Turkish Journal of Electrical Engineering and Computer Sciences

Volume 32 | Number 4 Article 5

7-26-2024

Ensemble Learning for Accurate Prediction of Heart Sounds using Gammatonegram Images

SINAM ASHINIKUMAR SINGH

SINAM Ajitkumar SINGH

AHEIBAM DINAMANI SINGH

Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2024) 32: 555 – 573

doi:10.55730/1300-0632.4087

Ensemble learning for accurate prediction of heart sounds using gammatonegram images

Sinam Ashinikumar SINGH[1*], **Sinam Ajitkumar SINGH**[2], **Aheibham Dinamani SINGH**[1]
[1]Department of Electronics and Communication Engineering, National Institute of Technology Manipur, Imphal, Manipu, India
[2]Department of Electronics, Dhanamanjuri University, Imphal, Manipur, India

Received: 09.01.2024 • **Accepted/Published Online:** 23.05.2024 • **Final Version:** 26.07.2024

Abstract: The analysis of heart sound signals constitutes a pivotal domain in healthcare, with the prediction of imbalanced heart sounds offering critical diagnostic insights. However, the inherent diversity in cardiac sound patterns presents a substantial challenge in predicting imbalanced signals. Many scientific disciplines have focused a great deal of emphasis on the problem of class inequality. We introduce an ensemble learning approach employing a convolutional neural network model-based deep learning algorithm to effectively tackle the challenges associated with predicting imbalanced heart sound signals. We use a gammatone filter bank to extract relevant features from the heard sound signal. Our approach leverages a pretrained convolutional neural network architecture, fine-tuning it with gammatonegram images to improve the classification performance. To overcome the challenges posed by imbalanced datasets, we integrate data augmentation into the image processing pipeline. The images are subsequently subjected to classification through deep convolutional neural network employing a transfer learning technique. This involves the utilization of convolutional neural network models such as AlexNet, SqueezeNet, GoogLeNet, and VGG19 to address concerns related to model overfitting. Our experimental results are rigorously validated using the publicly accessible PhysioNet 2016 dataset. The proposed ensemble methodology, incorporating AlexNet, SqueezeNet, and VGG19 models, demonstrated superior performance, attaining an accuracy of 99.51%, a sensitivity rate of 99.34%, and a specificity rate of 99.67%. These results emphasize the substantial clinical promise inherent in our methodology, particularly in the realm of identifying imbalanced and noisy heart sound signals. This, in turn, serves to advance the diagnosis of cardiovascular diseases.

Key words: PhysioNet, phonocardiogram, gammatonegram, deep learning, convolutional neural network

1. Introduction

About 17.9 million deaths annually, or 31% of all deaths worldwide, are caused by cardiovascular diseases (CVDs), which pose a serious threat to global public health [1]. According to data from the Centers for Disease Control (CDC), heart-related deaths occur in the US once every 36 seconds [2]. This concerning figure highlights how serious CVDs are as a public health issue. The impact of CVDs grows exponentially in low- and middle-income countries due to the lack of doctors with the necessary training and the limited availability of diagnostic

*Correspondence: ashini.sinam@gmail.com

[1]Cardiovascular diseases (2021). Website https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) [accessed 2023 Sep 24].

[2]CDC. Centers for Disease Control and Prevention. 2023. Heart Disease Facts | cdc.gov. Website https://www.cdc.gov/heartdisease/facts.htm [accessed 2023 Sep 24].

equipment [3]. Additionally, it has been reported that the range of auscultation efficiency among medical trainees and primary care physicians is between 20% and 40%, as cited in references [1–3]. Expert cardiologists, on the other hand, get an accuracy rate of about 80% [2]. Cost-effective screening tool usage is on the rise, indicating a recognition of the critical role early diagnosis plays in reducing the burden of CVDs. This involves the use of computerized stethoscopes that have machine learning algorithms for murmur analysis [4].

A great deal of study has been done in the area of automated cardiac sound analysis for CVD detection. The availability of publicly accessible datasets has significantly accelerated research progress in this field, as noted in [5], with particular attention to the PhysioNet 2016 Challenge [6] and the databases referenced in [7]. The 3240 Phonocardiogram (PCG) recordings in this dataset were obtained using 7 distinct stethoscopes and carefully annotated to distinguish between normal and pathological heart states.

1.1. Literature review

Various techniques have been examined to detect anomalies in heart sound (HS) by utilizing the aforementioned dataset. Numerous front-end features have been integrated into these methods, such as features obtained from temporal, frequency, and statistical data [8], mel-frequency cepstral coefficients (MFCC) [9, 10], and continuous wavelet transform (CWT) [11]. Recent studies have shown that time-frequency-based features, including CWT and short-time Fourier transform (STFT), are the most popular choice for feature extraction. Several classifiers have been studied in the literature, including the k-nearest neighbor (k-NN) [12], support vector machine (SVM) [13], random forest [14], multilayer perceptron (MLP) [15], and deep learning approaches using 1D and 2D convolutional neural networks (CNNs) [16, 17]. Though the study in [18] tackled the problem of domain variability in HS prediction, it did not directly address the problems associated with noise and imbalanced datasets.

In the literature, numerous researchers have employed CNN-based approaches utilizing time-frequency features for efficient heart sound classification. Ge et al. [19] extracted time-frequency features from cardiac cycles and employed a pretrained model for prediction, achieving a classification accuracy of 88.61%. Nia and Hesar [20] conducted heart sound segmentation followed by time-frequency feature extraction. Additionally, they utilized feature selection methods based on swarm optimization and sequential forward approach. Their machine learning approach achieved an accuracy of 98.03%, a sensitivity of 97.64%, and a specificity of 98.43%. Chen et al. [21] extracted time-frequency-based features using mel spectrum and log-mel spectrum from segmented heart sound signals, subsequently predicting heart sounds using a CNN model. It is evident from the literature that despite the implementation of a pretrained CNN model based on time-frequency features, the previous method requires significant improvement in terms of classification performance without increasing complexity.

Ismail [22] used spectrogram images to predict PCG signals employing chirplet Z-transform and CNN model. In a separate study, Jamil and Roy [23] used 1D and 2D HS features based on swarm optimization and genetic algorithm. They compute the classification performance using vision transformer. Zang [24] predict the HS based on 1D CNN model using an attentional mutiscale temporal network. They employed augmentation technique and deep learning approach to improve the classification model. Riccio [25] used 2D PCG images employing the partitioned iterated function system without preprocessing to classify the HS signal.

Conventionally, time-frequency-based approaches like STFT [26, 27] or techniques like MFCC [9, 10] are used to extract HS features. We deviate from the standard approach in this work by using gammatone

[3]Cardiovascular diseases (2021). Website https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) [accessed 2023 Sep 24].

features, taking cues from the human cochlea, and presenting them as gammatonegram images. Although mel frequency filters serve as the foundation for MFCC, our approach improves human auditory response modeling by integrating gammatone cepstral coefficients (GTCC) via gammatone filter (GF) bank implementation. The human ear is more sensitive to lower frequencies than to higher frequencies, so using fixed-length windows is not feasible. Based on empirical evidence presented in [28], GTCC performs 2.5% better than MFCC under a range of conditions.

