Journal of Informatics and Web Engineering

Vol. 2 No. 2 (September 2023) eISSN: 2821-370X

A Cost-Based Dual ConvNet-Attention Transfer Learning Model for ECG Heartbeat Classification

Johnson Olanrewaju Victor[1], XinYing Chew[1*], Khai Wah Khaw[2], Ming Ha Lee[3]

[1]School of Computer Sciences, Universiti Sains Malaysia, 11800 Pulau Pinang, Malaysia

[2]School of Management, Universiti Sains Malaysia, 11800 Pulau Pinang, Malaysia

[3]Faculty of Engineering, Computing and Science, Swinburne University of Technology Sarawak Campus, 93350 Kuching, Sarawak, Malaysia

**corresponding author: (xinying@usm.my, ORCiD: 0000-0001-5539-1959)*

Abstract – The heart is a very crucial organ of the body. Concerted efforts are constantly put forward to provide adequate monitoring of the heart. A heart disorder is reported to cause a lot of hidden ailments resulting in numerous deaths. Early heart monitoring using an electrocardiogram (ECG) through the advancement of computer-aided diagnostic (CAD) systems is widely used. Meanwhile, the use of human reading of ECG results are faced with many challenges of inaccurate and unreliable interpretations. Over two decades, studies provided artificial intelligence (AI) technique using machine learning (ML) algorithms as a fast and reliable technique for ECG heartbeat classification. Moreover, in recent times, deep learning (DL) techniques have been focused on providing automatic feature extraction and better classification performance. On the other hand, the challenge with the ECG data is its imbalance nature. Therefore, this paper proposes a cost-based dual convolutional attention transfer DL model for ECG classification. The proposed model uses PhysionNet-MIT-BIH and Physikalisch-Technische Bundesanstalt (PTB) Diagnostics datasets. The first part uses the MIT-BIH for ECG categorization, while representations learned from the first classifier are used for PTB analysis through transfer learning (TL). The proposed model is evaluated and compared with well-performing conventional ML models based on their F1-score and accuracy scores. Our experimental finding show that the proposed model outperformed the well-performing ML models as well as competitive with past studies for both the classification and TL part, having obtained 98.45% for both F1-score and accuracy. The proposed model is applicable to real-life trials and experiments for ECG heartbeat and other similar domains.

Keywords—Transfer learning, Cost-sensitive, Dual CNN, Attention, Heartbeat

Received: 30 June 2023; Accepted: 01 August 2023; Published: 16 September 2023

I. INTRODUCTION

Heart arrhythmia (Ha) is a major condition of the heart rate in which there is an irregular or one of the following appearances: too slow, too early, or too rapid. It is a series of abnormalities in the heartbeats depicted as electrical

Journal of Informatics and Web Engineering
https://doi.org/10.33093/jiwe.2023.2.2.7

Published by MMU Press. URL: https://journals.mmupress.com/jiwe

impulses. Individuals with shortness of breath, hypertension, palpitations, coronary artery disease, dizziness, and fainting are likely at risk of Ha. In as much as many diagnosed Ha reported are not harmful, there exist very dangerous, progressive, and often-diagnosed-too-late types leading to most cardiovascular disorders (CaDi), including stroke. Large numbers of deaths are recorded yearly as a result of CaDi. In 2019, reports showed 17.9 million death occurrences alone for CaDi, amounting to 32% of death globally [1–3]. Therefore, it is advantageous to constantly pursue early detection mechanisms of CaDi to provide timely control and prevention of its cases leading to perennial death.

Many medical diagnoses are based on accuracy and precision, where measurement and experiments are error-free and reliable. Hence, over the past two decades, ECG, a non-invasive, low-cost technology, has spurred much research interest in CaDi diagnosis. A typical 12-led ECG, is a process where receptive-signal objects (electrodes) are placed on different parts of the body to help detecting recurrent heartbeats. Figure 1 illustrates a signal reflect of the detected heartbeats. However, the complexity associated with accurate interpretation of the ECG necessitated the use of CAD to capture and annotate ECG data digitally. Reliability of CAD results for early diagnosis requires a long-term monitoring-more than 24 hrs [4, 5]. Not only does CAD speed up the collection process, but it also ensures that it is error-free and useful for further processes [1].

Advancements in AI and decision support systems have consistently provided engaging, innovative thoughts and ideas for analyzing medical data like the ECG with much precision and accuracy [6, 7]. Numerous intriguing models have been put forth in the literature for the automated categorization of ECG signals. These models comprise diverse ML and DL techniques or the hybrid of the methods to solve the categorization task, including DT, NB, KNN, SVM, and recent DL models [8, 9]. Researchers have recently created a variety of cardiac classification methods for detecting Ha using ECG data, largely depending on the growing learning capabilities of DL-based classifiers. The DL classifiers can convert the ECG signal to an image and then use two-dimensional convolutional neural networks (2D-CNN) [10] or one-dimensional CNN [11, 12] to learn the necessary features from the ECG signal. Long-term, short-term memory (LSTM) [13] is another DL technique frequently used for ECG signal analysis. Moreover, using 1D-CNN in several ECG classification domains has been extensively researched [11].

Meanwhile, the automatic classification of ECG heartbeat is a difficult undertaking often characterized by three core problems. The nature of the dataset comes first. Depending on the patient's physiological processes and emotional state, the ECG waveform's morphological properties always change from patient to patient. Hence, the intervals (RR, QRS) and ECG waveform segments (PR, ST) are affected by the physiological functions and autonomic nervous system activities. In light of this, the arrhythmia detection model may not work well for other patients if it was created using hand-crafted characteristics for a certain group of patients. Additionally, utilizing traditional ML models may not be reliable enough because they rely heavily on hand-crafted feature selection [14]. Consequently, preventing the classification model from having improved generalizability. In this case, DL models have the competitive advantage of providing automatic feature extraction coupled with the non-linear relationship learning ability inherent in them.

The second issue is about the class imbalance of the ECG data. It is widely known that ML/DL models perform well based on the data quality used as much as the learning algorithms applied. In this case, the ECG dataset's target class has a skewed distribution. It is, therefore, essential to provide a model that can learn with and through the imbalanced data in such a way that the minority class(es) is(are) significantly represented.

The third issue focuses on the Arrhythmia database [15], which consists of MIT-BIH and PTB diagnostic for ECG heartbeat classification. Reducing the cost and burden of training a reusable model from scratch necessitated developing a classifier on one dataset (in this case, MIT-BIH) and exploring such a classifier on the other (i.e., PTB) without retraining the classifier from scratch. The merits, no doubt include single-model-multiple tasks, reduction in cost and time for building a new model, and expanding re-usability of models in the same domain. Hence, a TL approach, resulting from an interesting pattern learned from a particular task and reusable in a similar study to boost model performance, is desirable [16, 17].

Therefore, this paper proposes a cost-based dual 1D-CNN attention (CB-DuConvNetA) transferable classifier that efficiently categorizes ECG heartbeat and adequately provides for automatic feature extraction. In addition, increases model performance and better generalizability of the minority classes following the AAMI standard for categorizing ECG MIT-BIH data. This paper's contribution may be summed up as follows:

i. Developing a dual 1D-CNN with an integrated attention mechanism in each convolutional block to boost the overall model performance via block-level performance and use the same as a pre-trained model.
ii. Developing a cost-based approach to solving the imbalance data problem associated with the ECG datasets. The method uses class-imbalance-ratio-weight, which leverages the proposed model loss function without additional overhead, to penalize the minority class, thereby achieving model generalizability.
iii. Using the pre-trained model as a transferable learning representation for PTB heartbeat classification.

The remainder of the paper is structured as follows. Section II offers a summary of related works. Section III discusses the methodology, including the data description, preprocessing, exploratory analysis, and the proposed CB-DuConvNetA model. Section IV provides the suggested method's classification results and performance comparisons. Section V presents the conclusion.

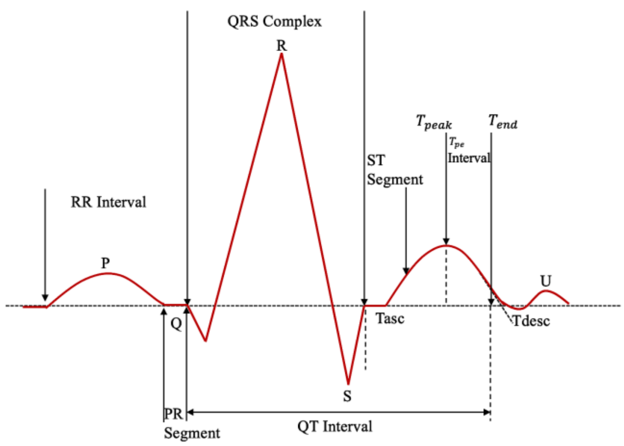


Figure 1. An example of ECG heartbeat impulse. Reproduced from [44].

II. RELATED WORKS

The fundamental steps to compute-diagnose Ha symptoms or its related abnormalities using ECG signal include: processing ECG signal, segmentation of heartbeat, feature extraction, and categorization. Identified in the literature, methods for ECG signal are but not limited to continuous wavelet transform (CWT), Empirical Mode Decomposition (EMD), Discrete wavelet Transform (DWT), Empirical Wavelet Transform (EWT), and Signal Energy Thresholding (SET) coefficients [2, 14, 18, 19]. In segmentation, an ECG signal based on an R-peak is utilized to detect the waves, segments, and intervals and compare them with well-known patterns using their temporal and morphological properties [12, 20, 21]. At the same time, Essa & Xie [13] discussed frequent feature extraction techniques to include wavelet, morphological, and n-order statistical features.

For the classification of ECG Ha, conventional ML models are widely used, including Gradient-Boosted Tree (GDBT) and random forest (RF) [14], KNN [9], [20], SVM, and multi-layer perceptron (MLP) MLP [22]. Meanwhile, recent studies focus on DL models for ECG classification [23, 24]. It may be due largely to the automatic feature, huge data handling, and the non-linear transformation capability inherent in most DL models [11]. The study of ECG classification is divergent and widely studied. One approach is converting the ECG data to an image and then using 2D-CNN or its variant as the classifier. Ahmad et al. [10] proposed variant multimodal fusion techniques for transforming ECG heartbeat classification. Three different images were first generated based on three other image-generating methods. The authors then performed image fusion on the three images to produce one image as input into the CNN. Similarly, Xie et al. [25] presented a feature enhancement framework to exploit the benefits of the CNN classifier. The authors introduce a time-frequency process to convert the ECG signal into images and feed it as input to CNN. Similarly, a time-frequency 2D-based approach was studied by Zhang et al. [11] to utilize the image-CNN technique by including a residual neural network (ResNet-101) TL for the PTB categorization. Other studies that implemented image-CNN using scalogram include [26, 27] and [28], where the classifier's learning on the PTB data is fine-tuned using weights collected during the initial training process.

