

# A Transfer-Learning Based Ensemble Architecture for ECG Signal Classification

Tareque Bashar Ovi\*, Sauda Suara Naba<sup>†</sup>, Dibaloke Chanda<sup>‡</sup> and Md. Saif Hassan Onim<sup>§</sup>

Department of Electrical, Electronic and Communication Engineering,

Military Institute of Science and Technology (MIST) Dhaka-1216, Bangladesh

\*ovitareque@gmail.com, <sup>†</sup>suaranaba@gmail.com, <sup>‡</sup>dibaloke66@gmail.com <sup>§</sup> onim.hassan16@gmail.com

**Abstract**—Manual interpretation and classification of ECG signals lack both accuracy and reliability. These continuous time-series signals are more effective when represented as an image for CNN-based classification. A continuous Wavelet transform filter is used here to get corresponding images. In achieving the best result generic CNN architectures lack sufficient accuracy and also have a higher run-time. To address this issue, we propose an ensemble method of transfer learning-based models to classify ECG signals. In our work, two modified VGG-16 models and one InceptionResNetV2 model with added feature extracting layers and ImageNet weights are working as the backbone. After ensemble, we report an increase of 6.36% accuracy than previous MLP based algorithms. After 5-fold cross-validation with the Physionet dataset, our model reaches an accuracy of 99.98%.

**Index Terms**—ensemble learning, deep learning, medical image processing, transfer learning, continuous wavelet transform

## I. INTRODUCTION

ECG documents the electrical function of the heart by encapsulating essential electrical activities. It is crucial for the preliminary detection of numerous cardiac abnormalities and cardiovascular diseases. Substantial facts like a patient's heart rate, heart rhythm, record of heart attack, probable narrowing of the coronary arteries, traces of reduced oxygen distribution to the heart, and so forth can be attained from an ECG report. ECGs are usually conducted in three basic ways: *resting ECG*, *ambulatory ECG*, and *cardiac stress test*. An ECG graph (Fig 1) can be characterized by a repeated series of *P*, *QRS*, *T*, and a conditional *U* wave. The peaks (*P*, *Q*, *R*, *S*, *T* and *U*), intervals (*PR*, *RR*, *QRS*, *ST*, and *QT*), and segments (*PR* and *ST*) represent the traits of an ECG. Deviation from the expected form of electrical signals signifies several issues including defects or abnormalities in the *heart's shape, size, electrolytes, ischemia, heart rhythm etc.* Once the anomaly is detected, risk factors for cardiovascular diseases can be diagnosed and treated, consequently upgrading the patient's medication.

However, unlike detecting, classifying the abnormalities in ECG precisely is an immensely complicated task considering the numerous challenges that arise throughout the process. One of the challenges can be examining and identifying cardiac problems from ECG paper. Again, applying ordinary visual analysis approaches can be challenging for physicians as well. Resultantly, intelligent computer systems that utilize sophisticated image and signal processing techniques have been

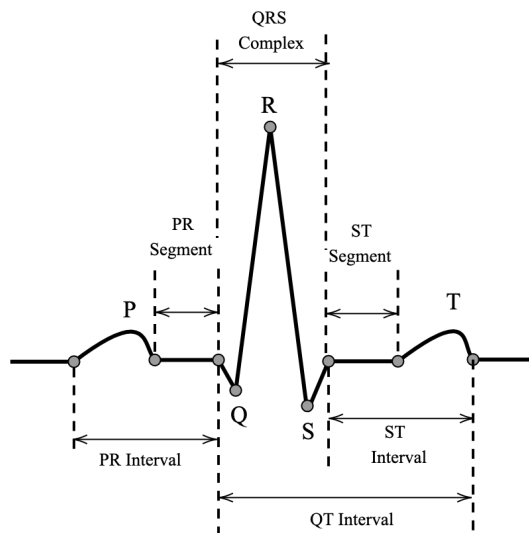


Fig. 1: Typical ECG Signal

introduced. This aids in more precise diagnosis of patients and offers better treatment than empirical procedures. Even in the case of a fully automated and computerized system, it's not an easy process, considering the presence of external noise and uneven classes in the dataset. The accuracy achieved applying machine learning is insufficient as well. Multiple attempts have been made to incorporate deep learning algorithms to address and solve the issue of precisely classifying cardiovascular diseases with a lower number of false positive detections.