1.2. Motivation and contribution

The PhysioNet 2016 dataset is characterized by imbalance and noise, presenting a challenge in accurately diagnosing heart conditions. The skewed distribution of data poses difficulties for most supervised classification algorithms, leading to suboptimal performance. The obstacles encountered by researchers in analyzing imbalanced datasets are outlined as follows:

- Majority class bias: Imbalanced datasets may introduce bias towards the majority class in classifiers.
- Limited minority class data: The insufficient size of the minority class can hinder a classifier's ability to generalize effectively.
- Evaluation metric misguidance: Traditional metrics such as accuracy may offer misleading insights in the context of imbalanced datasets.
- Bias selection: Sampling from unbalanced datasets introduces a skewed representation that could negatively affect the model's accuracy and gives rise to the possibility of selection bias.
- Overfitting risk: Overfitting occurs when a model performs well enough to categorize the training set but performs poorly when applied to the test set. This risk is introduced by imbalanced datasets.

S1 and S2, two vital HS that provide vital diagnostic information, are included in the audio recorded by a phonocardiogram. Alongside these vital signals, various noise sources and interference elements, such as ambient noise, muscle artifacts, and sensor-related anomalies, are present. Incorporating these additional variables may change the underlying cardiac signals, which makes it difficult to analyze the recorded PCG data appropriately. Using an effective band-pass filter denoising technique to identify and remove unnecessary noise elements from PCG recordings is the main challenge. Enhancing the accuracy and perceptual clarity of the HS encoded in the PCG signal is the main goal of this denoising technique. Thus, the goal of this innovation is to make it easier for qualified physicians to accurately diagnose and thoroughly analyze heart ailments.

The first HS usually has frequency components in the 25–150 Hz range, while the second HS usually has frequency components in the 150–250 Hz range. The automatic classification of these HS is complicated by the existence of both high-frequency and low-frequency disturbances. High-frequency external noises stem from various sources, including background noise, while low-frequency noise primarily consists of heart sounds in the 0–15 Hz range. The main sources of external noises are concurrent device activities, computer systems, LED lighting, AC duct interference, and electrical fans. The following is an outline of how auscultation sounds are represented acoustically:

$$X(t) = X_1(t) + X_2(t) + M_L(t) + M_H(t) \tag{1}$$

$H(t)$ is symbolically represented by equation (1) as a combination of the first HS ($X_1(t)$), the second HS ($X_2(t)$), the component linked to low-frequency noise ($M_L(t)$), and the part related to high-frequency

noise ($M_H(t)$). These components are classified automatically using a two-step procedure. The raw HS is first subjected to denoising, which removes unnecessary signals. The remaining signals are then classified into pathological and normal HS categories. For a computerized PCG classification to work, the following crucial steps must be taken: 1) preprocessing, 2) feature extraction, 3) augmentation, 4)ensemble model for PCG classification.

This study presents a novel deep learning CNN model, enhanced by effective preprocessing techniques utilizing gammatonegram images and ensemble-based transfer learning. These advancements aim to address challenges encountered in prior machine and deep learning investigations. Subsequent sections provide a comprehensive elucidation of each constituent depicted in the schematic diagram illustrated in Figure 1.

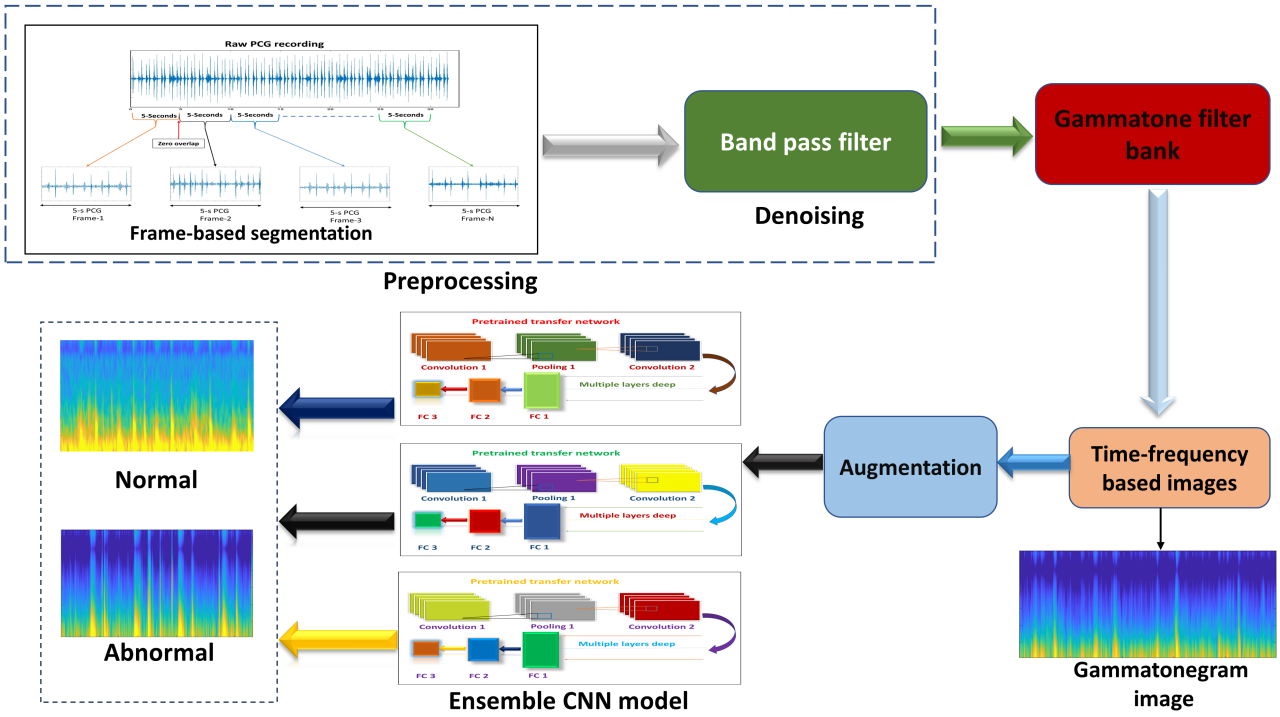


Figure 1. Schematic representation of the ensemble model proposed in this study

2. Materials and methods

2.1. Dataset

Numerous researchers have demonstrated the potential for assisting cardiologists and clinicians in accurately diagnosing cardiac anomalies with minimal effort, leveraging the PhysioNet 2016 challenge PCG dataset. In this study, we employed the PhysioNet 2016 PCG database [40] to assess cardiac anomaly conditions, highlighting the significant role of computational methods in enhancing diagnostic accuracy and clinical decision-making [6, 7]. The database contains 3240 HS in.wav files, which are divided into six separate databases labeled 'A' through 'F'. The HS recordings have lengths varying from 5 to 120 s and are sampled at 2000 Hz. The recordings are automatically categorized by the dataset into two primary groups: "Normal" and "Abnormal." The normal PCG recordings were sourced from individuals diagnosed with heart-related conditions, whereas the

healthy recordings originated from participants exhibiting optimal health. Among the patient cohort, conditions encompassed coronary artery disease and various cardiovascular ailments, including but not limited to heart valve defects such as aortic valve stenosis, valvular stenosis, mitral valve prolapse, mitral valve regurgitation, and aortic valve stenosis. A skilled listener manually annotated each recording, providing the dataset with accurate and extensive annotations for each HS segment, guaranteeing the highest degree of accuracy and thoroughness. Table 1 illustrated the PCG datasets distribution before preprocessing and after data augmentation.

