-class ECG consists of normal, PVC, atrial fibrillation, AFL, Bigeminy, LBBB, and APB. The simulation results generate the best accuracy of 92% and F1-score of 92%. Best performance is achieved using DenseNet architecture at 60 epochs. This study is expected to be a new reference technique in the classification of ECG signals.

Keywords: *transfer learning, heart abnormalities, ECG, image*

I. INTRODUCTION

ECG records the electrical activity produced by the heart throughout time. The foundations of this signal's cycle, i.e., the morphological sequence of its distinguishing waves: P, QRS complex, T, and U, can be understood to access it. The atrial myocardium's depolarization causes the P wave, followed by the QRS-complex, a rapidly spiking wave that induces ventricular contraction. The T wave, which de-notes the beginning of a new cycle with the repolarization of the ventricular myocardium, marks the cycle's conclusion. The U could signify the ventricular myocardium's late depolarization [1]. While sustaining this sequence, a heart rate between 60 to 100 beats per minute (bpm) at rest is required for the Signal-to-Noise Ratio (SNR). In our previous work, we proposed a biometric authentication based on ECG signals using time series analysis and Support Vector Machine (SVM) [2]. As a continuation, we also tried ECG-based biometric classification using Ensemble Empirical Mode Decomposition (EEMD) and Variational Mode Decomposition (VMD) methods, and our simulations show that the suggested strategy works and can attain an accuracy of up to 98.2% [3]. Recently, we proposed an abnormal ECG classification using Empirical Mode Decomposition (EMD) and Entropy [4]. The decomposed signal was utilized to determine the Shannon and Fuzzy entropies, which constitute the foundation of the feature extraction method. The three forms of simulated ECG signals are left bundle branch block, atrial fibrillation, and

regular sinus rhythm. The suggested method's validation performance was tested using the support vector machine and the k-Nearest Neighbor algorithm. According to the test results, the most accurate measurement is 81.1 %, with sensitivity and specificity measuring 89.8% and 79.4%, respectively. However, ECG devices generally do not provide access to retrieve the raw signal. This problem makes it difficult to develop automatic classification applications if the computation is based on time series data. Therefore, an alternative approach is needed in the ECG classification by analyzing the signal image.

On the other hand, deep learning is currently the most widely used by researchers to convey their ideas or concepts [5]. One of the techniques that sparked the development of deep learning was CNN. Regarding image classification, CNN accepts the input image or data, processes it, and then assigns it to one of several classes [6]. Many convolutional neural networks (CNNs) models operate by constructing a pre-trained model that is simple to set up and requires little prior processing. The model is trained using libraries with weights made up of millions of images before being applied to actual data. This procedure, namely as a transfer learning method [7]. The architectures used by CNN include VGGNet, Alexnet, DenseNet, ResNet, and others [8][9]. AlexNet architecture is a CNN model that is commonly used. Because AlexNet is a sizable CNN framework with a significant learning capability, such colon classification achieved 91.47% accuracy [10]. Another CNN architecture known as VGG16 is frequently employed in image classification

or recognition. It produces excellent accuracy in various instances of image classification [11]-[12]. While DenseNet was chosen because it has positive significance for applications in medical image analysis [12].

This study proposes a multi-class heart abnormalities detection based on ECG signal images using the transfer learning method. We employed some transfer learning methods to evaluate the best result. They are DenseNet, VGG16, and AlexNet. This essay describes the literature on transfer learning that includes CNN, DenseNet, AlexNet, and VGG-19 in Section 2. Finally, the experimental result of multi-class heart abnormalities detection based on ECG for all methods is briefly discussed in Section 3 and the conclusion is provided in Section 4.

II. MATERIAL AND METHODS

A. ECG Dataset

The ECG signals were received for research purposes from the MIT-BIH Arrhythmia database via the PhysioNet service. There were 45 patients with ECG signals: 19 females (ages 23 to 89) and 26 males (ages 32-89). There were 17 classifications of ECG signals, including 15 different types of cardiac dysfunctions, pacemaker rhythm, and normal sinus rhythm. This work used a gain of 200 adu/mV and used a sampling rate of 360 Hz to record every ECG signal. For the analysis, 1000 pieces of the ECG signal, each lasting 10 seconds and containing 3600 samples, were randomly chosen. Mat formatted data (Matlab). In this study, only seven classes from all 17 classes consider the common forms of the disease, as seen in Figure 1.