In another vein, the 1D variant of CNN is widely accepted and exploited for ECG classification, focusing on TL rather than training models for individual ECG classification problems. Kachuee et al. [29] suggested 1D-CNN for heartbeat classification to diagnose five distinct arrhythmias in the AAMI standard. A data augmentation technique was used in the study to balance the datasets. Additionally, they proposed a strategy for applying the information learned from the previously trained model to the myocardial infarction (MI) classification problem by utilizing PhysionNet-MIT-BIH and PTB Diagnostics datasets. Findings show that the proposed CNN and its transferable representation can classify arrhythmias and MI with average prediction accuracies of 0.934 and 0.959, respectively. In another research, the authors categorize six distinct types of physiological signals using a deep neural network (DNN) of CNN and LSTM. They provided examples of applying the TL method and stressed effectively exploiting the suggested DNN and other areas [30]. Pham et al. [3] investigated 1D-CNN for ECG classification. The 1D-CNN is enhanced with three techniques: evolving normalization–activation, squeeze-and-excitation, and gradient clipping. To solve the imbalance data problem, a variant of 5-fold and 10-fold cross-validation (CV) was implemented. Despite a lot of hyperparameter tuning required for their proposed model, it surpasses other models, including the PhysioNet Challenge 2017 dataset, with an amazing F1-score of about 86.71%.

One of the critical challenges with the ECG classification is its data imbalance nature. An imbalanced data problem happens when the class distribution in a dataset is skewed. In this instance, there is the presence of the majority (with higher class numbers) and the minority (with fewer class numbers). In most cases, the majority is the negative class (-ve), and the minority is the positive class (+ve). The problem with this skewness is that it compels the classifier to be biased in favor of the majority class during training due to the high representation of the majority class [31, 32]. As a result, the model will not provide a true reflection of the underlying problem. Many real-life applications fall into this category, including medical [33], fraud detection [34], and Employee retention [35]. Either as a multi or binary class classification task, less frequent instances are significant and of paramount interest. It is cost-expensive and life-endangering to diagnose a patient's heart condition as normal when it is not. As a result, many imbalanced data techniques are proposed in the literature, including data resampling, algorithm-based, and hybrid [32, 36]. In order to have a greater detection rate for irregular heartbeats, it is important for a heartbeat model to manage the issue of unbalanced data.

Ahamed et al. [8] presented ensembles of ML techniques, including using artificial neural networks (ANN) and LSTM. Owning to the highly imbalanced nature of the ECG (MIT-BIH and PTB) dataset, the authors proposed penalizing the loss of an ANN by assigning class weights. The ensemble outperforms conventional ML to obtain 0.9806% and 0.9766% accuracies. Similarly, Zubair & Yoon [37] proposed a cost-sensitive loss function (LF) of 1D-CNN. A LF ensures that deep representations of the classes are not skewed, helping to boost the model's ability to generalize and perform better in terms of discrimination. However, the LF used may result in computational overhead for the classifier since it is a custom LF based on a quadratic mean. Liu et al. [38] proposed a loss optimization ECG classifier based on dynamic task prioritization. The method adjusts task loss weights based on how different tasks affect learning. Nevertheless, the approach results in task bias (where few tasks get more attention or resources than others), abrupt changes in task importance (causing instability), and computational overhead. In another paper, stacked-denoising autoencoders (SDAE) and bidirectional (Bi-LSTM) were presented to learn the semantic encoding of heartbeats automatically without any tedious feature extraction. The authors used the Bi-LSTM for heartbeats categorization with semantic encoding, whereas used the SDAE as a noise-reduction mechanism. To relieve impacts from unbalanced data, the authors employed a cost-matrix LF. The MIT-BIH Arrhythmias Database and Noise Stress Test Database (NSTDB) were used to evaluate the proposed model [39]. Likewise, Khan et al. [40] suggested a cost-matrix technique for imbalanced data. Meanwhile, the cost-matrix technique is prone to design issues thereby increasing the model complexity and parameter tuning.

Also, Romdhane et al. [12] proposed an optimization step for 2-sequential 1D-CNN blocks using focal LF. A LF emphasizes minority pulse classes by giving them more weight. The authors used MIT-BIH and INCART datasets to evaluate the proposed model. Findings show that the focal LF enhanced the classifier's accuracy for the positive classes and the overall metrics. Pham et al. [3] also used focal LF as the cost-sensitive technique in the proposed ECG classification model. However, focal LF may be faced with high computational cost, potential gradient perturbation, too much emphasize on difficult samples. In another view, Khan et al. [4] used a synthetic minority oversampling technique (SMOTE) approach to solve the imbalance problem of ECG. The authors further implement a 1D convolutional ResNet to classify the balanced datasets. To evaluate the model's performance, the authors used a 10-fold CV, and the findings indicate that the suggested ResNet outperforms other 1D-CNNs. Further studies of heartbeats categorization include but are not limited to, a fall-detection using ECG [19], Belief networks application [41], Neuro-Fuzzy System [42], smart-decision based model [6].

III. METHODOLOGY

A. Dataset

The ECG dataset widely used in research is the MIT-BIH Arrhythmia Dataset. It consists of 5 classes following the AAMI standard and the PTB, with 2 classes considered for the proposed transferable CB-DuConvNetA. The MIT-BIH Arrhythmia Dataset ECG is sourced from heartbeat recordings [43]. The dataset incorporates significant events that are well represented by a small random sample of the recordings comprising Twenty-Three observations (randomly picked from One-Hundred to One-Hundred Twenty-Four, inclusive, again with some numbers missing) and Twenty-Five observations (Two-Hundred to Two-Hundred Thirty-Four with some missing numbers). The first record group is intended to be a typical sample of the waves and artifacts that might be observed during a clinical study. In contrast, the second record group comprises complex ventricular, supraventricular, and junctional arrhythmias and other defects. A large diversity in QRS shape makes nearly all selected recordings usable. Therefore, the adjusted limb lead-II (MLII) is preferably used.

The digitization process of MIT-BIH includes the “bandpass-filtered” impulses digitalized at 360Hz per signal, the sampling frequency set at 60Hz to identify arrhythmia, and likely noise peaked at 30Hz. The 11-bit samples were originally captured in a first-difference-format (FDF) of 8-bit. The analog-to-digital range used is roughly $\pm 225 mV/s$. A comprehensive annotated list of the heartbeats classes is presented in Table 1 below. It comprises 20 different overlapping classes. To make ECG classification, AAMI provided five distinct classes from the list in Table 1, detailing fifteen relevant categories as presented in Table 2.

Table 1. Heartbeat Symbols and Explanation [15]

Heartbeat Notation	Description
· or N	Normal beat
L	Left bundle branch block beat (B4)
R	Right B4
A	Atrial premature beat (APB)
a	Aberrated APB
J	Nodal (junctional) PB
S	Supraventricular PB
V	Premature ventricular contraction
F	Fusion of ventricular and normal beat
[	Start of ventricular flutter/fibrillation
!	Ventricular flutter wave
]	End of ventricular flutter/fibrillation
e	Atrial escape beat (EB)
j	Nodal (junctional) EB
E	Ventricular EB
/	Paced beat
f	Fusion of paced and normal beat
x	Non-conducted P-wave (blocked APB)
Q	Unclassifiable beat
|	Isolated QRS-like artifact

Table 2. Arrhythmia heartbeats in the MIT-BIH database based on five classes [29]

Heartbeat Category	Refined Notations
N	Normal Left/Right B4 Atrial EB Nodal EB
S	APB Aberrant APB Nodal PB Supraventricular PB
V	Premature ventricular contraction Ventricular EB
F	Fusion of ventricular and normal
Q	Paced beat Fusion of paced and normal Unclassifiable beat

In the PTB ECG, the collections were obtained based on a16 input-channel (14-ECGs, 1-respiration, 1-line voltage), ±16 mV input, ± 300 mV offset, 100 Ω (DC) resistance, and 0 to 1kHz bandwidth. The PTB ECG comprises 290 records, 148 as myocardial infarction (MI), 18 as Cardiomyopathy/Heart failure, and seven other types. PTB is generally classified into two classes: normal and abnormal.

This paper, therefore, uses the ECG datasets sourced from a Kaggle site1 consisting of a collection of both the MIT-BIH and PTB. Both MIT-BIH and PTB consist of two separate files. Using DNN technique, both datasets have been utilized to investigate the categorization of heartbeats and to examine some of the possibilities of TL. The normal and MI cases are carefully preprocessed and segmented so that a record can adequately represent a heartbeat. The classes are encoded from 0 to 4. The MIT-BIH training and test sets consist of 87,554 and 21,982 observations, each with 188 features, including the class label. Also, the PTB consists of 10506 and 4046, classified as abnormal and normal observations, respectively, with the same number of features. The detailed descriptions of the two datasets are presented in Figure 2.

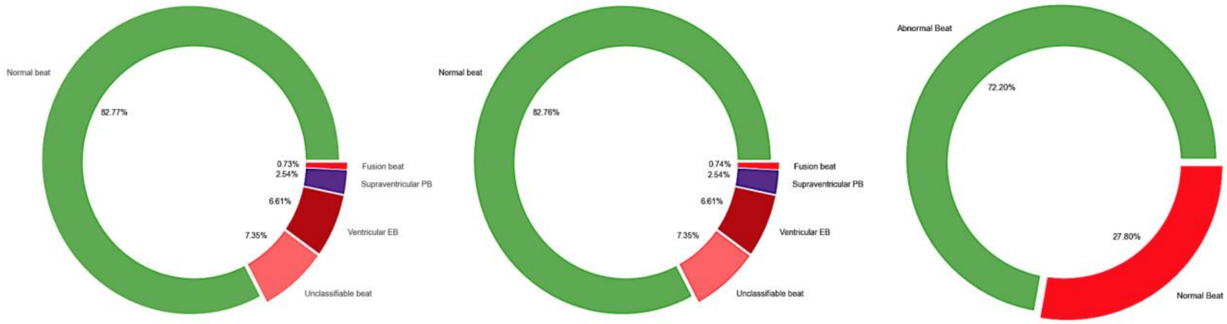
B. Preprocessing

Preprocessing provides an initial understanding of the dataset under study. It is a vital stage in the lifecycle of building a successful ML model. Pham et al. [3], in preprocessing ECG heartbeat, highlighted fundamental steps to extract meaningful beats from ECG waves shown in Figure 3. Meanwhile, this paper relies on the ECG datasets mentioned on the Kaggle site. A missing value process was performed on the four datasets; none contained any missing values. The PTB "normal" and "abnormal" datasets were combined and shuffled so that there is a good representation of the 'normal' and 'abnormal' when it is split into train, validation, and test sets for the TL model. The two datasets were further preprocessed by reshaping them to fit the proposed approach.