Table 1. Distribution of PCG datasets before preprocessing and after data augmentation.

PhysioNet HS data				**HS after frame-based segmentation**				**HS after data augmentation**			
Dataset name	PCG		Total	Dataset name	PCG		Total	Dataset name	PCG		Total
	Abnormal	Normal			Abnormal	Normal			Abnormal	Normal	
A	292	117	409	A	1852	738	2590	A	1852	1852	3704
B	104	386	490	B	104	386	490	B	386	386	772
C	24	7	31	C	240	51	291	C	240	240	480
D	28	27	55	D	87	51	138	D	87	87	174
E	183	1958	2141	E	665	8129	8794	E	8129	8129	16,258
F	34	80	141	F	210	502	712	F	502	502	1004
Total	665	2575	3240	Total	3158	9857	13,015	Total	11,264	11264	22,528

2.2. Preprocessing

2.2.1. Frame-based segmentation

The PhysioNet dataset consist of imbalanced PCGs, with nearly four times as many instances of healthy heart sounds compared to abnormal HS. As outlined in the reference [29], the intrinsic class imbalance presents a substantial hurdle to classification performance, as models frequently manifest a predisposition towards the majority class. Consequently, it is crucial to address the risk of classifying all instances into the majority class while neglecting the minority class. To overcome the above problem, we employed a frame-based technique, as depicted in the figure 2, where we partitioned the HS using a 1000-sample window (5-second window) with no overlap.

2.2.2. Denoising

As indicated in the literature [29], we applied a 4th-order bandpass filter (Butterworth) in the 25–400 Hz range to successfully reduce noise in cardiac sound data. The application of the Butterworth filter allows for a substantial reduction in unwanted noise components within HS signals. Employing this filter enhances the quality and reliability of the signal, consequently improving subsequent operations related to feature extraction and classification. The selection of a bandpass range from 25 to 400 Hz ensures a precise emphasis on the critical frequency spectrum linked to HS signals. This approach effectively preserves essential signal components while minimizing noise interference.

2.3. Feature extraction/gammatonegram image generation

The transformation of HS signal into an image format is crucial for the established transfer learning method. This was accomplished by converting HS signals into gammatonegram images using GF bank procedures. The transformation of a PCG image from an HS signal is shown in Figure (3), which was achieved by using GF bank methods.

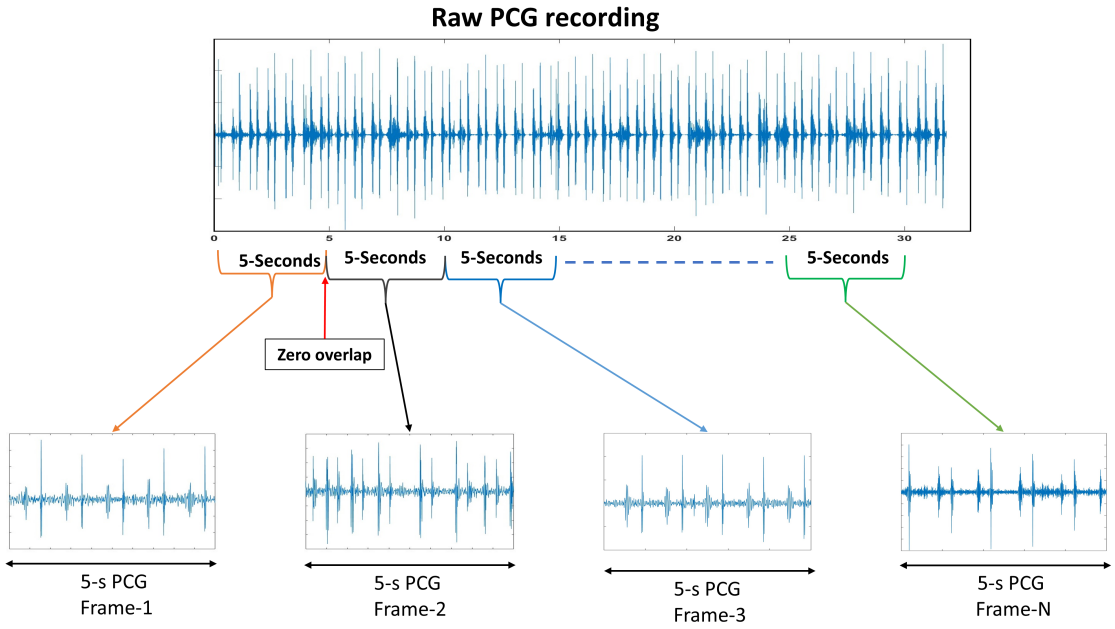


Figure 2. Partitioning HS recordings into frames for analytical processing.

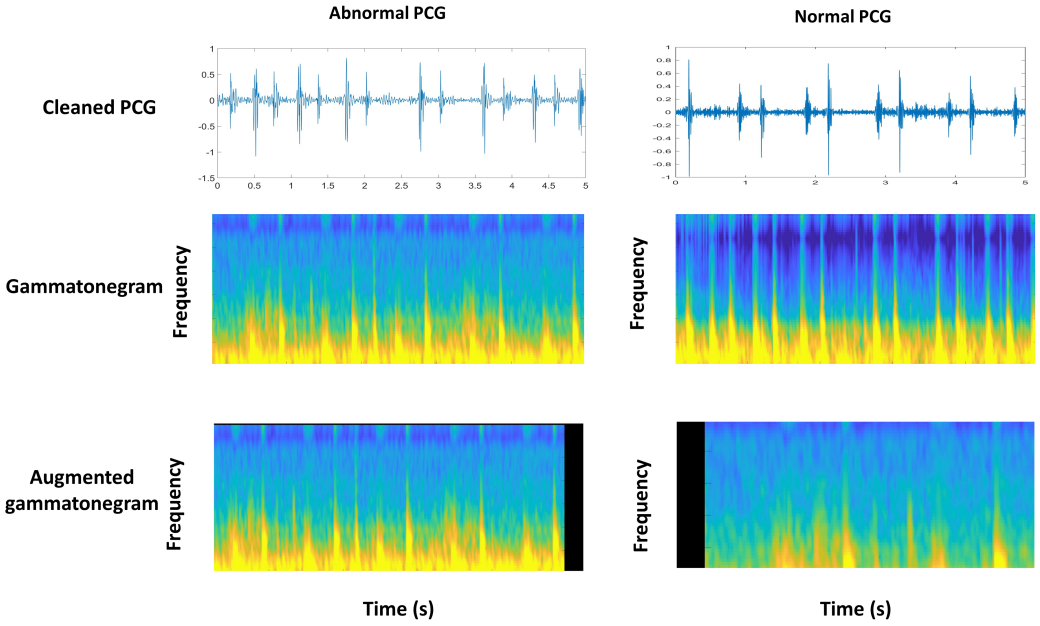


Figure 3. Gammatonegram images along with augmented images generated using GF bank.