B. Convolutional Neural Network

The CNN architecture consists of four phases: convolutional, max pooling, flattering, and full connection, as shown in Figure 2. Numerous feature maps are constructed in the first step convolutional procedure to employ detector features [13]. Next, the pooling layer creates new filters based on rules, whereas the convolutional layer is utilized to extract features. Finally, at the output step, an utterly connected layer of several neurons serves as a decision-making structure [14].

C. Dense Convolutional Networks (DenseNet)

Huang and colleagues proposed the DenseNet architecture in 2018 [15]. Although the vanishing gradient problem is addressed, this network also promotes feature reuse. It can distinguish between information that has been kept and information that has been added. The gradient from the source image and the loss function is directly accessible to each layer [16]. The DenseNet enables even more advanced network training. Fewer parameters are needed because redundant features don't need to be learned. DenseNet merges features using a concatenation

operation. Before a 3x 3 convolutional layer, each layer reads the state and obtains output to reduce the number of layers from the previous layers [16].

D. Alexnet

CNN is a capable deep learning classification technique that can encode specific attributes or features from input images at various feature levels. AlexNet architecture is a CNN model that is commonly used. This design saves time, can avoid over-fitting, and is strong enough to learn several polyp traits. The first convolutional layer applies 96 11 x 11 3 kernels to the 256 x 256 3 input image to
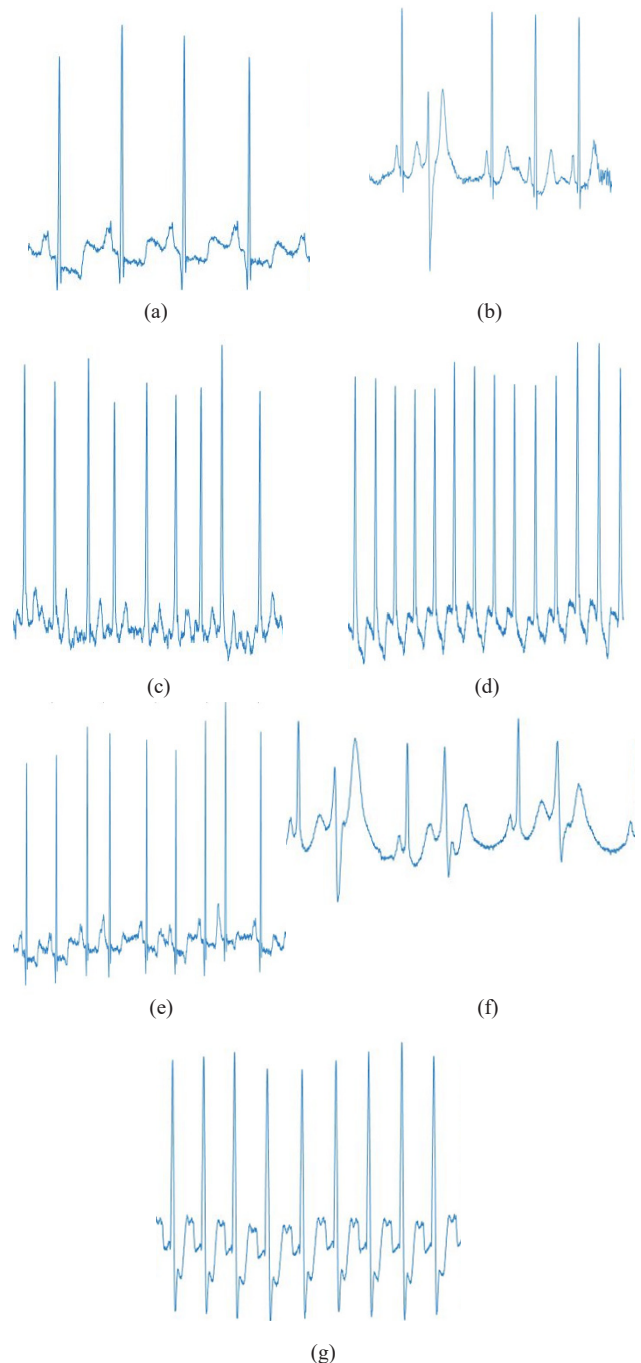