Zihlmann *et al.* [1] presented two neural network architectures, a convolutional neural network (CNN) and a convolutional recurrent neural network (CRNN), for ECG classification. On the hidden challenge testing set, the second architecture exceeded the first in terms of accuracy, with an F1 score of 82.1 percent for the PhysioNet/CinC Challenge 2017 atrial fibrillation (AF) classification data set. Rajkumar *et al.* [2] employed CNN to attain the features automatically from the time domain ECG signals collected from the MIT-BIH database [3]. They achieved 93.6 percent accuracy using the ELU activation function. Syama *et al.* [4] implemented a multilayer-perceptron neural network (MLPNN) for the classification of ECG signals collected from the MIT-BIH

database and achieved 94 percent accuracy. ESSA *et al.* [5] came up with a deep learning-based multi-model ensemble by implementing a combination of a convolutional neural network (CNN) and long short-term memory (LSTM) network in order to classify ECG. The proposed approach was evaluated on the MIT-BIH database to derive the experimental findings with an overall accuracy of 95.81 percent. Boussaa *et al.* [6] employed Mel Frequency Coefficient Cepstrum algorithm (MFCC) and artificial neural network (ANN) classifier for classifying ECG signal extracted from the MIT-BIH database. Hannun *et al.* [7] collected 91,232 single-lead ECGs from 53,549 individuals who used a single-lead ambulatory ECG monitoring device to create a deep neural network (DNN) to distinguish 12 rhythm categories. ŞEN *et al.* [8] proposed ECG time-series signal classification and ECG spectrogram pictures to classify heart-beat arrhythmia. Ullah *et al.* [9] proposed a two-dimensional (2-D) convolutional neural network (CNN) model in order to classify ECG signals into eight classes and achieved an accuracy of 99.11 percent. Cordoş *et al.* [10] proposed a strategy to use ECG signal for biometric identification employing Inception-v3, Xception, MobileNet and NasNetLarge models. They found Inception-v3 model to be the most effective model for ECG classification having an accuracy of 99.5 percent. Gajendran *et al.* [11] used different modern deep networks trained on the ImageNet database to categorize scalograms (2D representations) of ECG signals collected from three PhysioNet databases. Rahuja *et al.* [12] introduced a classifier for automatic ECG classification employing continuous wavelet transform and AlexNet CNN. They used PhysioNet research dataset for their experiment and attained an overall accuracy of 97.3 percent with a limited dataset. Their works demonstrate that modern CNN architectures performs well for multi-class ECG signal classification [13]–[16].

Our major contribution is introducing modified VGG-16 and InceptionResNetV2 models pretrained with ImageNet weights and finally constructing a soft voting based ensemble architectures for taking maximum accuracy for respective classes.

## II. METHODOLOGY

### A. Dataset Description

We used a pre-existing open-source dataset for this work which consists of three distinct types of ECG signals from three different types of heart condition *Arrhythmia (ARR)*, *Normal Sinus Rhythm (NSR)*, and *Congestive Heart Failure (CHF)*. The original version of the dataset is available at PhysioNet [17] in three separate sections which are MIT-BIH Arrhythmia Database [18], MIT-BIH Normal Sinus Rhythm Database [3] and BIDMC Congestive Heart Failure Database [19]. The modified version of the dataset combines these three separate databases in a common format as shown in fig 2.