C. Exploratory Data Analysis

According to Kersey et al. [44], using feature schemes inherent in the ECG data can provide gainful insight into better understanding the underlying problem in a heartbeat. The paper considered some EDA techniques for this purpose. The ECG data is known for its imbalanced nature. The extent and details of the class distribution are shown in Figure 2. In Figure 2, subfigure (a), the normal beat is 82.77% in the entire MIT-BIH training set, followed by an Unclassifiable beat at 7.35%, Ventricular 6.61%, Supraventricular 2.54 and the least Fusion beat at 0.73%. The same trend is observed for the test set, as shown in subfigure b. It shows that normal is average, $1000 \times \textit{the minimum class distribution}$ indicating a densely skewed class distribution. The implication is that a developed model for ECG classification will be 100% biased toward the normal beats. The interest of this paper is to see how significantly the minority classes can be adequately represented in the classification task. The imbalance in the PTB data is also shown in subfigure c, representing a 2.5965:1 ratio of normal to abnormal beats.

1 https://www.kaggle.com/datasets/shayanfazeli/heartbeat



(a) MIT-BIH train set (b) MIT-BIH test set (c) PTB dataset

Figure 2. Class distribution of MIT-BIH and PTB Diagnostics datasets

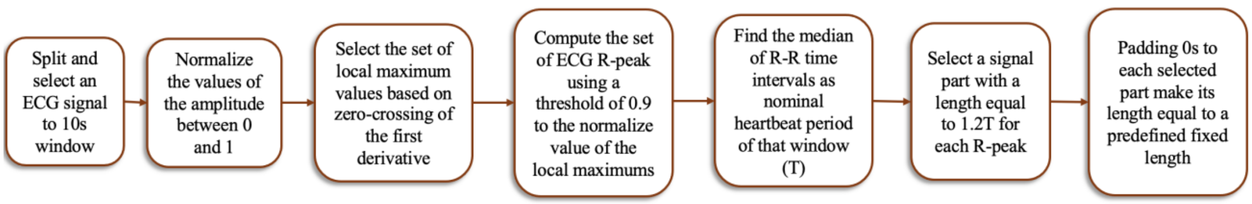


Figure 3. ECG data preprocessing stages

After carefully preprocessing and segmenting ECG data, an instance of the observation therein should provide an adequate representation of a typical heartbeat. Figure 4 illustrates such a regular heart from plotting a record in the dataset. Compared with Figure 1, it shows clearly that Figure 4 resembles a typical heartbeat. It further indicates that the raw ECG data is well-preprocessed and segmented.

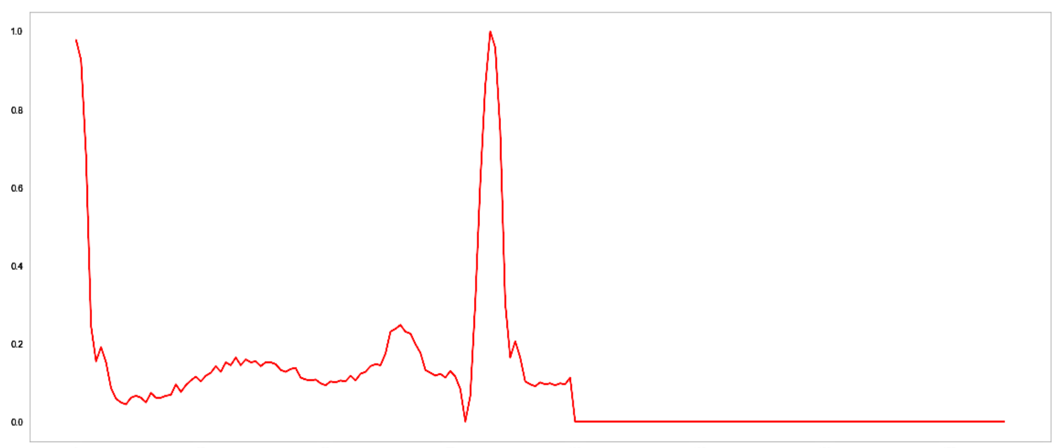


Figure 4. A typical heartbeat from the already processed and segmented ECG data

Another gray area begging for clarity is how distinguishable is the normal heartbeat to other classes. Figure 5 provides vivid clarity of normal heartbeat comparison with the other four categories.

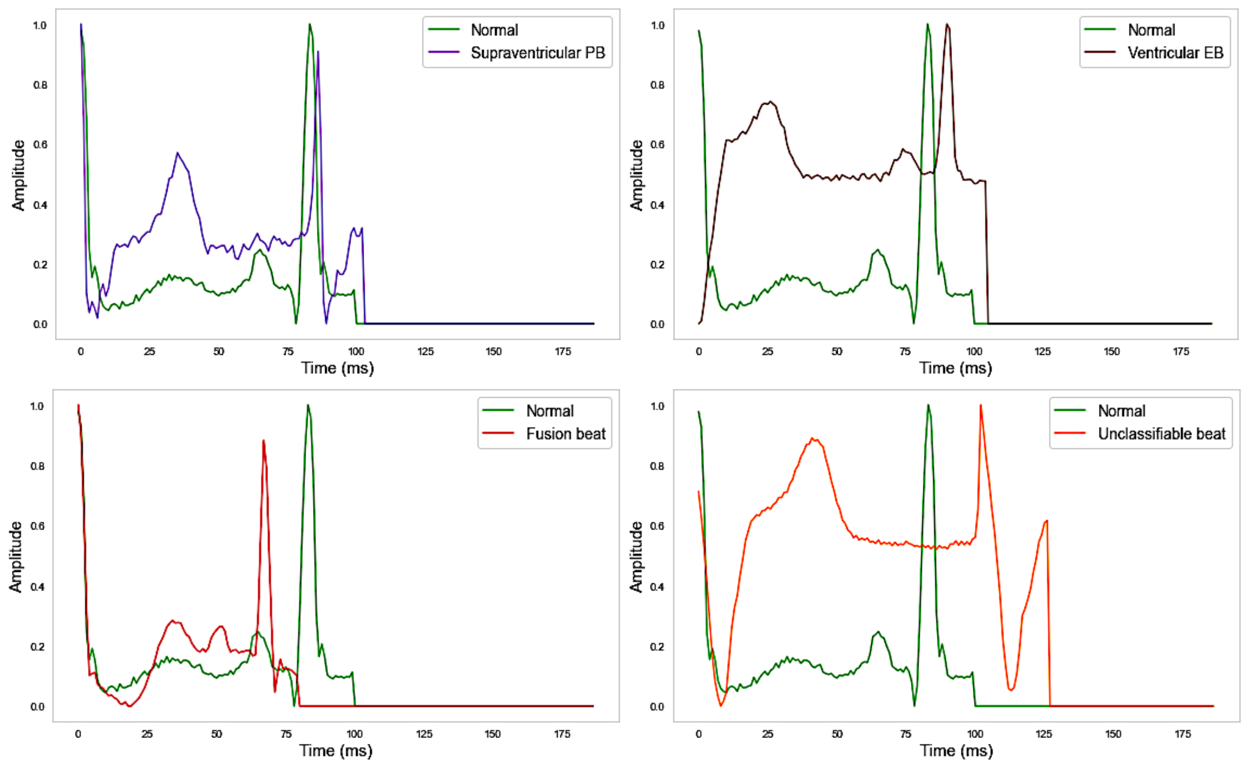


Figure 5. Comparison of the normal heartbeat with the four other classes

D. 1D-CNN

Vast studies of heartbeat classification using CNN outperforms conventional models of ML. Many of these studies have showcased the strength inherent in 2D-CNN, where ECG image data is fed to the 2D-CNN. Meanwhile, 1D-CNN remains competitive with its 2D counterpart in as much as the underlying data can be reshaped to 1D or a sequence-like format. Also, using 1D-CNN removes the overburden of converting ECG waveforms to an image without hampering the expected accuracy of the model. However, recent advances in vision transformer DL models showed impressive improvement in image classification, which applies to ECG image data or ECG waveform-image data, thereby removing any envisaged bottlenecks in ECG waveform-to-image classifiers [45, 46]. 1D-CNN convolution has also been investigated for temporal data to capture data patterns. Generally, CNN models comprise automatic feature extractors, complex non-linear relators, and classifier parts. High-level feature and pattern representation are carried out in the feature extraction process, coupled with complex non-linear transformation to learn meaningful patterns. The CNN model comprises an input channel and a series of convolutional and pooling layers (used for feature extraction and complex non-linear learning). The other part is the dense layers which classify the learned features. The concept of CNN is expressed in Equation (1) – (5):

$$Conv:\ Z_i^{k,l} = c_l + \sum_d^D \omega_d^l\, x_{i+d-1}^{ml} \tag{1}$$

$$Max-pooling:\ h_i^{k,l} = \max_{s\in S}\left(Z_{ixT+s}^{k,l}\right) \tag{2}$$

$$Dense\ layer: Z_i = \omega_i * h_i \tag{3}$$

$$RELU(Z_i) = \max\,(0, Z_i) \tag{4}$$

$$Softmax(Z_i) = {e^{Z_i}}/{\sum_l e^{Z_l}} \tag{5}$$

Where c is the bias constant, ω is the weight, S is the pooing size, and T is the stride rate. The convolution and pooling can be stacked as required, but care must be taken not to overburden the model with too many trainable parameters. The function of the pooling, then, is to decrease the feature's map size and model distortion. Equations (1) to (5) explicit the feature mappings through to classification, a series of convolutional and pooling layers (in the forward direction during training), activation layer-rectifier linear unit (RELU), dense layer, and of course, the output layer embellished with Softmax function for a multi-class task (or sigmoid for binary classification). A cost function is used to compute the loss at the output, and the result is propagated back to update the weights. The type of cost function implemented in the proposed model as one of the contributions in the paper is explained in detail in the later section. This process is repeatedly performed until the halting requirements or several epochs are satisfied.