Figure 1. The 7 classes of ECG signal images used in this study (a) Normal (b) PVC (c) AFIB (d) AFL (e) APB (f) Bigeminy (g) LBBB

Table 1. Confusion Matrix

		Predicted	
		Negative	Positive
Actual	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

filter it. The output of the first convolution layer has been normalized and aggregated and then serves as the input for the second convolution layer. Then the second convolutional layer works to filter through 256 kernels of size 5548, then connected to 384 kernels of size 3x3 256 in the third convolutional layer. The fourth convolutional layer contains 384 3x3 192 kernels, while the fifth convolutional layer contains 256 3x3 192 kernels. Each completely connected layer contains 4096 neurons, as seen in Figure 3 [17].

E. VGG-16

Simonyan and Zisserman originally introduced VGG16 in 2014 [18]. The VGG16 improved AlexNet architecture, reducing kernel filter size to 3x3. Figure 4 shows the VGG16's fundamental design and depicts the model parameters employed in this investigation, and Table 1 presents the steps involved in creating each layer of the VGG-16 model. During the training phase, a two-dimensional convolution with a 3x3 kernel size is used to process the input RGB image, which measures 224x224. Then, each layer undergoes the convolution process to produce a feature vector that serves as a decision-making parameter during the classification step.

A. Performance Evaluation

Parameters for evaluating model performance include accuracy, precision, recall and f1-score. The fundamental calculation for these parameters is provided in Table 2 and uses a confusion matrix. Positive data that is expected to be accurate is referred to as True Positive (TP), and negative data that is expected to be correct is referred to as True Negative (TN). False Positive (FP) refers to negative data anticipated to be positive, and False Negative (FN) refers to positive data predicted to be negative.

Accuracy is the ratio of correct predictions, positive or negative, to the overall data. Precision is True Positive (TP) with the amount of data predicted to be positive. The recall compares True Positive (TP) with the number of positive data. F1-score is the average of recall and precision. Equations 1, 2, 3 and 4 are the formulas for precision, recall, f1-score and accuracy[20].

$$Precision = \frac{TP}{TP + FP} \qquad (1)$$

$$Recall = \frac{TP}{TP + FN} \qquad (2)$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \qquad (3)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (4)$$

Then the TP, TN, FP, and FN of NSR calculated using equation (5).

$$TP = C1 \qquad (5)$$
$$FP = C2 + C3 + C4 + C5 + C6 + C7$$
$$TN = C9 + C10 + C11 + C12 + C13 + C14 + C16 +$$
$$C17 + C18 + C19 + C20 + C21 + C23 + C24 + C25 +$$
$$C26 + C27 + C28 + C30 + C31 + C32 + C33 + C34 +$$
$$C35 + C37 + C38 + C39 + C40 + C41 + C42 + C44 +$$
$$+C45 + C46 + C47 + C48 + C49$$
$$FN = C8 + C15 + C22 + C29 + C36 + C43$$

III. Result and Discussion

In this section, we simulated the performance classification of the proposed system based on precision, recall, f1-score dan accuracy parameters. The system model is run on google colab pro using TPU v2 with 64 GB High Bandwidth Memory (HBM). A total of 1233 ECG datasets consisting of 7 classes, including: AFIB with 115 images, AFL with 160 images, APB with 230 images, Bigeminy with 252 imag-es, LBBB with 240 images, NSR with 103 images and PVC with 133 images that were simulated at this testing stage. The image dataset is divided into 80% for training data and 20% for test