A common scaling factor was used to normalize the range of the magnitude followed by a sampling stage with a common sampling rate of  $128Hz$  and number of samples equal to 65536. This allowed all signals to have the same length. Number of instances for ARR, NSR, and CHF class are

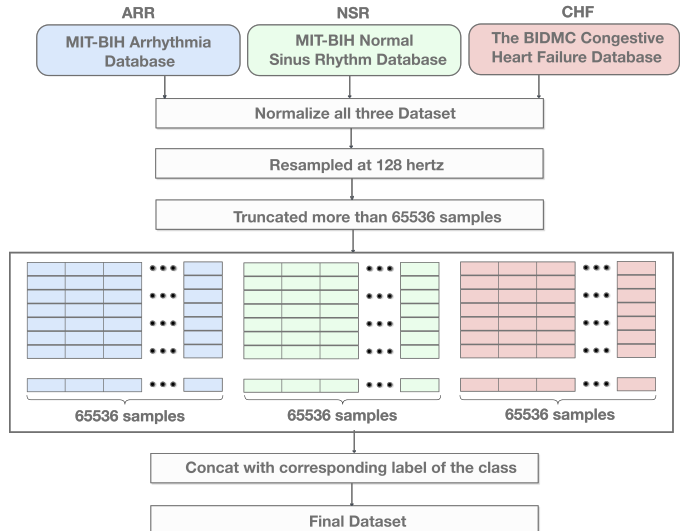


Fig. 2: Modified Dataset Combining MIT-BIH Arrhythmia Database, MIT-BIH Normal Sinus Rhythm Database, BIDMC Congestive Heart Failure Database

respectively 96, 30, and 36 which equates to a total of 162 ECG recordings.

### B. Dataset Preprocessing

Time series data is not compatible with CNN architecture. To make it compatible, CWT(Continuous Wavelet Transform) is utilized to convert the one-dimensional signal to a two-dimensional scalogram.

$$C(a, b; f(t), \psi(t)) = \int_{-\infty}^{\infty} f(t) \frac{1}{a} \psi^* \left( \frac{t-b}{a} \right) dt \quad (1)$$

CWT is generally expressed by the equation shown in (1) defined by a scale parameter  $a > 0$  and position parameter  $b$ . This is used as an analyzing function to measure the similarity between the wavelet and the signal by equating the inner product between them. The CWT is defined as the sum of the signal multiplied by scaled and shifted versions of the wavelet function  $\psi(t)$  also known as the mother wavelet. A small value of  $a$  signifies a compressed wavelet which captures rapid changing details whereas a large value of  $a$  captures slowly changing details. Generated scalograms after CWT applied to 1D signals can be interpreted as a time-frequency representation of the time-series signal. Fig 4 shows sample scalogram images after applying CWT to three ECG classes. The generated scalograms are of dimension of  $224 \times 224 \times 3$  each.

### C. Proposed Methods

Two different approaches are explored to correctly categorize the three classes. Firstly, a simple CNN architecture is used as a reference to classify the scalograms. In the second approach, we made use of both transfer learning and ensemble methods combining three different pre-trained deep