It is necessary to use time-frequency domain techniques to analyze the nonstationary characteristics of HS signals. In comparison to traditional spectrograms, GF banks which are intended to mimic the aural properties of the human ear produce gammatonegrams from the filter responses. The basilar membrane, which is located in the cochlea of the human auditory system, is essential for translating incoming sound frequencies into matching

vibrational responses and energy patterns. $G_m(t)$, the impulse response of GF, is determined by Equation (2). By convolutioning a sinusoidal tone with a frequency centering at F_c with a gamma distribution, this response is obtained [28]. These finely tuned filters provide an audio representation called a Gammatonegram, which is similar to a spectrogram, by mimicking the vibrational responses present in the human ear.

Broader bandwidths (BW) at higher frequencies and narrower BW at lower frequencies represent the characteristics of gammatonegram bandpass filters, which set them apart from the STFT. These filters' impulse response, represented by $G_m(t)$ and shown in equation (2), provides additional insight into their architecture. These finely tuned filters mimic the vibrational reactions that take place in the human ear, creating an audio representation that is similar to a spectrogram, or gammatonegram. A characteristic that sets gammatonegram bandpass filters apart from the STFT is their increasingly wider BW at higher frequencies and narrower bandwidths at lower ones.

$$G_m(t) = Pt^{(b-1)}e^{-2\pi dt}cos(2\pi F_c t + \phi) \qquad t > 0 \tag{2}$$

Additional parameters further define the filter's properties: P stands for the amplitude factor, d for the filter bandwidth, b for the filter order, and ϕ for the phase shift. It is noteworthy that the decay factor controls the filter's bandwidth and impulse response, and that the distribution of F_c follows the Equivalent Rectangular Bandwidth scale, as shown in equation (3).

$$ERB(F_c) = 24.7 + 0.108F_c \tag{3}$$

By superimposing the response of each time frame, $X_d(t)$, on the entire GF bank, $G_m(t)$, the gammatonegram spectrogram is created. Equation (4) describes the implementation of this technique over overlapping time frames.

$$F_m = S_d(t) * G_m(t) \tag{4}$$

2.4. Augmentation

Adequately balancing the dataset used for training machine learning models can significantly improve their performance and accuracy. Augmenting the data has the potential to further enhance the effectiveness of these models [30]. Data augmentation comprises a collection of methods aimed at artificially expanding the dataset by generating additional data points from existing ones, thereby addressing class imbalance. A straightforward technique for generating synthetic data involves making minor alterations, such as flipping, transformations, or rotations, particularly applicable in the case of image data [31].

The limited quantity of samples within the dataset presents issues with overfitting when deep neural networks are being trained. By using approaches similar to those used in audio augmentation, data augmentation—a technique that expands the dataset—offers an achievable solution to the overfitting issue. The method of augmentation makes it easier to create synthetic data from the current collection. Although there are many different ways to amplify audio signals, including stretching, noise injection, background distortion, time-shifting, pitch alteration, and speed adjustment, this study only used two of these augmentation techniques, which are described below:

i. Time-shifting: Time-shifting, when applied to audio files, is the displacement of an image in both the horizontal and vertical directions by a randomly chosen distance between [-100, 100] pixels.

ii. Data stretching: In order to achieve the desired effect, this augmentation method mainly modifies the temporal duration of the signal, allowing compression or the expansion while maintaining pitch. Data

stretching has several uses in the field of HS analysis, such as simulating slower heartbeats and expanding the amount of the dataset. The goal of producing data that closely resembles the original, slower heartbeat PCG signals is what drives the stretching of PCG signals. By enriching the dataset with a variety of PCG signal changes, this augmentation approach effectively grows the dataset and produces a deep learning model that is more resilient and has better generalization. As part of the data stretching method in this study, the original heart sound signals are stretched along the temporal axis by a factor of 0.5.

2.5. Transfer learning models

To address specific challenges in training models with domain-specific datasets, we employed an efficient transfer learning method in this work. This approach utilizes CNN models pretrained on the ImageNet dataset. The database has been used for pretraining a number of CNN-based models that are commonly used in the field of transfer learning, such as AlexNet [32], GoogLeNet [33], SqueezeNet [34], and VGG19 [35]. Figure 4 represents the transfer learning model based on deep learning algorithm.

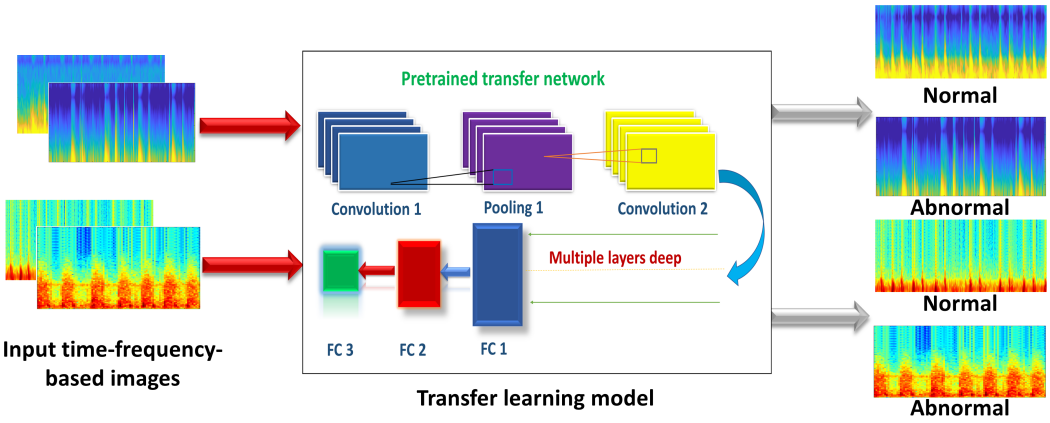


Figure 4. PCG classification mechanism using transfer learning approach.

2.6. Ensemble model

In real-world applications characterized by nonuniform distributions of the target variable, imbalanced data is a common occurrence. In such scenarios, the minority class often carries greater significance than the dominant class, and inaccurate classification of the minority class can lead to substantial consequences. By reducing the influence of the dominant class on the model and increasing the focus on the minority class, ensemble models offer a useful approach to overcoming this difficulty. Figure 5 represents the block diagram of the proposed ensemble model.