Table 2. Confusion Matrix of Multi-class ECG

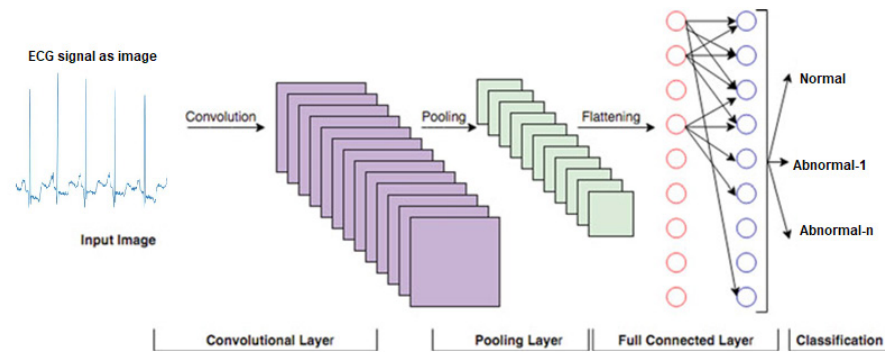|		Predicted						
		NSR	PVC	Bigeminy	LBBB	AFIB	AFL	APB
	NSR	C 1	C 2	C 3	C4	C 5	C 6	C 7
	PVC	C 8	C 9	C 10	C 11	C 12	C 13	C 14
	Bigeminy	C 15	C 16	C 17	C 18	C 19	C 20	C 21
Actual	LBBB	C 22	C 23	C 24	C 25	C 26	C 27	C 28
	AFIB	C 29	C 30	C 31	C 32	C 33	C 34	C 35
	AFL	C 36	C 37	C 38	C 39	C 40	C 41	C 42
	APB	C 43	C 44	C 45	C 46	C 47	C 48	C 49

Figure 2. The basic convolutional neural network architecture [13].
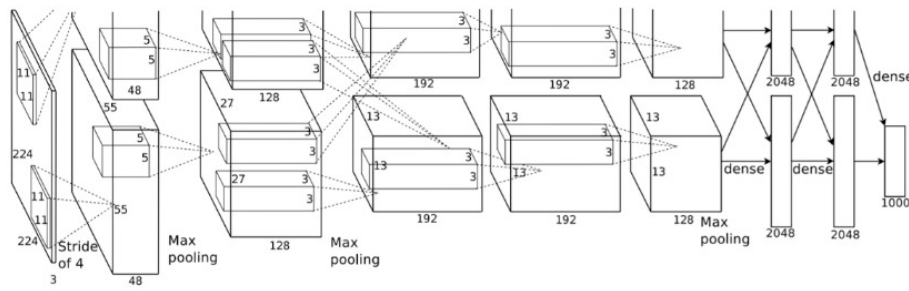

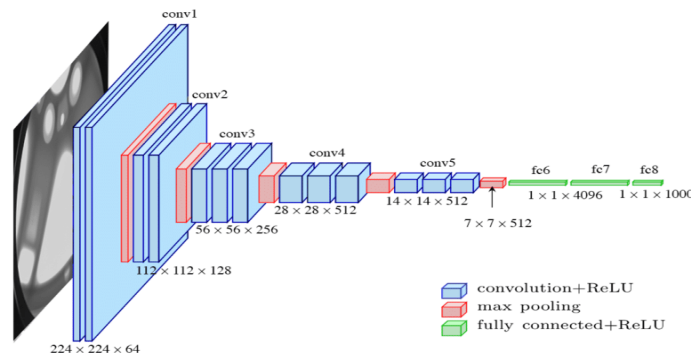Figure 3. The schematic of AlexNet architecture [18].

Figure 4. The VGG-16 architecture [20]

data. The augmentation process is applied to classes with a few datasets, i.e. the AFL, LBBB, Bigeminy and APB. The augmentation is used to overcome the data imbalance. The detailed parameters for image augmentation can be seen in Table 3. Figure 5 (a) illustrates the original signal while Figure 5 (b) shows the results of the augmentation. Tests were carried out using 3 CNN architectures, which consists of: DenseNet201, AlexNet and VGG. In the simulation, a learning rate parameter of 0.01 is added to avoid overfitting. The proposed system performance is improved by using Adam optimizers. In addition, ReLU and Softmax activation functions are also applied, with a dense layer using a value of 512. In the training process, the scenario is using batch size 32 with several epochs: 10, 20, 30, 40, 50 and 60.