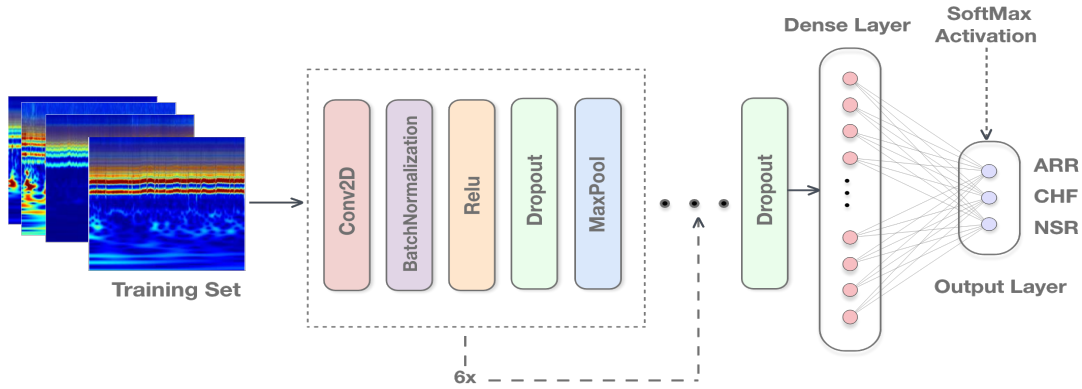


Fig. 3: Custom Convolutional Neural Network Architecture

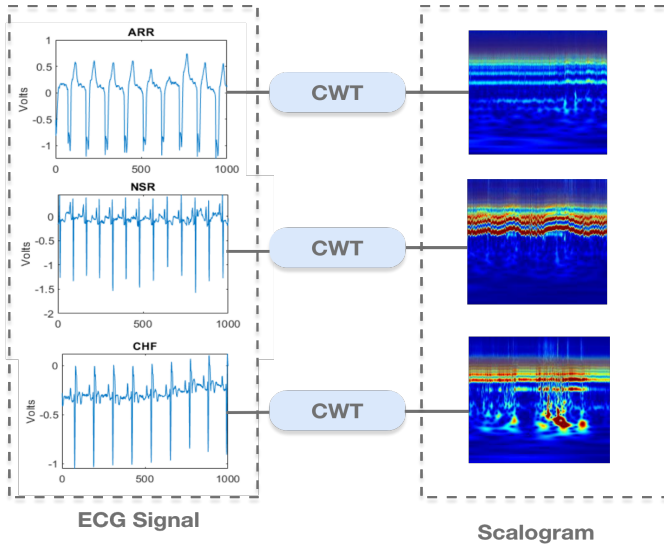


Fig. 4: Conversion from ECG Signal to Scalogram Using Continuous Wavelet Transform

CNN architectures. The following sections will give a detailed overview of these two approaches.

For the first approach, a CNN is constructed by stacking several repeated blocks as shown in fig 3. A single block consists of a *convolutional layer* followed by a *Batch Normalization* layer. The purpose of the *convolutional layer* is to extract relevant features from the scalogram. *Batch normalization* layers significantly speed up the training process as it standardizes the input data for each batch. The next layer is a *dropout layer* whose purpose is to prevent overfitting by randomly turning off some learning parameter. The final layer in the block is a *pooling layer* which reduces the spatial size of the previous layer output. These repeated blocks are finally terminated with an additional *dropout layer*. After that, flattening is done to convert to *dense layer* which is required for classification. The output layer is activated by *Softmax Activation Function*. Hyper-parameters used for this model is summarized in Table I. They were fine tuned after several test and trial to keep high accuracy and low loss.

TABLE I: Training Parameters for Convolutional Neural Network

Parameter	Value
Base Architecture	CNN
Classes	3
Input Dimension	$224 \times 224 \times 3$
Number of Epochs Trained	200
Model Type	Multiclass Classification
Hidden Layer Activation	Rectified Linear Unit (ReLU)
Output Layer Activation	Softmax
Optimizer	Adam
Initial Learning Rate	0.05
Dropout Rate	0.5
Padding Type	Same
Shuffle Frequency	Each Epoch

TABLE II: Augmentation parameters for augmentation of training images

Augmentation Parameter	Value
Random Flip	- <i>Horizontal</i>
Random Rotation	- 0.2

The final layer is composed of three perceptions and activation function for this layer is softmax, which is a common practice for multi-class classification. Softmax generates a probability value between 0 to 1 based on the confidence of prediction. The model was trained for 4 min 36 sec with a batch size of 128.

In the second approach, prior to building the model, augmentation was performed to increase the quantity of training data and to incorporate variance into the dataset. The dataset's size expanded from 162 to 486 as a result of the augmentation.

Two different state-of-the-art pre-trained CNN architectures are incorporated for the ensemble model which are, respectively, VGG-16 and InceptionResNetV2. A brief description about these two models are given in the following section.

1) *Modified VGG-16*: VGG-16 is a CNN Architecture that won the 2014 ImageNet Large Scale Visual Recognition Chal-