In the proposed model, two additional techniques were considered to impact the model's generalizability and better performance, including attention mechanism and Pooling in each block.

E. Cost-Based Approach to Data Imbalance

As mentioned earlier, imbalanced data negatively impacts the true performance of any ML model. It is further stated that different methods have been studied to solve data imbalance problems in classification tasks. One method, as mentioned above, is the data-level approach of resampling. The use of oversampling is to increase samples of minority labels to the number of the majority. At the same time, the down-sampling reduces majority class(es) samples to the number of the minority class. It is expected the method would generate balanced data. However, resampling may lead to over-fitting or underfitting [37], whereas computational overhead is identified with the cost-sensitive learning techniques reviewed in the literature. Therefore, this paper presents a novel class imbalance ratio weight penalty (CIRWP) to use the inherent DL classifier's LF call. A DL model uses an in-built LF to compute the loss at the output, and the result is propagated back to update the weights (these weights are the weight of each class). The LF in DL is classified into two: categorical cross-entropy (CCE) (SparseCategoricalCrossentropy is applicable also) for multi-class and binary cross-entropy (BCE). The CCE is expressed as in Equation (6) – (8):

$$CCE = -\log\left(\frac{e^{s+ve}}{\sum_k^M e^{s_k}}\right) \tag{6}$$

Where s_{+ve} represents the +ve class, s_k represents the score of the +ve class and M as the class

$$BCE = -\frac{1}{m}\sum_{k=1}^{m}(y_k * \log \hat{y}_k + (1-y_k) * \log(1-\hat{y}_k)) \tag{7}$$

The class imbalance ratio (CIR) is expressed as:

$$CIR = \frac{\max_k\{|M_k|\}}{\min_k\{|M_k|\}} \tag{8}$$

It is expected that $\max_k\{|M_k|\}$ and $\min_k\{|M_k|\}$ returns maximum and minimum class number in all k classes. For instance, if a dataset's maximum class is 400 samples and its minimum class only contains 25, then CIR is 16. It means the +ve class will be penalized 16 times the negative class. There is, therefore, the need to formulate a class weight function from equation 8, which is expressed as in Equation (9):

$$CW_k = \frac{M_T}{k \times M_k} \tag{9}$$

Where k is the number of classes. M_T represents the total observations, and M_k is the samples in each class k in the dataset.

Since the focus is not on the majority class, then the weight penalty can be obtained by penalizing the majority class by 1 and the rest of the minority classes with the magnitude, CW_k of the majority class. The penalty is expressed as in Equation (10):

$$CIRWP_k = \frac{CW_k}{CW_{majority_k}} \tag{10}$$

Where $CW_{majority_k}$ represents the CW of the majority class

So, the penalty can be used as a cost list of the form, refer Equation (11):

$$\left[\frac{CW_j}{CW_{majority_k}}, \frac{CW_{j+1}}{CW_{majority_k}}, \dots, \frac{CW_k}{CW_{majority_k}}\right] \tag{11}$$

Where $j = 1, \dots, k$

The cost list is then assigned to the loss_weights parameter in the compile function of the proposed CB-DuConvNetA model to adjust the CCE or BCE during training dynamically.

F. Proposed Classifier Architecture

The proposed CB-DuConvNetA model comprises two parts. One part is the classifier, and the second is the transferable part. The classifier part consists of a dual 1D-CNN in parallel. The first convolutional block has three stacked convolution layers with 64 filters, 3 kernel sizes, and a RELU activation function. In addition, each convolution layer includes a Batch Normalization and Dropout rate = 0.2 layers. GlobalMaxPooling was used in the pooling process. The pooling output was reshaped to fit into the attention mechanism.

In contrast to Zubair & Yoon [37] and Kachuee et al. [29], the attention technique was introduced to improve the model's performance by focusing on the important part of the ECG data and drastically reducing persistent noise inherent in the data. Furthermore, the attention technique benefits the TL proposed in the paper by its improved and better generalizability to new or unseen instances. Attention is more robust to input variations and adept at extracting important features to provide informative patterns [47]. The output of the attention is, after that, flattened.

The second convolutional block consists of the same layer components as in block one but with some changes, including changing the filter and kernel size to 128 and 6, respectively (thereby improving the block 2 capacity and feature extraction) and using GlobalAveragePooling. Using different pooling in the two convolutional blocks allows each block flexibility and robustness and increases the benefits of each pooling technique. GlobalMaxPooling is known for dimensionality reduction by aggregating maximum activation value from the convolution layer, better invariance translation, and important feature encapsulation. In contrast, GlobalAveragePooling is known for dimensionality reduction by averaging feature map value within each pooling region, noise-variations adaptiveness, and better computational efficiency [48].

The filter, 64, used in the first convolutional block, allows for capturing low-level features in the ECG data, and the kernel size, 3, provides for local feature extraction in the low-level region. Making the first convolution a lightweight. Whereas the filters, 128, and the kernel size, 3, used in the second convolutional block ensure the model seeks a deeper high-level abstraction and discriminative representations of the data, leading to better model performance. A dropout rate of 0.2 is employed in the two convolutional blocks, including the fully-connected layers, to reduce the over-fitting of the model to an unseen validation set. In contrast, batch normalization was employed to lessen the shift within the complex non-linear interaction [37].

The flattened attention output from each convolutional block was then concatenated. The concatenated output is fed into three consecutive fully-connected layers, each with 64, 32, and 16 neurons, respectively. The output layer consists of 5 units each for the 5 classes of ECG classification and Softmax activation function. The 2 convolutional blocks used the same source of input data. The class weight penalty described in Section III (E) solves the imbalanced problem during the model training.

The second part of the proposed model is the TL. The classifier in the first part is trained and saved as a pre-trained model. After that, the import of the weights of the pre-trained model is performed. Subsequently, a layer freezing technique is performed to make the initial task trained on the model non-trainable but retain the valuable knowledge learned from it. To further adjust the model to the new target task, in this instance, the PTB, the layer-freezing approach only permits changing the weights of the last layers (usually the top layers). This process achieves the transferable representation on the second dataset for MI prediction. The TL model also used the proposed CIRWP. The overall architecture is shown in Figure 6, while the detail of the dual 1D-CNN attention is shown in Figure 7.