Figure 6, Figure 7, and Figure 8 show the accuracy and loss curve pre-trained models. From the degenerate curve, it is known that the accuracy value tends to increase with increasing epoch. Meanwhile the training and validation loss curves show that the larger the epoch, the loss curve

tends to be zero. Validation training on DenseNet and VGG achieves the best accuracy compared to AlexNet. In this simulation, AlexNet generates the lowest training and validation accuracy of about 0.6.

From Figure 6 and Figure 7, the conditions with the best performance and steady state began to be reached at 40 epochs. The validation accuracy of DenseNet architecture shows the best performance compared to others. CNN architecture performance is evaluated based on parameters precision, recall, F1-score, and accuracy as shown in Table 4 for DenseNet, Table 5 for VGG, and Table 6 for AlexNet.

The precision matrix, the highest values, is obtained by DenseNet with 0.92, fol-lowed by VGG with 0.9 and the lowest one is AlexNet with 0.56. For Recall score, the highest performance is taken by DenseNet in 0.91, followed closely by VGG in 0.9 and AlexNet get poor in 0.62. On the other hand, F1-score for DenseNet is 0.92, VGG is 0.89 and AlexNet is 0.58. Finally, for the accuracy, the DenseNet gets 0.92, while VGG reaches 0.89 and AlexNet receives 0.63. From all epoch test, the Dense-Net

model has the highest value on all parameters, followed by the VGG and AlexNet models. The best performance of each model was obtained at 60 epoch for DenseNet, 40 epoch for VGG and 60 epoch for AlexNet.

Figure 9 represents the heatmap of confusion matrices for the test data with 20% testing and 80% training. This confusion matrix represents the highest accuracy achieved using DenseNet at 60 epochs. The proposed classifier accurately distinguished 7 classes with the average accuracy 0.92. Bigeminy and LBBB produce the highest accuracy. Meanwhile, most misclassifications were found in APB which was predicted to be AFL.

Table 7 shows comparison of classification results between proposals methods and previous research. The comparison results show that research [21] resulted in an average classification accuracy of 99.84% for 17 classes but for different datasets. In research [21], 2D data was used while other studies used 1D. In the same dataset, research [22] yielded a higher accuracy of 99.01% for distinguishing between the four categories of heart rate. While in the proposed method, the system performance is used for a larger number of classes, 7. When compared with research [23], the accuracy of the proposed method with more classification classes results in higher accuracy. In fact, direct comparison of studies cannot be carried out because the use of the CNN technique in previous studies was applied to one-dimensional ECG signals. While in this study, ECG classification was carried out on ECG images where this technique has not been widely explored in previous studies.

IV. CONCLUSION

This study presents multi-class heart abnormalities

Table 3. Image augmentation parameters

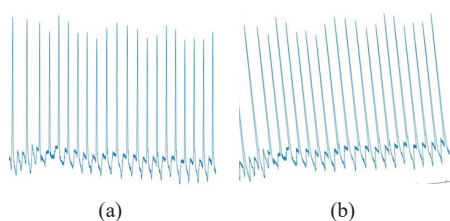| Parameter	Description
Rotation range	Rotation image 18
Zoom	zoom range 0.15
Image Size	150 x 150
Color mode	Color mode = RGB
Class mode	Class mode = Categorical
Seed & Random State	Seed = 42 Random State = 42
Width shift range	Wide shift range = 0.15
Height shift range	High shift range = 0.15
Shear range	Shift range = 0.15

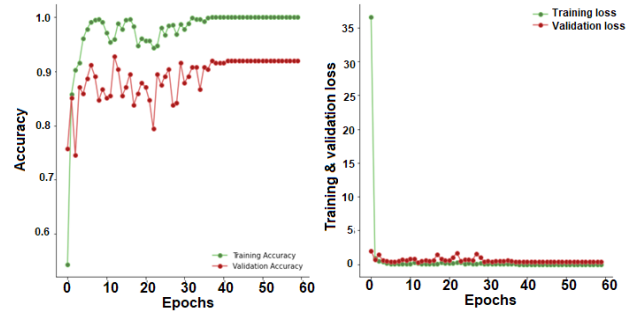Figure 5. (a) Original ECG graph, (b) Augmented ECG graph.