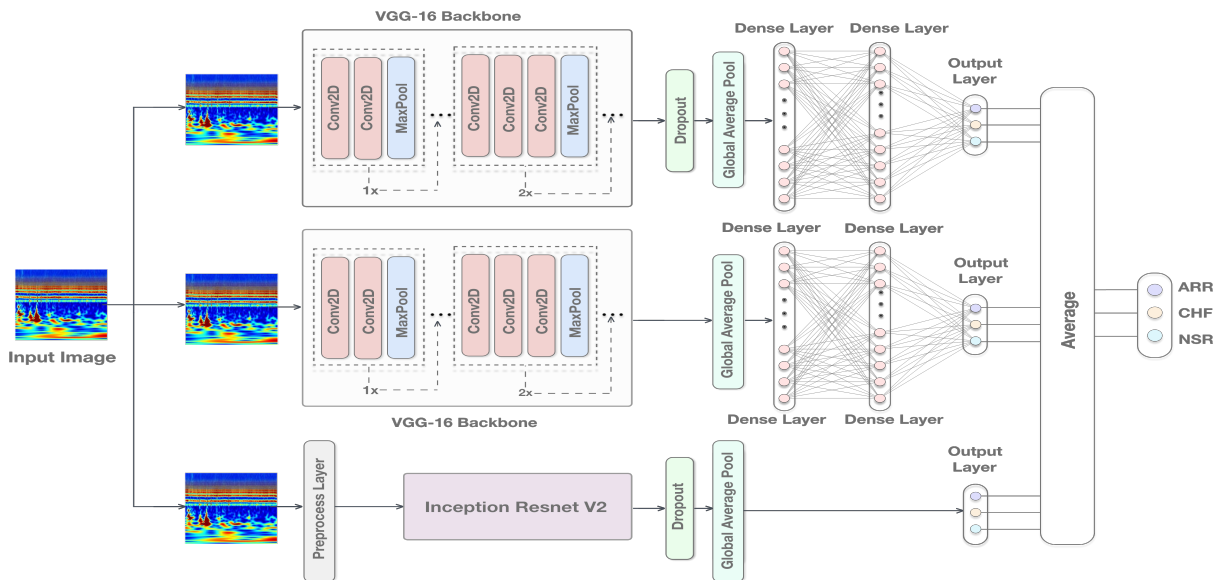


Fig. 5: Architecture of the ensemble model

lenge (ILSVRC). The VGG-16 architecture is an enhancement to the AlexNet design. Instead of directly employing the VGG-16, we made some slight modification, but kept the feature extraction part of the model which consists of consecutive convolutional and pooling layers. Instead of directly transitioning to fully connected layers, two intermediate layers, dropout and global average pooling are used. Fully connected layers are vulnerable to overfitting, limiting the entire network’s generalization capabilities. To avoid this, dropout is added as an intermediate layer for regularization. The reason behind using the global average pooling layer is it enforces a strong correspondence between the feature maps and categories. Finally, three fully connected dense layers are used, where the final layer is comprised of three perceptions for three categories.

2) *Modified InceptionResnetV2*: InceptionResNetV2 is a convolutional neural network that extends the Inception family but includes residual connections. It was trained on millions of images from the *ImageNet* Database. The architecture accepts data in a certain format. Because of this, a preprocessing layer is used to convert to that particular format. Dropout and global average pooling are used for the same reason mentioned above. Finally a single dense layer with softmax activation is used for classification. The complete ensemble architecture is shown in fig 5.

Ensemble method combines multiple architectures or algorithms to make predictions by combining the outputs. For classification type problem there are two major ways to combine the outputs, one is voting and another is averaging. In our work, we implemented the latter approach. The averaging mechanism works by combining each model’s prediction equally. The hyperparameters for the ensemble model are summarized in Table III.

TABLE III: Training Parameters for Ensemble Architecture

Parameter	Value
Number of Epochs Trained	80
Batch Size	1
Output Layer Activation	Softmax
Optimizer	Adam
Initial Learning Rate	0.05
Loss Function	Categorical Crossentropy
Total Number of Parameters	91,391,817
Trainable Parameters	37,020,649
Non-Trainable Parameters	54,371,168

### III. RESULT ANALYSIS

With the custom CNN approach, the accuracy value is not satisfactory enough. The reason behind such performance is mainly the dataset size. Such a small dataset is not suitable for machine learning models because they tend to overfit, meaning they perform really well on training data, but poorly on test data. Even though the training accuracy is nearly 100%, the validation accuracy is 59.38%. We tried several empirical methods to regularize the model, but the accuracy value did not improve.