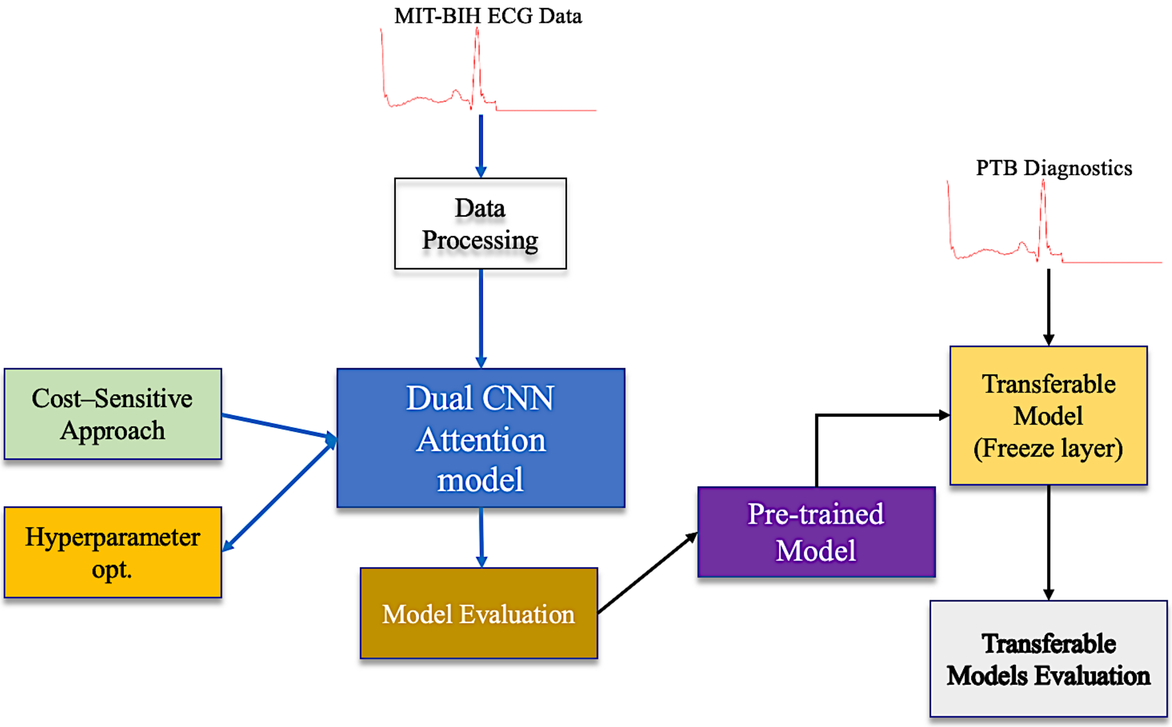


Figure 6. The overall framework of the proposed ECG transferable model

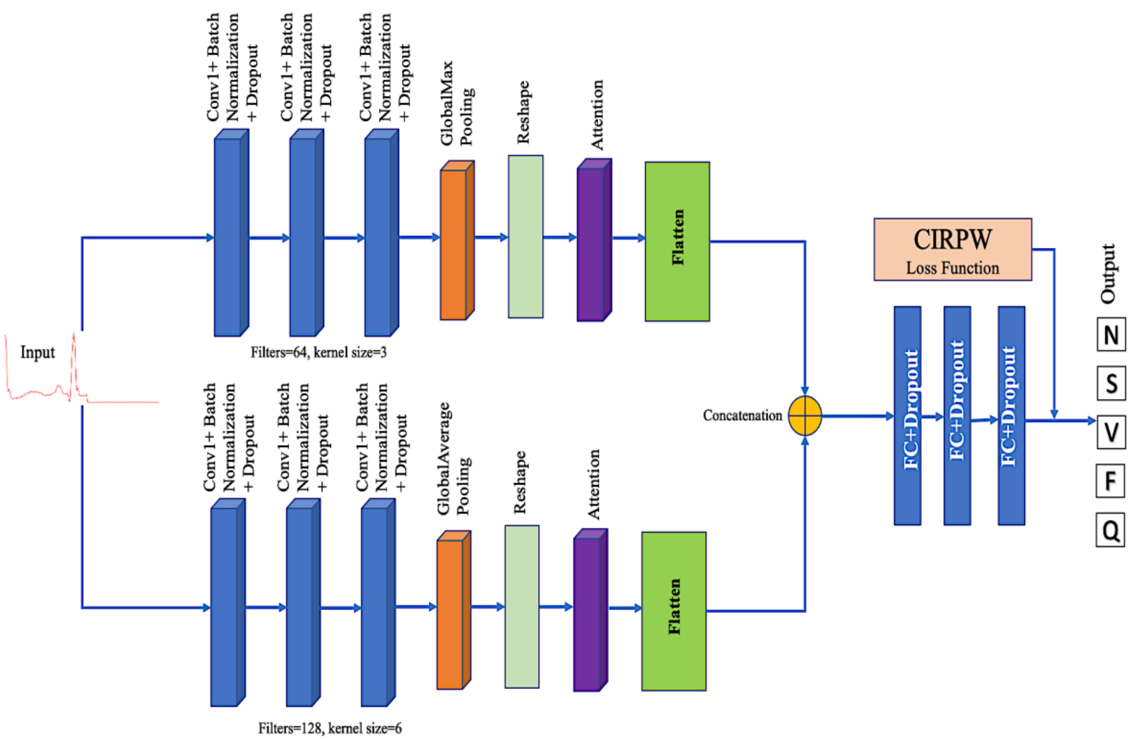


Figure 7. The proposed cost-sensitive dual 1D-CNN with Attention Network

G. Experiment setups

The proposed model is developed using Python 3.9, TensorFlow, and the Keras library [49]. Categorical cross_entropy loss with Softmax activation was used for the classifier part, while binary cross_entropy and sigmoid were used for the TL part. The Adam optimization method was used to train the classifier with a batch size of 128 for 30 epochs, while the TL runs with 100 epochs. Before the training, the MIT-BIH train set is split into a separate training and validation set with a ratio 0f 80:20. The test set is already provided. In the same approach, the PTB data was combined and shuffled. After that, it was split into three: training, validation, and test sets (train-test:80:20 and again train-validation:80:20). Earlystopping and Modelcheckpoint were implemented to obtain the best accuracy by monitoring the 'validation loss' and allow for saving the model as a pre-trained model. The model summary and hyperparameters are presented in Table 3.

Table 3. The proposed model summary and hyperparameters

CNN+Attention Model	1st 1D-CNN block			2nd 1D-CNN block			FC			Output Layer
Parameters	Conv layer 1	Conv layer 1	Conv layer 1	Conv layer 1	Conv layer 1	Conv layer 1	1st	2nd	3rd	
Input shape	(187,1)	(187,1)	(187,1)	(187,1)	(187,1)	(187,1)				
Filters	64	64	64	128	128	128				
Kernel size	3	3	3	6	6	6				
BatchNormalization			Y							
Dropout				0.2						
Pooling		GlobalMaxPooling			GlobalAveragePooling					
Attention		Y			Y					
AF			Relu							softmax
Units							64	32	16	5

IV. RESULTS AND DISCUSSION

A. Cost-Based Approach

The proposed CIRWP for a cost-sensitive approach allows the classifier to use its inherent LF by assigning the weight penalty obtained from equations 9, 10, and 11. The computed weight penalty for both MIT-BIH and PTB datasets is presented in Table 4.

Table 4. Weight penalty computation for the cost-based model

MIT-BIH+					PTB Diagnostics++				
Classes	Category	Counts	Class weight	Penalty	Classes	Category	Counts	Class weight	Penalty
0	N	72471	0.24162	1	0	Normal	4046	1.79831	2.5965
1	S	2223	7.8771	32.6005	1	Abnormal	10506	0.6926	1
2	V	5788	3.02536	12.5209					
3	F	641	27.3179	113.0593					
4	Q	6431	2.72287	11.2690					
Total		87554					14552		

+ for Class 0 Penalty: (87554/(5x72471) = 0.24162/0.24162=1 ++for Class 0 Penalty: (14552/(2x4046) =1.79831/0.6926=2.5965)

B. Performance Metrics

In evaluating the proposed model performance, the paper considered the confusion matrix and other metrics obtainable from it, including recall, precision, F1-score, and accuracy expressed as in Equation (12) – (15):

$$accuracy = \frac{T_p+T_n}{T_p+F_n+T_n+F_p} \quad (12)$$

$$recall = \frac{T_p}{T_p+F_n} \quad (13)$$

$$precision = \frac{T_p}{T_p + F_p} \tag{14}$$

$$F1 - score = \frac{2 * precision * recall}{precision + recall} \tag{15}$$

Where T_p is true +ve, T_n is true -ve, F_p is false +ve, and F_n is false -ve. Because accuracy considers all classes equally, a model trained using imbalanced data can be biased, yielding high classification accuracy; therefore, attention is placed more on recall, precision, and F1-Score.

C. MIT-BIH ECG Classification

The proposed model was trained using a batch size of 128 over 30 epochs. Earlystopping and Modelcheckpoint were implemented to obtain the best accuracy by monitoring the 'validation loss' and allowing for saving the model as a pre-trained model. The proposed model performance for the 5 categories of ECG is presented in Table 5, indicating the cost-based approach implemented within the classifier provided a fair representation of the minority classes to obtain an F1-score and accuracy of 98.08% and 98.14%, respectively. The model accuracy and loss are shown in Figure 8. It shows that the model's overfitting is well reduced, and the best accuracy is obtained.

Furthermore, the performance of the proposed model is compared with well-performing ML algorithms, including RF, classification and regression tree (CART), KNN, NB, SVM, XGBOOST, and LightGBM. Some of the ML implemented are built with the capability to automatically handle imbalanced data, such as XGBOOST and LightGBM, while SVM can be trained as a cost-sensitive model. SMOTE was used to balance the data for the rest of the ML model with no inherent ability to handle imbalanced data. Therefore, the proposed model was also trained with SMOTE-balanced ECG data.

Table 5. The proposed model performance for the 5 classes of MIT-BIH ECG data

Classes	Precision	Recall	F1-score	Support
N	0.99	0.99	0.99	18118
S	0.89	0.75	0.82	556
V	0.96	0.94	0.95	1448
F	0.81	0.65	0.72	162
Q	0.99	0.98	0.99	1608

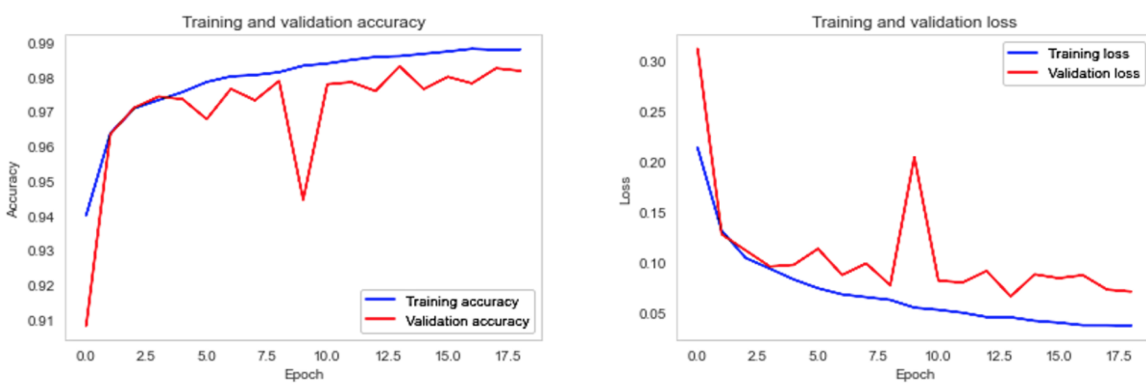
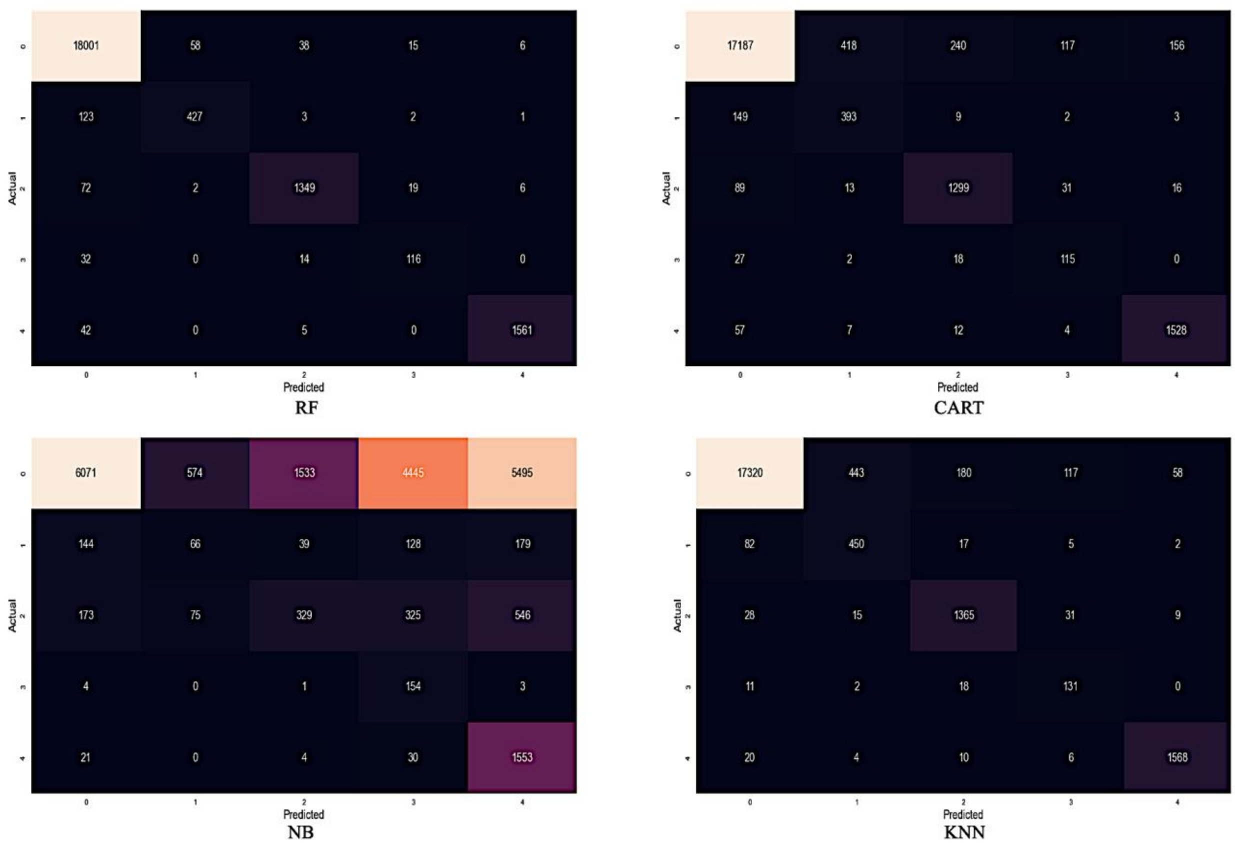