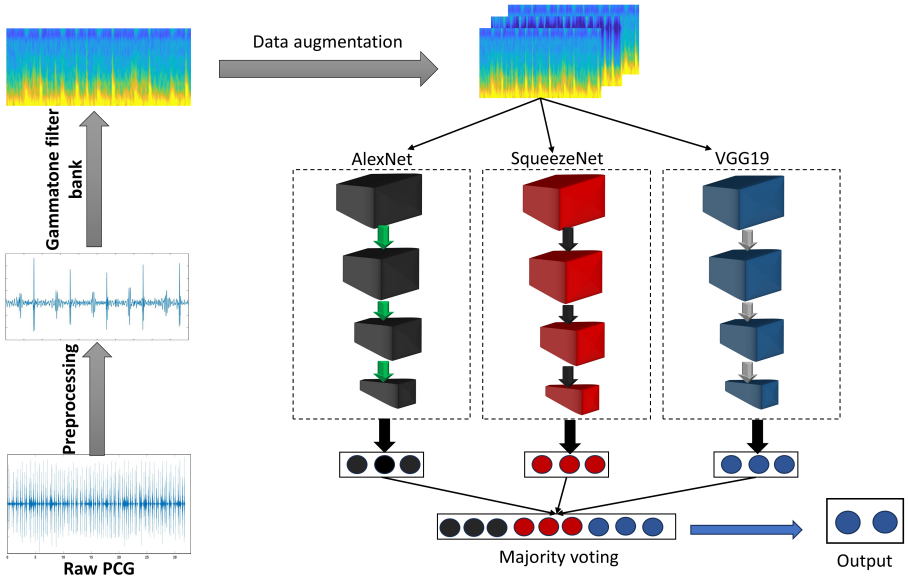


Figure 5. The block diagram of the proposed ensemble model.

Ensemble models have gained widespread acceptance as a popular and effective technique to improve the efficiency of the deep learning models, especially CNN-based models [32–35]. This paper analyzes the application of ensemble models, which include SqueezeNet, AlexNet, and VGG19, for gammatonegram images-based unbalanced HS classification and explains how they work. Our proposed model combines the results of several CNN models to improve classification performance overall. The identical dataset is employed to train each of these individual models, with varying initial weights applied. The ensemble model attains superior accuracy compared to any single model when predictions from multiple models are amalgamated. The basic concept underlying ensemble models is to maximize each individual model's strengths while minimizing its shortcomings. The principal objective is to reduce the variance and bias that are present in individual models, which will lead to increased accuracy, generalization, and robustness in the end.

CNNs, such as VGG19, AlexNet, and SqueezeNet, are commonly used in image classification, each with inherent constraints influencing their performance. SqueezeNet, prioritizing memory efficiency, suits embedded systems but may struggle with complex datasets. Despite AlexNet's proven success, its shallow architecture may limit accurate representation, particularly with smaller datasets, raising the risk of overfitting. VGG19, with its deep 19-layer structure, excels in capturing intricate data but demands more computational resources and time for training. To overcome individual limitations, we adopted an ensemble approach, combining the strengths of SqueezeNet, AlexNet, and VGG19. This integration yields a robust, accurate model for hyperspectral classification, mitigating the drawbacks of individual models.

Through the successive training of weak models and the amalgamation of their estimations, boosting methods improve the predictive power of an ensemble model. In this work, we used the boosting approach described in Algorithm 1 to create an ensemble model. To find incorrectly classified samples, the boosting approach first trains a weak model and then assesses its efficacy on the training set. The training data is then updated to give these incorrectly categorized samples larger weights. The updated data is then used to train a new weak model, which focuses on the samples that were previously incorrectly classified. Iteratively repeating steps 2-4 proceeds until a predefined stopping threshold is reached.

Algorithm 1 Boosting

Input: Data to be trained (a, b)**, Weak Model (**M_w**), Sample (**D**)**
Result: Ensemble Model (E**)**
Start Procedure:

$l = p$ models
for $l \geq 1$: $p \rightarrow$ The maximum number of weak models acts as the stopping criterion
$X = Train(M_w)$
$Y = Compute(X)$
$Z = Identify(D, Y)$
$C = Update(a, b, Z)$
$X_x = TrainWeak(C)$
end for
$F = Concatenate(X_1, X_2, ..., X_n)$
$E = Predict(F)$

[4]The 'Train' function is responsible for training the weak model and undergoes evaluation on 'X' using the 'Compute' function to designate the weak model as 'Z'. Next, based on samples that were incorrectly identified, the training data is updated, and the 'TrainWeak' function is used to train the weak models.

3. Evaluation metrics

3.1. Model training

To tackle the issue of PCG classification, our approach employed transfer learning with pretrained models (AlexNet, SqueezeNet, and VGG19). This approach expedited task adaption while simultaneously reducing the total amount of time needed for learning and training. CNN models initially emerged for the analysis of images, but by using transfer learning, they have shown to be flexible enough to be used in a variety of applications, including HS classification.

To utilize our deep-learning model, we incorporated 2D gammatonegram images. We were able to efficiently manage the intricacies involved in HS analysis and get the data ready for model training by using the GF bank technique. The input images were scaled to fit the network's input specifications before the training process began.

3.2. Performance parameter

Accuracy, sensitivity, and specificity were computed in order to evaluate the effectiveness of the proposed method; the relevant mathematical equations for each of these parameters are given in equations (5), (6), (7), respectively.

$$Accuracy = \frac{(T\ P + T\ N)}{(T\ P + T\ N + F\ P + F\ N)} \tag{5}$$

$$Sensitivity = \frac{(T\ P)}{(T\ P + F\ N)} \tag{6}$$

$$Specificity = \frac{(T\ N)}{(T\ N + F\ P)} \tag{7}$$

Here TP, TN, FP, and FN denotes true positive, true negative, false positive, and false negative, respectively.

4. Results

This section reveals the experimental data that were acquired using our proposed methods, with careful performance comparisons that highlight the results. The evaluation leverages the PhysioNet PCG datasets to assess the effectiveness of the proposed ensemble model in classifying heart sound signals. To accommodate testing and training, the dataset has been split into 70:30 ratios. The training dataset consists of 15,770 audio recordings, equally divided into 7885 instances of normal HS and 7885 instances of abnormal HS.

Four popular CNNs—AlexNet, SqueezeNet, GoogleNet, and VGG19—are used as base models for transfer learning in the proposed model. A new fully connected layer updates the preceding layer while keeping the weights of the training models constant in order to identify HS as normal or abnormal. Training with two distinct learning rates, $3e^{-4}$ and $3e^{-3}$, is used to assess classification performance. The gammatonegram images are divided into groups at random, with 30% of the data for test validation and 70% for learning the models (SqueezeNet, AlexNet, GoogleNet, and VGG19). Conversely, the testing dataset comprised 6758 images, with an equal distribution of 3379 normal and 3379 abnormal instances. To assess the efficacy and robustness of the proposed model, a rigorous validation procedure employing 10-fold cross-validation methodology was used. This approach ensures comprehensive evaluation across diverse subsets of the dataset, enhancing confidence in the model's performance and generalization capabilities. The model underwent training with a maximum of 10 epochs and the Adam optimizer with momentum, with the mini-batch size set to 10. A mini-batch of 1577 observations is utilized in each cycle.