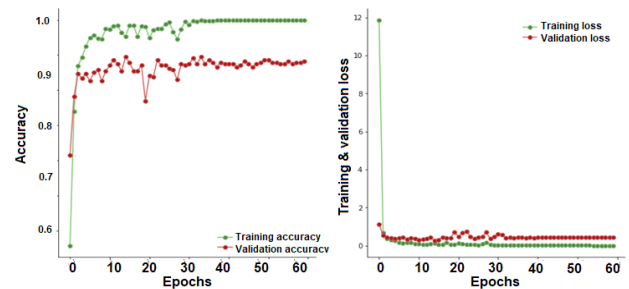Figure 6. Training dan validation on DenseNet
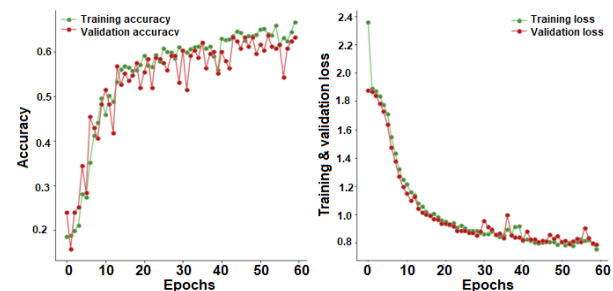

Figure 7. Training dan validation on VGG-16

Figure. 8. Training dan validation on AlexNet

Table 4. Precision, Recall, F1-score dan Accuracy DenseNet Architecture

Epochs	Precision	Recall	F1-score	Accuracy
10	0.89	0.88	0.88	0.89
20	0.91	0.90	0.90	0.91
30	0.90	0.89	0.90	0.91
40	0.91	0.91	0.91	0.91
50	0.90	0.90	0.90	0.91
60	0.92	0.91	0.92	0.92

Table 5. Precision, Recall, F1-score dan Accuracy VGG Architecture

Epochs	Precision	Recall	F1-score	Accuracy
10	0.86	0.87	0.86	0.85
20	0.87	0.85	0.85	0.86
30	0.87	0.86	0.87	0.87
40	0.89	0.90	0.89	0.89
50	0.89	0.89	0.89	0.89
60	0.90	0.88	0.89	0.88

Table 7. Comparison with previous studies

Reference	Methods	Dataset	Classes	Acc.
Liang, Y., el. al. (2020) [23]	CNN-BiLSTM	https://data.mendeley.com/datasets/7dybx7wyfn/3	17	85.00%
Sonain Jamil & MuhibUr Rahman (2020) [22]	2D CWT-ArrhythmiaNet-CoF	MIT-BIH https://www.physionet.org/cgi-bin/atm/ATM	17	99.84%
Shasha Ji, et. al. (2021) [25]	Stacking-DWKNN	MIT-BIH https://www.physionet.org/cgi-bin/atm/ATM	4	99.01%
Sandra˙Smigiel, et. al. (2021)[24]	CNN-Entropy	PTB-XL https://physionet.org/content/ptb-xl/1.0.1/	2	88.90%
			5	76.50%
			20	69.80%
This work	**DenseNet VGG AlexNet**	**https://data.mendeley.com/datasets/7dybx7wyfn/3**	**7**	**92%**
				89%
				63%

Table 6. Precision, Recall, F1-score dan Accuracy AlexNet Architecture

Epochs	Precision	Recall	F1-score	Accuracy
10	0.20	0.39	0.26	0.47
20	0.21	0.36	0.25	0.42
30	0.29	0.39	0.31	0.47
40	0.45	0.43	0.35	0.52
50	0.44	0.42	0.34	0.51
60	0.56	0.62	0.58	0.63

detection based on ECG images using transfer learning method. This proposed system is unique com-pared to previous studies which are commonly signal processing based. Several CNN methods are implemented to classify 7 class ECG signal consisting of NSR, AFIB, AFL, APB, Bigeminy, LBBB, and PVC. System performance is measured based on the parameters of precision, F1-socre, recall, accuracy, and confusion matrix. The test results on 1233 datasets consisting of 80% training data and 20% testing data show that the best performance for all test parameters, DenseNet model is the best architecture followed by VGG and AlexNet. The maximum accuracy obtained at 60 epoch is 0.92. Needs to be developed for further research is the ability of a comprehensive abnormality classification system that connects clinical features and technical use so that it can identify abnormalities more quickly and provide appropriate treatment.