Figure 8. The proposed model accuracy and loss

The models' performances are presented in Table 6. It is observed that RF obtains F1-score of 89.45% to outperform other models in the same category, while XBOOST and LightGBM obtain 88.89% and 88.56%, respectively, showing their natural ability to make good predictions for imbalanced data. Meanwhile, DuConvNetA+SMOTE obtains a 97.78% F1-score outperforming the conventional models. It indicates that the DuConvNetA base classifier has a significant performance. The proposed model obtains an F1-score of 98.08% outperforming all the other models. The performance of the proposed model is also validated by the positive predictions in the confusion matrix in Figure 9 compared to other models.

Table 6. Comparison of the proposed model with conventional ML models

Models	Precision	Recall	F1-score	Accuracy
RF	0.9149	0.8756	0.8945	0.9800
CART	0.7203	0.8425	0.7692	0.9374
KNN	0.7505	0.8984	0.8061	0.9517
NB (Multinomial)	0.2883	0.5195	0.2370	0.3733
SVM	0.6958	0.9154	0.7516	0.9302
XGBOOST	0.9527	0.8416	0.8889	0.9790
LightGBM	0.9334	0.8490	0.8856	0.9781
DuConvNetA+SMOTE	0.9788	0.9778	0.9782	0.9778
Proposed Model	**0.9807**	**0.9814**	**0.9808**	**0.9814**



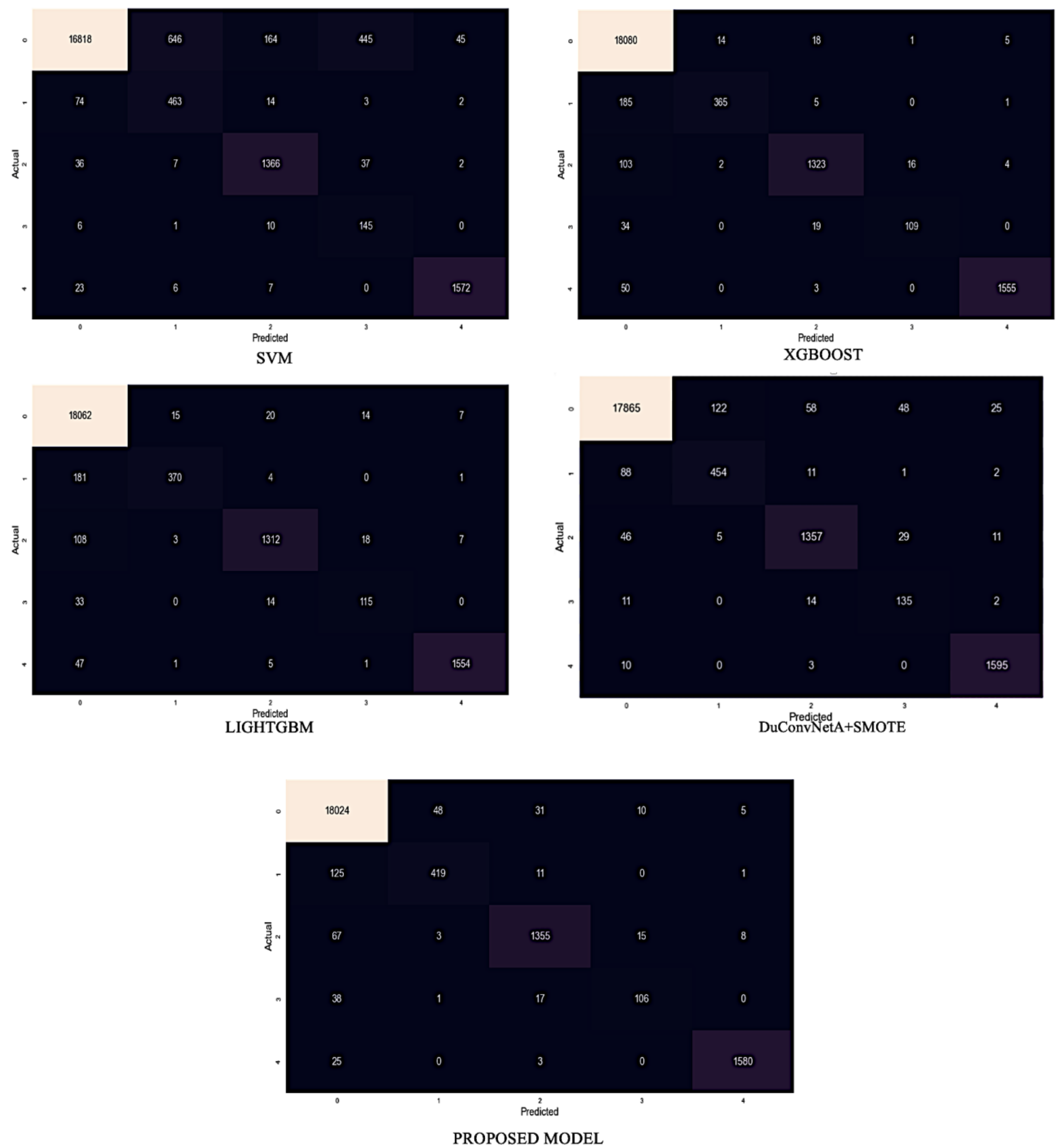


Figure 9. Confusion matrix of models considered in the paper

D. Ablation Study

Meanwhile, an ablation study was performed to show the effect of varying the batch size of the base pre-trained model as a manual approach to hyperparameter tuning. The result shows both the accuracy and F1-score of the base pre-trained improve to a certain point, stabilize and decrease as the batch size increases. It typically indicates that the model performs better than the unseen data for batch sizes 8, 16, 32, and 64. but sharply declines at batch sizes equal to 256, while accuracy reaches its best point of 0.9814 at batch sizes equal to 128, the highest among all batch sizes, as shown in Figure 10. The variations observed here may be due to various factors, including the characteristics of ECG data, the proposed model's complexity, and the optimizer used during training. Moreover, the scope of the paper did not extensively cover some well-known hyperparameter tuning such as random, grid, Bayesian and genetic search algorithms [50], [51] aside from the manual approach implemented.

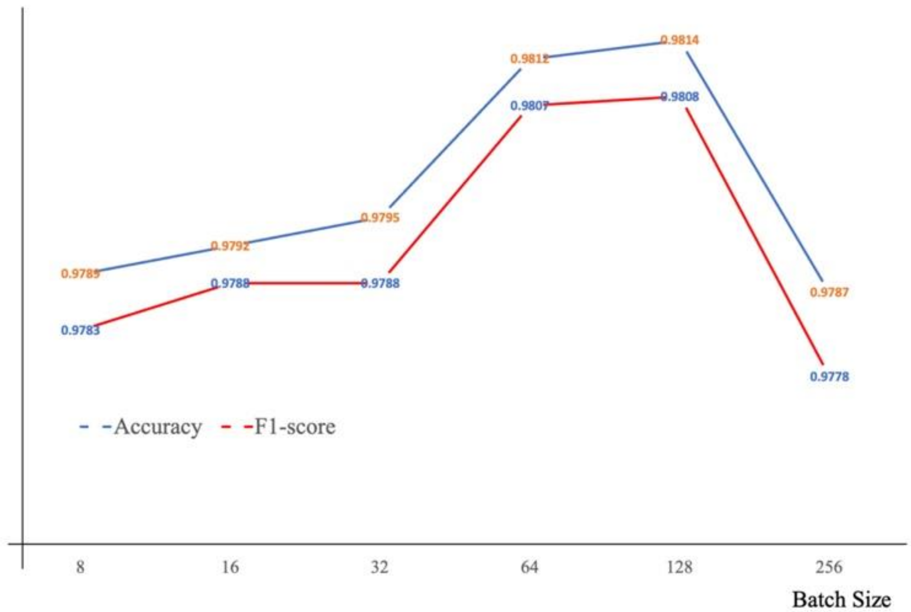


Figure 10. Comparison of batch sizes with base classifier's performance

E. Transfer Learning: PTB Diagnostics

The TL described in the last part of Section III (F), comprising the layer-freezing method, was implemented to allow the transferable representation on the PTB dataset for MI prediction. Adam optimizer, batch size 32, and binary cross_entropy parameters were implemented in the TL and ran for 100 epochs. The prediction result is presented in Table 7. Normal obtains an F1-score of 97%, while abnormal has 99%. It shows that the proposed classifier as a pre-trained model has a greater transferable representation on the PTB. The accuracy and loss of the TL model are shown in Figure 11, indicating better generalization to the unseen PTB data. The TL on PTB obtains F1-score and accuracy of **98.45%** and **98.45%**, respectively. The confusion matrix in Figure 12 further validates the model's robust performance indicating a reduced F_n rate compared to F_p rate (meaning practitioners are more interested in cases predicted as normal heartbeats but are actually unhealthy ones).