Table 2. Performance comparison of different models before and after data augmentation with learning rates of 0.0001 and 0.001.

Model	**Performance before data augmentation**			**Performance after data augmentation**		
	Learning rate = 0.0001					
	Accuracy %	**Sensitivity %**	**Specificity %**	**Accuracy %**	**Sensitivity %**	**Specificity %**
AlexNet	96.59	95.04	97.38	98.15	98.28	98.01
SqueezeNet	95.97	93.66	97.01	98.44	98.66	98.22
VGG19	93.31	90.28	94.28	95.30	95.23	95.38
GoogleNet	94.41	92.39	95.06	96.68	96.62	96.15
ResNet50	94.33	95.2	94.04	96.92	96.19	97.15
DenseNet	90.47	90.49	90.46	95.31	92.50	96.21
Ensemble 1	97.38	95.35	98.10	99.51	99.34	99.67
Model	**Performance before data augmentation**			**Performance after data augmentation**		
	Learning rate = 0.001					
	Accuracy %	**Sensitivity %**	**Specificity %**	**Accuracy %**	**Sensitivity %**	**Specificity %**
AlexNet	95.15	93.24	95.77	97.94	98.01	97.81
SqueezeNet	96.05	94.29	96.58	97.54	97.57	96.80
VGG19	93.11	90.07	94.08	94.54	94.76	94.32
GoogleNet	94.13	91.97	94.82	95.17	95.56	94.79
ResNet50	91.80	94.19	91.03	94.08	95.56	94.08
DenseNet	88.31	90.92	87.48	93.04	91.34	94.62
Ensemble 2	97.38	95.35	98.03	99.12	99.29	98.96

Table 3. Performance evaluation of base models and proposed ensemble model.

Model architecture	Learning rate	Test set 6758 PCG	Predicted		Performance %		
			Abnormal	Normal	Acc	Sen	Spec
AlexNet	0.0001	Abnormal	3321	58	98.15	98.28	98.01
		Normal	67	3312			
SqueezeNet		Abnormal	3334	45	98.44	98.66	98.22
		Normal	60	3319			
GoogleNet		Abnormal	3285	114	96.68	96.62	96.15
		Normal	130	3249			
VGG19		Abnormal	3218	161	95.30	95.23	95.38
		Normal	156	3223			
AlexNet	0.001	Abnormal	3314	65	97.94	98.01	97.81
		Normal	74	3305			
SqueezeNet		Abnormal	3321	82	97.54	97.57	96.80
		Normal	108	3271			
GoogleNet		Abnormal	3229	150	95.17	95.56	94.79
		Normal	176	3203			
VGG19		Abnormal	3202	177	94.54	94.76	94.32
		Normal	192	3187			
Proposed Model 1	0.0001	Abnormal	3366	13	99.51	99.34	99.67
		Normal	11	3368			
Proposed Model 2	0.001	Abnromal	3355	24	99.12	99.29	98.96
		Normal	35	3344			

Acc: Accuracy, Sen: Sensitivity, Spec: Specificity

We conducted a comparative analysis of performance evaluation before and after data augmentation to assess its efficacy with a learning rate of 0.0001 and 0.001, as demonstrated in Table 2. The learning rates that yielded the highest classification performance levels were selected to evaluate the efficacy of the augmentation technique. Specifically, the increases in accuracy percentages after augmentation were observed as follows: 1.56% and 2.79% for AlexNet with learning rates of 0.0001 and 0.001, respectively; 2.47% and 1.49% for SqueezeNet with learning rates of 0.0001 and 0.001, respectively; 1.99% and 1.43% for VGG19 with learning rates of 0.0001 and 0.001, respectively; and 2.27% and 1.04% for GoogleNet with learning rates of 0.0001 and 0.001, respectively; 4.84% and 4.73% for DenseNet with learning rates of 0.0001 and 0.001, respectively; lastly, 2.59% and 2.28% for ResNet50 with learning rates of 0.0001 and 0.001, respectively. In Table 2, we incorporated additional traditional CNN models such as GoogleNet, ResNet50, and DenseNet. Despite the potential for these models to surpass our base models (AlexNet, SqueezeNet, VGG19) in terms of classification performance, they were excluded from our proposed ensemble model. This decision was based on their multilayered architecture, which has the potential to elevate the computational complexity of the classification system.

A detailed summary of the experimental results for the proposed models is provided in Table 3. This summary takes into account two distinct learning rates (0.001 and 0.0001) over 10 epochs. Each of the four base models has been thoroughly evaluated with respect to performance metrics including accuracy, specificity, and sensitivity. The network was trained using the momentum-based Adam optimizer with a 10-epoch mini-batch size and a 10-epoch training termination condition. A benchmark database has been made available by PhysioNet to help with PCG classification algorithm validation. Figures 6 and 7 show a comparative analysis of the basis models using gammatonegram and spectrogram images at learning rates of 0.0001 and 0.001, respectively.

An optimal outcome on the testing dataset was achieved by effectively employing a learning rate of 0.0001. The SqueezeNet classifier emerged as the top-performing model, showcasing an exceptional validation accuracy of

98.44% when applied to gammatonegram images. Furthermore, the model demonstrated a sensitivity of 98.66% and specificity of 98.22%, affirming its robust and effective capabilities in the realm of image categorization tasks.

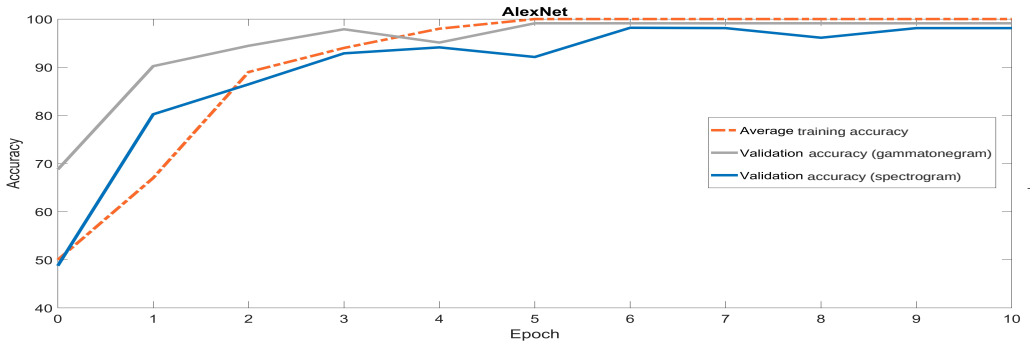
We propose an ensemble approach, integrating predictions from AlexNet, SqueezeNet, and VGG19, intentionally excluding GoogleNet to prevent increased computational complexity. Using learning rates of 0.003 and 0.0003 for Proposed Models 1 and 2, each base model undergoes independent training. Following Algorithm 1, ultimate predictions are obtained by a weighted average of individual model predictions. Our methodology involves model selection and fine-tuning to enhance overall classification.