Because of this, in the second approach we utilized data augmentation. Additionally, after analyzing prior studies, we came to the conclusion that transfer learning works well for scalograms and can achieve quite good accuracy which motivated us to incorporate transfer learning in ensemble methods.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

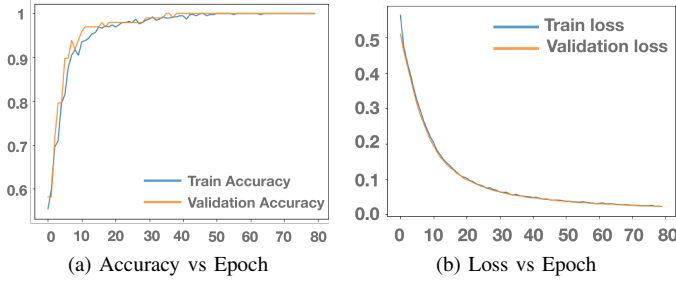


Fig. 6: Training and validation Evaluation Curves for Ensemble Architecture

Fig 6 represents how the accuracy and loss value change with increase in epoch for the ensemble approach. From the curves, it is obvious that the validation performance of the model is almost as good as the training performance which means the model is able to generalize. To make sure the model did not overfit, we performed k-fold cross validation for 20 epoch with  $k = 5$  which resulted in around 99.98% accuracy.



Fig. 7: Illustration of 5-fold cross validation

Along with accuracy, other evaluation metrics like loss, accuracy, precision, and recall are calculated.

TABLE IV: Evaluation Metrics of the Validation data for the Ensemble model

Evaluation Metric		Value
Loss	-	0.226
Accuracy	-	1.0
Precision	-	1.0
Recall	-	1.0

*Precision* and *recall* are defined by the ratio shown in equation (2) and (3) respectively. Where  $TP = True\ Positive$ ,  $FP = False\ Positive$  and  $FN = False\ Negative$ . All the evaluation metrics are summarized in Table IV.

Table V compares the performance of some state-of-the-art research works with our work.

#### IV. CONCLUSION

Our model achieves 99.98% accuracy with 5-fold cross validation to classify three classes ARR, NSR and CHF experimented on Physionet dataset. Time series ECG data was transformed into image data with wavelet transformation for compatibility with CNN models. CNN based architectures generated comprehensive results whereas transfer learning-based model performed better. But single transfer learning-based model was not adequate enough to command perfect accuracy and reliability. Hence, our proposed transfer learning based ensemble model with soft voting outperforms all previous literature with a cost of minor computational run-time.