Table 7. TL results on the PTB data

Classes	Precision	Recall	F1-score	Support
Normal	0.97	0.97	0.97	809
Abnormal	0.99	0.99	0.99	2102

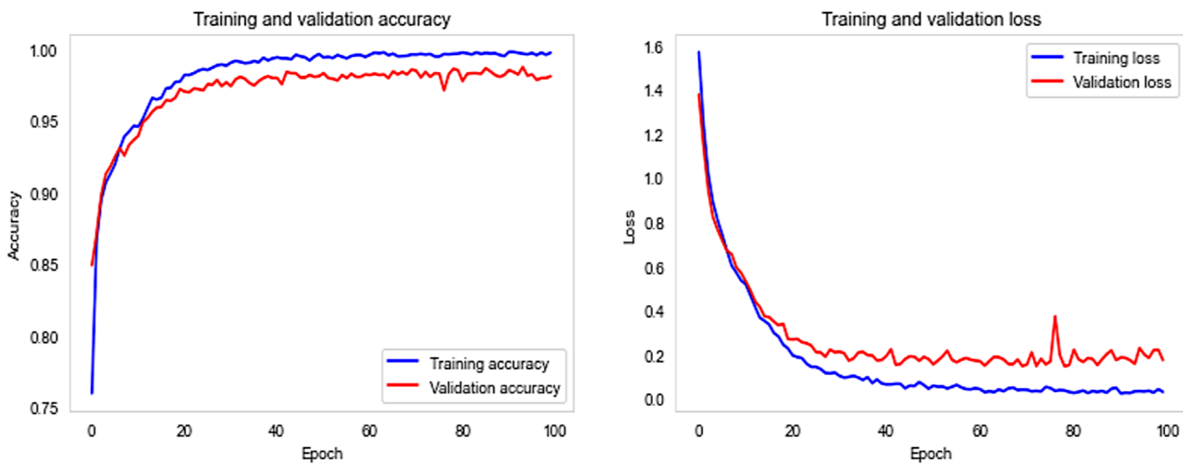


Figure 11. The proposed TL model accuracy and loss

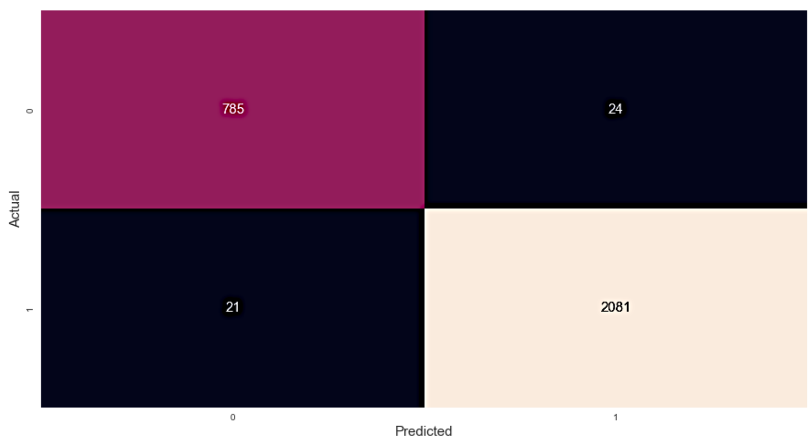


Figure 12. Confusion matrix showing the performance of the TL model on the PTB dataset

In addition to comparing the proposed model with convectional ML models, a comparison with existing works in the literature is carried out. The result is presented in Tables 8 and 9, indicating the proposed model's competitiveness with past studies.

Table 8. Comparison of classifier for MIT-BIH ECG classification

Work	Approach	Accuracy (%)	F1-score (%)
[29]	Augmentation+ Deep residual CNN	93.4	-
[8]	Class weight+ANN	98.06	-
[3]	Focal loss+CNN (squeeze-and-excitation)	98.56	-
[18]	Exponential-political optimizer trained+Deep quantum neural network	91.40	-
This work	**CIRWP+DuConvNetA**	**98.14**	**98.08**

Table 9. Comparison of TL approach on PTB

Work	Accuracy (%)	F1-score (%)	Precision (%)	Recall (%)
[29]	95.9	-	95.2	95.1
[39]	98.25	-	-	96.32
[8]	97.66	-	96.90	97.06
[12]	98.41	98.38	98.37	98.41
[38]	96.5	77.7	-	-
[3]	98.28	-	99.9	97.72
[18]	90.9	-	-	91.7
This work	**98.45**	**98.45**	**98.45**	**98.45**

V. CONCLUSION

This paper explored the cost-sensitive approach of the DL model for ECG classification. A CIRWP approach that allows the base classifier to use its inherent loss function by assigning class weight was proposed and implemented. The CIRWP overcomes computational overhead common to most cost-sensitive techniques. In addition, a dual 1D-CNN with an attention mechanism with each block of the convolution exploring different pooling techniques was developed to implement the CIRWP. It also noted that the attention and varying pooling techniques contributed immensely to the model performance. Two ECG datasets (MIT-BIH and PTB Diagnostics were used in all the experiments). Furthermore, the advantage of TL was advanced in this paper, and a practical experiment demonstrated that a well-performing classifier could provide optimal transferable representation on another task (unseen data). To compare the proposed model performance, well-performing conventional ML algorithms were used.

In all experiments, the proposed model obtains an outstanding performance compared to other models with an F1-score and accuracy of 98.08% and 98.14% on the MMIT-BIH data, while the TL model obtains 98. 45% for both F1-score and accuracy.

A comparison of CIRWP with other cos-sensitive techniques in terms of complexity and stability is suggested for future work. Also, future research can explore well-known hyperparameter tuning such as random, grid, Bayesian and genetic search algorithms [50], [51] to improve probable misclassifications difficulties in the proposed model.

ACKNOWLEDGEMENT

This work is funded by Ministry of Higher Education Malaysia, Fundamental Research Grant Scheme [Grant Number: FRGS/1/2022/STG06/USM/02/4], for the Project entitled “Efficient Joint Process Monitoring using a New Robust Variable Sample Size and Sampling Interval Run Sum Scheme”.

REFERENCES

[1] T. T. Mayabee, K. T. Haque, S. B. Alam, R. Rahman, M. A. Amin, and S. Kobashi, “ECG Signal Classification Using Transfer Learning and Convolutional Neural Networks,” M. S. Kaiser, S. Waheed, A. Bandyopadhyay, M. Mahmud, and K. Ray, Eds., in Lecture Notes in Networks and Systems, vol. 618. Singapore: Springer Nature Singapore, 2023. doi: 10.1007/978-981-19-9483-8.

[2] Mohebbanaaz, L. V. R. Kumar, and Y. P. Sai, “A new transfer learning approach to detect cardiac arrhythmia from ECG signals,” Signal Image Video Process, vol. 16, no. 7, pp. 1945–1953, 2022, doi: 10.1007/s11760-022-02155-w.

[3] B.-T. Pham, P. T. Le, T.-C. Tai, Y.-C. Hsu, Y.-H. Li, and J.-C. Wang, “Electrocardiogram Heartbeat Classification for Arrhythmias and Myocardial Infarction,” Sensors, vol. 23, no. 6, p. 2993, 2023, doi: 10.3390/s23062993.

[4] F. Khan, X. Yu, Z. Yuan, and A. U. Rehman, “ECG classification using 1-D convolutional deep residual neural network,” PLoS One, vol. 18, no. 4, p. e0284791, 2023, doi: 10.1371/journal.pone.0284791.

[5] M. Alfaras, M. C. Soriano, and S. Ortín, "A Fast Machine Learning Model for ECG-Based Heartbeat Classification and Arrhythmia Detection," Front Phys, vol. 7, no. July, pp. 1–11, 2019, doi: 10.3389/fphy.2019.00103.

[6] S. Tiwari et al., "A smart decision support system to diagnose arrhythymia using ensembled ConvNet and ConvNet-LSTM model," Expert Syst Appl, vol. 213, no. PA, p. 118933, 2023, doi: 10.1016/j.eswa.2022.118933.

[7] K. Y. Ngiam and I. W. Khor, "Big data and machine learning algorithms for healthcare delivery," Lancet Oncol, vol. 20, no. 5, pp. e262–e273, May 2019, doi: 10.1016/S1470-2045(19)30149-4.

[8] M. A. Ahamed, K. A. Hasan, K. F. Monowar, N. Mashnoor, and M. A. Hossain, "ECG heartbeat classification using ensemble of efficient machine learning approaches on imbalanced datasets," 2020 2nd International Conference on Advanced Information and Communication Technology, ICAICT 2020, no. November, pp. 140–145, 2020, doi: 10.1109/ICAICT51780.2020.9333534.

[9] M. S. Bin Sinal and E. Kamioka, "Early abnormal heartbeat multistage classification by using decision tree and K-nearest neighbor," ACM International Conference Proceeding Series, pp. 29–34, 2018, doi: 10.1145/3299819.3299848.

[10] Z. Ahmad, A. Tabassum, L. Guan, and N. Khan, "ECG heartbeat classification using multimodal image fusion," ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, vol. 2021-June, pp. 1330–1334, 2021, doi: 10.1109/ICASSP39728.2021.9414709.

[11] Y. Zhang, J. Li, S. Wei, F. Zhou, and D. Li, "Heartbeats Classification Using Hybrid Time-Frequency Analysis and Transfer Learning Based on ResNet," IEEE J Biomed Health Inform, vol. 25, no. 11, pp. 4175–4184, 2021, doi: 10.1109/JBHI.2021.3085318.

[12] T. F. Romdhane, H. Alhichri, R. Ouni, and M. Atri, "Electrocardiogram heartbeat classification based on a deep convolutional neural network and focal loss," Comput Biol Med, vol. 123, no. July, p. 103866, 2020, doi: 10.1016/j.compbiomed.2020.103866.

[13] E. Essa and X. Xie, "An Ensemble of Deep Learning-Based Multi-Model for ECG Heartbeats Arrhythmia Classification," IEEE Access, vol. 9, pp. 103452–103464, 2021, doi: 10.1109/ACCESS.2021.3098986.

[14] F. I. Alarsan and M. Younes, "Analysis and classification of heart diseases using heartbeat features and machine learning algorithms," J Big Data, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0244-x.

[15] A. L. Goldberger et al., "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," Circulation, vol. 101, no. 23, pp. e215--e220, 2000, [Online]. Available: https://www.physionet.org/content/physiobank

[16] F. Zhuang et al., "A Comprehensive Survey on Transfer Learning," Proceedings of the IEEE, vol. 109, no. 1, pp. 43–76, 2021, doi: 10.1109/JPROC.2020.3004555.

[17] K. Weiss, T. M. Khoshgoftaar, and D. D. Wang, A survey of transfer learning, vol. 3, no. 1. Springer International Publishing, 2016. doi: 10.1186/s40537-016-0043-6.