The data depicted in Table 3 reveals a convergence in performance across all models. However, it is worth noting that the SqueezeNet model exhibits inferior performance compared to the other base models when utilizing a learning rate of 0.0001. Conversely, the AlexNet model demonstrates superior classification performance when employing a learning rate of 0.001. Table 3 presents confusion matrices for each base model and the ensemble-based transfer model. For instance, in proposed Model 2, predictions encompass 3366 abnormal data out of 3379 and 3368 out of 3379 normal data. Results indicate improved classification with the ensemble model, addressing the influence of the predominant class and emphasizing the minority class. The ensemble proposed model, utilizing a learning rate of 0.0001, attains exceptional classification performance with a 99.51% accuracy, 99.34% sensitivity, and 99.67% specificity. Our findings underscore the efficacy of ensemble models in handling imbalanced datasets, with implications for machine learning and data analytics.

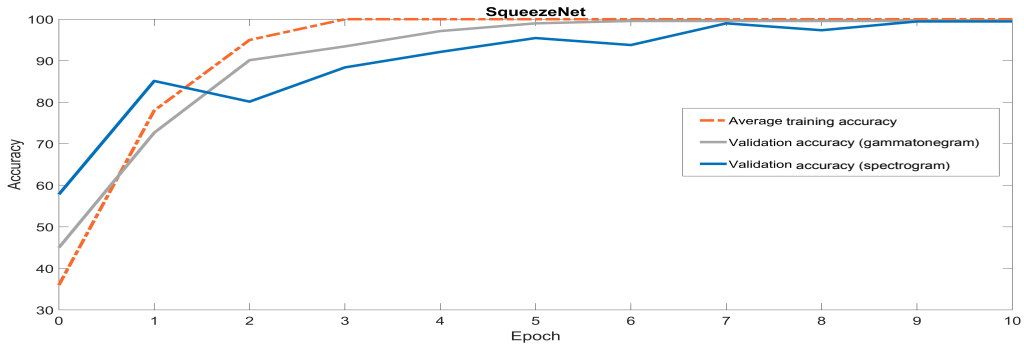
5. Discussion

This study introduces a novel approach utilizing gammatone filters depicted in images for adept feature extraction, aiming to efficiently classify heart sounds into two distinct categories. The primary goal is to enhance the accuracy and effectiveness of methods applied in processing biomedical data. To address limitations like a small dataset and class imbalance, two solutions are implemented: data augmentation pretraining and the introduction of a dropout layer with a rate of 0.5 during training. The study innovatively employs a biologically derived GF bank for generating gammatone images, mimicking the mechanics of the human ear's cochlea to improve HS categorization efficacy. Table 4 presents a comprehensive comparative analysis of our proposed ensemble model, the ensemble model using spectrogram images, and conventional transfer learning methods. Gammatonegrams and spectrograms are utilized in Figures 6 and 7, illustrating the training and validation accuracy of base models with learning rates of 0.0001 and 0.001. Remarkably, the SqueezeNet model achieved the highest classification accuracy among the considered models.

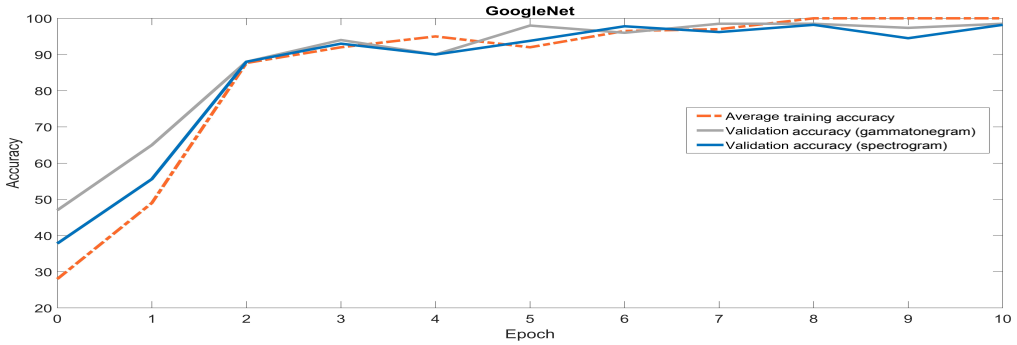
Our research validates the superior classification accuracy of our proposed ensemble model compared to traditional approaches, specifically outperforming conventional transfer learning models. The integration of AlexNet, SqueezeNet, and VGG19 as base models played a pivotal role, resulting in an impressive 99.51% classification accuracy. Future CNN research should consider adopting this ensemble approach to enhance accuracy and address limitations of conventional transfer learning. Table 5 comprehensively compares our approach, utilizing the PhysioNet database, with the existing literature. Unlike prior studies emphasizing accuracy alone, our model prioritizes a balanced sensitivity and specificity, crucial for minimizing the costs of misclassification. Our ensemble strategy aims to overcome unbalanced dataset challenges noted in previous studies, demonstrating superior performance in predicting cardiac abnormalities. The model's feature extraction capabilities can be further enhanced by integrating diverse model designs, ensuring robustness against overfitting, noise, and data variations. Our approach achieved outstanding results, boasting a 99.51% accuracy, alongside an impressive 99.34% sensitivity and 99.67% specificity.



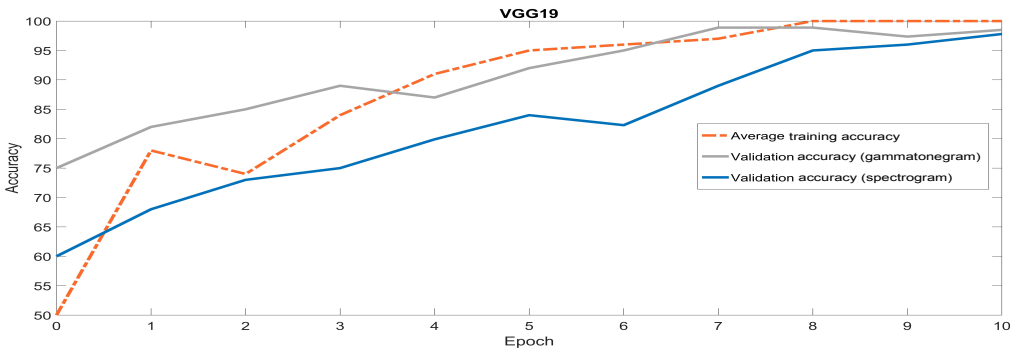
(a) Assessment of classification accuracy utilizing AlexNet.