REFERENCES

[1] E. M. Dewi, S. Aulia, and S. Hadiyoso, "Multi-abnormal ECG signal classification using dispersion entropy and statistic feature," TELKOMIKA Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, dan Teknik Elektronika, vol. 10, no. 3, pp. 677, 2022.

[2] S. Hadiyoso, S. Aulia, and A. Rizal, "One-lead electrocardiogram for biometric authentication using time series analysis and Support Vector Machine," International Journal of Advanced Computer Science and Application, vol. 10, no. 2, 2019.

[3] S. Hadiyoso, I. Wijayanto, A. Rizal, and S. Aulia, "Biometric systems based on ECG using ensemble empirical mode decomposition and variational mode decom-position," Journal of Applied Engineering Science, vol. 18, no. 2, pp. 181–191, 2020.

[4] S. Aulia and S. Hadiyoso, "Abnormal ECG classification using empirical mode decomposition and entropy," Jurnal Rekayasa Elektrika, vol. 17, no. 3, pp. 191–198, 2021.

[5] A. Wroblewska et al., "A Deep learning approach for automatic detection of qualitative features of lecturing," Jordanian Journal of Computers and Information Technology, vol. 06, no. 03, pp. 1–9, 2022.

[6] N. Kalaivani, N. Manimaran, S. Sophia, and D. D. Devi, "Deep learning based lung cancer detection and classification," IOP Conference Series: Material Science Engineering, vol. 994, no. 1, pp. 1-5, 2020.

[7] M. Ahsan, R. Gomes, and A. Denton, "Application of a convolutional neural network using transfer learning for tuberculosis detection," in Proc. IEEE International Conference on Electro Information Technology, 2019, pp. 427–433.

[8] M. Tan and Q. V. Le, "EfficientNet: rethinking model scaling for convolutional neural networks," in Proc. 36th International Conference on Machine Learning, 2019, pp. 10691–10700.

[9] S. Hadiyoso and S. Aulia, "Classification of koilonychia, beaus lines, and leu-konychia based on nail image using transfer learning VGG-16," Jurnal Rekayasa Elektrika, vol. 18, no. 2, pp. 109–114, 2022.

[10] M. Masud, N. Sikder, A. Al Nahid, A. K. Bairagi, and M. A. Alzain, "A machine learning approach to diagnosing lung and colon cancer using a deep learning based classification framework," Sensors, vol. 21, no. 3, pp. 1–21, 2021.

[11] M. Liebenlito, Y. Irene, and A. Hamid, "Classification of tuberculosis and pneumonia in human lung based on chest x-ray image using convolutional neural network," Inprime: Indonesian Journal of Pure and Applied Mathematics, vol. 2, no. 1, pp. 24–32, 2020.

[12] D. Bendarkar, P. Somase, P. Rebari, R. Paturkar, and A. Khan, "Web-based recognition and translation of american sign language with CNN and RNN," International Journal of online and Biomedical Engineering, vol. 17, no. 1, pp. 34–50, 2021,