[18] A. Choudhury, S. Vuppu, S. Pratap Singh, M. Kumar, and S. Nakharu Prasad Kumar, "ECG-based heartbeat classification using exponential-political optimizer trained deep learning for arrhythmia detection," Biomed Signal Process Control, vol. 84, no. January, p. 104816, 2023, doi: 10.1016/j.bspc.2023.104816.

[19] F. S. Butt, L. La Blunda, M. F. Wagner, J. Schäfer, I. Medina-Bulo, and D. Gómez-Ullate, "Fall detection from electrocardiogram (ECG) signals and classification by deep transfer learning," Information (Switzerland), vol. 12, no. 2, pp. 1–22, 2021, doi: 10.3390/info12020063.

[20] Q. U. A. Mastoi et al., "Novel DERMA Fusion Technique for ECG Heartbeat Classification," Life, vol. 12, no. 6, pp. 1–13, 2022, doi: 10.3390/life12060842.

[21] S. Nurmaini, A. Darmawahyuni, A. N. S. Mukti, M. N. Rachmatullah, F. Firdaus, and B. Tutuko, "Deep learning-based stacked denoising and autoencoder for ECG heartbeat classification," Electronics (Switzerland), vol. 9, no. 1, pp. 1–17, 2020, doi: 10.3390/electronics9010135.

[22] S. Aziz, S. Ahmed, and M. S. Alouini, "ECG-based machine-learning algorithms for heartbeat classification," Sci Rep, vol. 11, no. 1, pp. 1–14, 2021, doi: 10.1038/s41598-021-97118-5.

[23] A. M. Alqudah, S. Qazan, L. Al-Ebbini, H. Alquran, and I. A. Qasmieh, ECG heartbeat arrhythmias classification: a comparison study between different types of spectrum representation and convolutional neural networks architectures, vol. 13, no. 10. Springer Berlin Heidelberg, 2022. doi: 10.1007/s12652-021-03247-0.

[24] F. Murat, O. Yildirim, M. Talo, U. B. Baloglu, Y. Demir, and U. R. Acharya, "Application of deep learning techniques for heartbeats detection using ECG signals-analysis and review," Comput Biol Med, vol. 120, no. March, p. 103726, 2020, doi: 10.1016/j.compbiomed.2020.103726.

[25] Q. Xie, S. Tu, G. Wang, Y. Lian, and L. Xu, "Feature enrichment based convolutional neural network for heartbeat classification from electrocardiogram," IEEE Access, vol. 7, pp. 153751–153760, 2019, doi: 10.1109/ACCESS.2019.2948857.

[26] M. K. Gajendran, M. Z. Khan, and M. A. K. Khattak, "ECG classification using deep transfer learning," Proceedings - 2021 4th International Conference on Information and Computer Technologies, ICICT 2021, pp. 1–5, 2021, doi: 10.1109/ICICT52872.2021.00008.

[27] I. Moran, D. T. Altilar, M. K. Ucar, C. Bilgin, and M. R. Bozkurt, “Deep Transfer Learning for Chronic Obstructive Pulmonary Disease Detection Utilizing Electrocardiogram Signals,” IEEE Access, vol. 11, no. March, pp. 40629–40644, 2023, doi: 10.1109/ACCESS.2023.3269397.

[28] A. Pal, R. Srivastva, and Y. N. Singh, “CardioNet: An Efficient ECG Arrhythmia Classification System Using Transfer Learning,” Big Data Research, vol. 26, p. 100271, 2021, doi: 10.1016/j.bdr.2021.100271.

[29] M. Kachuee, S. Fazeli, and M. Sarrafzadeh, “ECG heartbeat classification: A deep transferable representation,” Proceedings - 2018 IEEE International Conference on Healthcare Informatics, ICHI 2018, pp. 443–444, 2018, doi: 10.1109/ICHI.2018.00092.

[30] A. Bizzego, G. Gabrieli, and G. Esposito, “Deep neural networks and transfer learning on a multivariate physiological signal dataset,” Bioengineering, vol. 8, no. 3, pp. 1–12, 2021, doi: 10.3390/bioengineering8030035.

[31] Y. Cui, M. Jia, T. Y. Lin, Y. Song, and S. Belongie, “Class-balanced loss based on effective number of samples,” Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2019-June, pp. 9260–9269, 2019, doi: 10.1109/CVPR.2019.00949.

[32] J. M. Johnson and T. M. Khoshgoftaar, “Survey on deep learning with class imbalance,” J Big Data, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0192-5.

[33] M. Toğaçar, K. B. Özkurt, B. Ergen, and Z. Cömert, “BreastNet: A novel convolutional neural network model through histopathological images for the diagnosis of breast cancer,” Physica A: Statistical Mechanics and its Applications, vol. 545, p. 123592, 2020, doi: 10.1016/j.physa.2019.123592.

[34] H. He, W. Zhang, and S. Zhang, “A novel ensemble method for credit scoring: Adaption of different imbalance ratios,” Expert Syst Appl, vol. 98, pp. 105–117, 2018, doi: 10.1016/j.eswa.2018.01.012.

[35] A. C. Kae, C. Xinying, J. O. Victor, and K. K. Wah, “Employee Turnover Prediction by Machine Learning Techniques,” Journal of Telecommunication, Electronic and Computer Engineering, vol. 13, no. 4, pp. 49–56, 2021, [Online]. Available: https://jtec.utem.edu.my/jtec/article/view/6148/4063

[36] S. Wang, W. Liu, J. Wu, L. Cao, Q. Meng, and P. J. Kennedy, “Training deep neural networks on imbalanced data sets,” Proceedings of the International Joint Conference on Neural Networks, vol. 2016-Octob, pp. 4368–4374, 2016, doi: 10.1109/IJCNN.2016.7727770.

[37] M. Zubair and C. Yoon, “Cost-Sensitive Learning for Anomaly Detection in Imbalanced ECG Data Using Convolutional Neural Networks,” Sensors, vol. 22, no. 11, 2022, doi: 10.3390/s22114075.

[38] J. Liu, M. Zeng, K. Shan, and L. Tian, “Loss Optimization Based Algorithm for Multi-classification of ECG Signal,” ACM International Conference Proceeding Series, pp. 537–541, 2022, doi: 10.1145/3570773.3570856.

[39] E. K. Wang, X. Zhang, and L. Pan, “Automatic Classification of CAD ECG Signals with SDAE and Bidirectional Long Short-Term Network,” IEEE Access, vol. 7, no. Ld, pp. 182873–182880, 2019, doi: 10.1109/ACCESS.2019.2936525.

[40] S. H. Khan, M. Hayat, M. Bennamoun, F. A. Sohel, and R. Togneri, “Cost-sensitive learning of deep feature representations from imbalanced data,” IEEE Trans Neural Netw Learn Syst, vol. 29, no. 8, pp. 3573–3587, 2018, doi: 10.1109/TNNLS.2017.2732482.

[41] S. M. Mathews, C. Kambhamettu, and K. E. Barner, “A novel application of deep learning for single-lead ECG classification,” Comput Biol Med, vol. 99, no. May 2017, pp. 53–62, 2018, doi: 10.1016/j.compbiomed.2018.05.013.

[42] M. L. Talbi, A. Charef, F. Abdelliche, and P. Ravier, “Heartbeat Classification Using ANFIS System and QRS Complex Features Extraction,” ACM International Conference Proceeding Series, 2020, doi: 10.1145/3432867.3432892.

[43] 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, doi: 10.1109/51.932724.

[44] L. Kersey, K. Lilly, and N. Park, “ECG heartbeat classification: An exploratory study,” ACM International Conference Proceeding Series, pp. 20–27, 2018, doi: 10.1145/3314487.3314491.

[45] A. Vaid et al., “A foundational vision transformer improves diagnostic performance for electrocardiograms,” NPJ Digit Med, vol. 6, no. 1, 2023, doi: 10.1038/s41746-023-00840-9.

[46] Y. Dong, M. Zhang, L. Qiu, L. Wang, and Y. Yu, “An Arrhythmia Classification Model Based on Vision Transformer with Deformable Attention,” Micromachines (Basel), vol. 14, no. 6, p. 1155, 2023, doi: 10.3390/mi14061155.

[47] Z. Niu, G. Zhong, and H. Yu, “A review on the attention mechanism of deep learning,” Neurocomputing, vol. 452, pp. 48–62, 2021, doi: 10.1016/j.neucom.2021.03.091.

[48] S. N. Chakravarthy, H. Singhal, and N. R. P. Yadav, “DR-NET: A Stacked Convolutional Classifier Framework for Detection of Diabetic Retinopathy,” Proceedings of the International Joint Conference on Neural Networks, vol. 2019-July, no. 1, 2019, doi: 10.1109/IJCNN.2019.8852011.

[49] M. Abadi et al., "TensorFlow: A system for large-scale machine learning," in Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation, OSDI 2016, 2016, pp. 265–283.

[50] T. Agrawal, Hyperparameter Optimization in Machine Learning. 2021. doi: 10.1007/978-1-4842-6579-6.

[51] R. Turner et al., "Bayesian Optimization is Superior to Random Search for Machine Learning Hyperparameter Tuning: Analysis of the Black-Box Optimization Challenge 2020," Proc Mach Learn Res, vol. 133, pp. 3–26, 2020.

APPENDIX: List of some abbreviations used in the paper

2D-CNN	Two-dimensional Convolutional Neural Networks
1D-CNN	One-dimensional Convolutional Neural Networks
AAMI	Advancement of Medical Instrumentation
AI	Artificial Intelligence
CAD	Computer-Aided Diagnostic
CaDi	Cardiovascular Disorders
CB-DuConvNetA	Cost-Based Dual ConvNet-Attention
CIR	Class Imbalance Ratio
CIRWP	Class Imbalance Ratio Weight Penalty
DL	Deep Learning
DT	Decision Tree
ECG	Electrocardiogram
GDBT	Gradient-Boosted Tree
Ha	Heart arrhythmia
KNN	K-Nearest Neighbor
LF	Loss Function
LSTM	Long-term, Short-Term Memory
MI	Myocardial Infarction
ML	Machine Learning
MLP	Multi-layer Perceptron
NB	Naïve Bayes
PTB	Physikalisch-Technische Bundesanstalt
RELU	Rectifier Linear Unit
RF	Random Forest
SMOTE	Synthetic Minority Oversampling Technique
SVM	Support Vector Machine
TL	Transfer Learning
+ve	Positive
-ve	Negative