#### REFERENCES

- [1] M. Zihlmann, D. Perekrestenko, and M. Tschannen, "Convolutional recurrent neural networks for electrocardiogram classification," in *2017 Computing in Cardiology (CinC)*. IEEE, 2017, pp. 1–4.
- [2] A. Rajkumar, M. Ganesan, and R. Lavanya, "Arrhythmia classification on ecg using deep learning," in *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*. IEEE, 2019, pp. 365–369.
- [3] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. Ivanov, R. Mark, J. Mietus, G. GB, C. Peng, and H. Stanley, "The mit-bih normal sinus rhythm database," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.
- [4] S. Syama, G. S. Sweta, P. Kavyasree, and K. J. M. Reddy, "Classification of ecg signal using machine learning techniques," in *2019 2nd International Conference on Power and Embedded Drive Control (ICPEDC)*. IEEE, 2019, pp. 122–128.
- [5] E. Essa and X. Xie, "An ensemble of deep learning-based multi-model for ecg heartbeats arrhythmia classification," *IEEE Access*, vol. 9, pp. 103 452–103 464, 2021.
- [6] M. Boussaa, I. Atouf, M. Atibi, and A. Bennis, "Ecg signals classification using mfcc coefficients and ann classifier," in *2016 International Conference on Electrical and Information Technologies (ICEIT)*. IEEE, 2016, pp. 480–484.
- [7] A. Y. Hannun, P. Rajpurkar, M. Haghpanahi, G. H. Tison, C. Bourn, M. P. Turakhia, and A. Y. Ng, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature medicine*, vol. 25, no. 1, pp. 65–69, 2019.
- [8] S. Y. ŞEN and N. ÖZKURT, "Ecg arrhythmia classification by using convolutional neural network and spectrogram," in *2019 Innovations in Intelligent Systems and Applications Conference (ASYU)*. IEEE, 2019, pp. 1–6.
- [9] A. Ullah, S. M. Anwar, M. Bilal, and R. M. Mehmood, "Classification of arrhythmia by using deep learning with 2-d ecg spectral image representation," *Remote Sensing*, vol. 12, no. 10, p. 1685, 2020.
- [10] C. Cordoş, L. Mihailă, P. Faragó, and S. Hintea, "Ecg signal classification using convolutional neural networks for biometric identification," in *2021 44th International Conference on Telecommunications and Signal Processing (TSP)*, 2021, pp. 167–170.
- [11] M. K. Gajendran, M. Z. Khan, and M. A. K. Khattak, "Ecg classification using deep transfer learning," in *2021 4th International Conference on Information and Computer Technologies (ICICT)*. IEEE, 2021, pp. 1–5.
- [12] N. Rahuja and S. K. Valluru, "A deep neural network approach to automatic multi-class classification of electrocardiogram signals," in *2021 International Conference on Intelligent Technologies (CONIT)*. IEEE, 2021, pp. 1–4.
- [13] A. Demonbreun and G. M. Mirsky, "Automated classification of electrocardiograms using wavelet analysis and deep learning," in *2020 Computing in Cardiology*. IEEE, 2020, pp. 1–4.
- [14] W. Zhang, L. Yu, L. Ye, W. Zhuang, and F. Ma, "Ecg signal classification with deep learning for heart disease identification," in *2018 International Conference on Big Data and Artificial Intelligence (BDAI)*. IEEE, 2018, pp. 47–51.

TABLE V: Performance Comparison with Existing Work

Year	Reference	Dataset	Method	Performance (Accuracy)
2018	Zhang <i>et al.</i> [14]	MIT-BIH	12-layer 1-d CNN	97.70%
2019	Syama <i>et al.</i> [4]	MIT-BIH	MLPNN	94.00%
2020	Huang <i>et al.</i> [16]	MIT-BIH	FCResNet using MOWPT	98.79%
2020	Ahamed <i>et al.</i> [15]	MIT-BIH	KNN	97.64%
			Decision Tree	96.11%
			ANN	98.06%
			Support Vector Machine	97.58%
			LSTM	95.53%
			Ensemble Approach	97.78%
2021	Cordoş <i>et al.</i> [10]	Fantasia	Transfer Learning with Inception-v3	99.5%
<b>2022</b>	<b>Ours</b>	<b>Physionet</b>	<b>Transfer Learning based Ensemble with Data Augmentation</b>	<b>99.98%</b>

- [15] M. A. Ahamed, K. A. Hasan, K. F. Monwar, N. Mashnoor, and M. A. Hossain, "Ecg heartbeat classification using ensemble of efficient machine learning approaches on imbalanced datasets," in *2020 2nd International Conference on Advanced Information and Communication Technology (ICAICT)*. IEEE, 2020, pp. 140–145.
- [16] J.-S. Huang, B.-Q. Chen, N.-Y. Zeng, X.-C. Cao, and Y. Li, "Accurate classification of ecg arrhythmia using mowpt enhanced fast compression deep learning networks," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–18, 2020.
- [17] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals," *circulation*, vol. 101, no. 23, pp. e215–e220, 2000.
- [18] G. B. Moody and R. G. Mark, "The impact of the mit-bih arrhythmia database," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45–50, 2001.
- [19] D. S. Baim, W. S. Colucci, E. S. Monrad, H. S. Smith, R. F. Wright, A. Lanoue, D. F. Gauthier, B. J. Ransil, W. Grossman, and E. Braunwald, "Survival of patients with severe congestive heart failure treated with oral milrinone," *Journal of the American College of Cardiology*, vol. 7, no. 3, pp. 661–670, 1986.