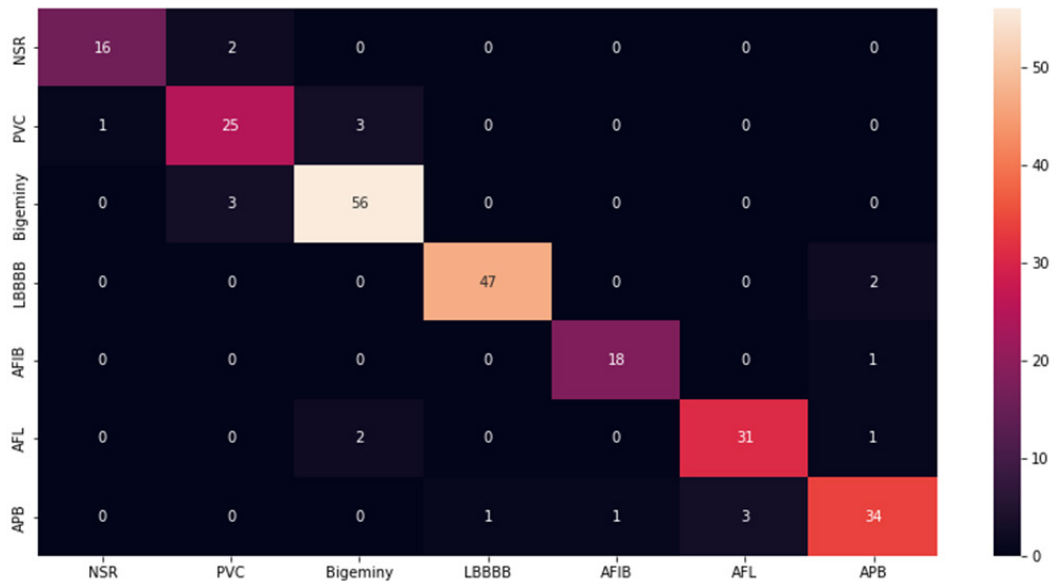[13] S. Aulia, S. Hadiyoso, T. L. E. R. Mengko, and A. B. Suksmono,

Figure 9. Heatmap of confusion matrix

"Covid-19 and tuberculosis classification based on chest x-ray using convolutional neural network," in Proc. International Conference on Electronics, Biomedical Engineering, and Health Informatics, 2021, pp. 407–420.

[14] K. Ramasubramanian and A. Singh, *Machine Learning Using R*, Berkeley, CA: Apress, 2019.

[15] O. Rochmawanti, F. Utaminingrum, and F. A. Bachtiar, "Analisis performa pre-trained model convolutional neural network dalam mendeteksi penyakit tuberkulosis," Jurnal Teknologi Informasi dan Ilmu Komputer, vol. 8, no. 4, pp. 805, 2021.

[16] I. Wahlang, P. Sharma, S. Sanyal, G. Saha, and A. K. Maji, "Deep learning techniques for classification of brain MRI," International Journal of Intelligent System Technologies and Applicayions, vol. 19, no. 6, pp. 571–588, 2020,

[17] Z. Yuan et al., "Automatic polyp detection in colonoscopy videos," IEEE Transaction on Medical Imaging, vol. 10133, pp. 101332K, 2017.

[18] Y. T. Li and J. I. Guo, "A VGG-16 based faster RCNN model for PCB error in-spection in industrial AOI applications," in Proc. IEEE International Conference on Consumer Electronics-Taiwan, 2018, pp. 1–2.

[19] B. R. Nanditha, A. G. Kiran, H. S. Chandrashekar, M. S. Dinesh,

and S. Murali, "An ensemble deep neural network approach for oral cancer screening," Int. J. online Biomed. Eng., vol. 17, no. 2, pp. 121–134, 2021.

[20] A. Elnakib, H. M. Amer, and F. E. Z. Abou-Chadi, "Early lung cancer detection using deep learning optimization," International Journal of Online and Biomedical Engineering, vol. 16, no. 6, pp. 82–94, 2020.

[21] S. Jamil and M. Rahman, "A novel deep-learning-based framework for the classification of cardiac arrhythmia," Journal of Imaging, vol. 8, no. 3, 2022.

[22] Y. Liang, S. Yin, Q. Tang, Z. Zheng, M. Elgendi, and Z. Chen, "Deep learning algorithm classifies heartbeat events based on electrocardiogram signals," Frontiers in Physiology, vol. 11, pp. 1-10, 2020.

[23] S. Śmigiel, K. Pałczyński, and D. Ledziński, "ECG signal classification using deep learning techniques based on the PTB-XL dataset," Entropy, vol. 23, no. 9, pp. 1–20, 2021.

[24] S. Ji, R. Li, S. Shen, B. Li, B. Zhou, and Z. Wang, "Heartbeat classification based on multifeature combination and stacking-DWKNN algorithm," Journal of Healthcare Engineering, vol. 2021, pp. 1-14, 2021.

[25] Physionet Database. Available: http://www.physionet.org