(b) Assessment of classification accuracy utilizing SqueezeNet

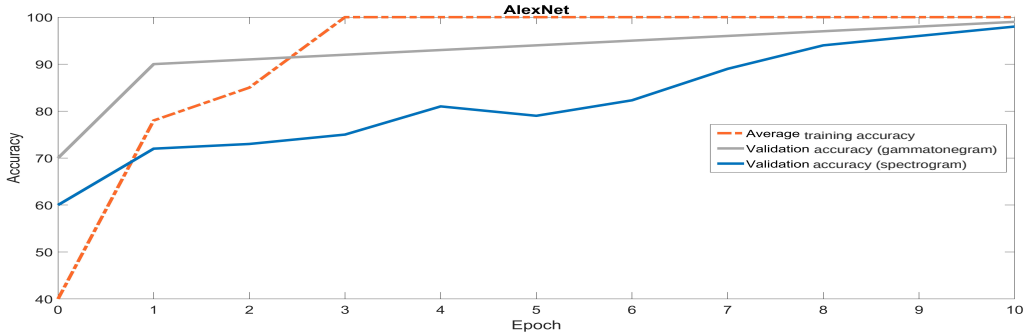


(c) Assessment of classification accuracy utilizing GoogleNet

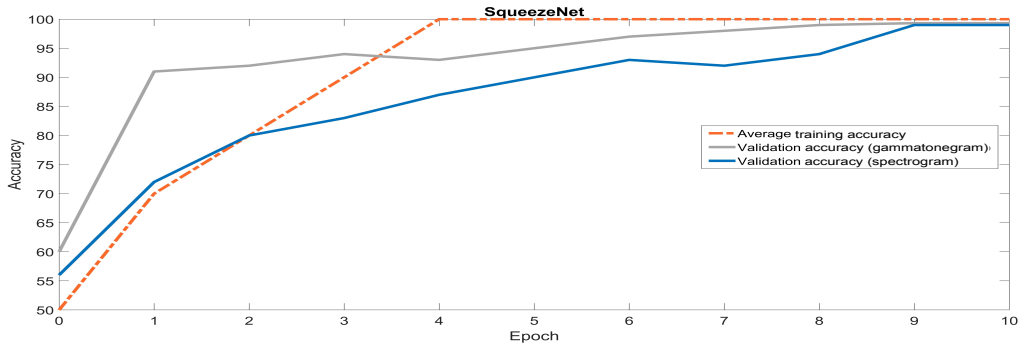


(d) Assessment of Classification accuracy utilizing VGG19

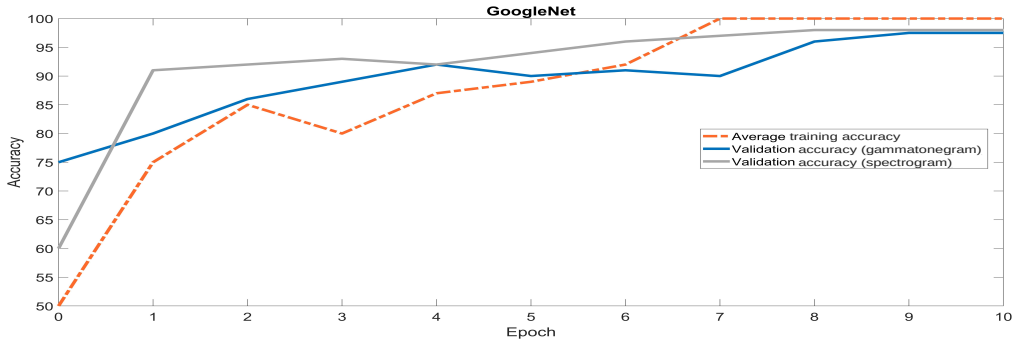
Figure 6. Evaluating the classification accuracy of the base models using gammatonegram and spectrogram images, employing a learning rate of $1e^{-4}$.



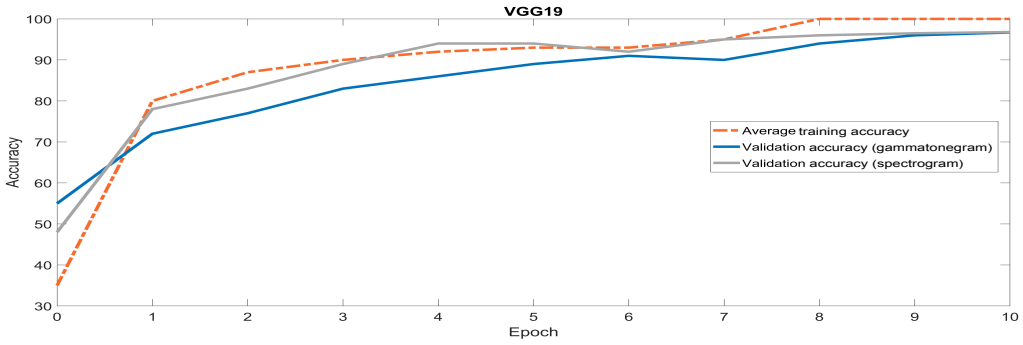
(a) Assessment of classification accuracy utilizing AlexNet



(b) Assessment of classification accuracy utilizing SqueezeNet



(c) Assessment of classification accuracy utilizing GoogleNet



(d) Assessment of classification accuracy utilizing VGG19

Figure 7. Evaluating the classification accuracy of the base models using gammatonegram and spectrogram images, employing a learning rate of $1e^{-3}$.

Table 4. Comparative performance evaluation of the proposed ensemble model with traditional models.

Model	Feature	Learning rate	Acc %	Sen %	Spec %
AlexNet	Gammatonegram images	0.0001	98.15	98.28	98.01
SqueezeNet			98.44	98.66	98.22
GoogleNet			96.68	96.62	96.15
VGG19			95.30	95.23	95.38
AlexNet		0.001	97.94	98.01	97.81
SqueezeNet			97.54	97.57	96.80
GoogleNet			95.17	95.56	94.79
VGG19			94.54	94.76	94.32
AlexNet	Spectrogram images	0.0001	97.69	98.10	97.28
SqueezeNet			97.82	98.25	97.39
GoogleNet			95.64	95.77	95.53
VGG19			96.37	96.68	96.06
AlexNet		0.001	96.92	97.86	95.97
SqueezeNet			95.90	96.03	95.76
GoogleNet			94.83	94.93	94.73
VGG19			94.46	95.02	**94.19**
Ensemble 1	Spectrogram images	0.0001	99.27	99.29	99.26
Ensemble 2	Spectrogram images	0.001	98.84	98.87	98.81
Proposed 1	Gammatonegram images	0.0001	**99.51**	**99.34**	**99.67**
Proposed 2	Gammatonegram images	0.0001	99.12	99.29	98.96

Table 5. Comparative evaluation of our proposed approach with state-of-the-arts method employing PhysioNet 2016 database.

Ref	Classification model	Method/features employed	Performance %		
			Accuracy	Sensitivty	Specificity
[9] 2018	CNN	MFCC and Mel-Spectrogram	81.50	84.50	78.50
[38] 2019	LSTM	Spectrogram Images	94.66	96.15	93.18
[42] 2020	Neural network	Multidomain	97.89	97.73	98.05
[40] 2020	CNN	Time-frequency	94.00	95.00	93.00
[10] 2021	Ensemble	Multidomain	92.47	94.08	91.95
[39] 2021	Gradient boosting	Multidomain	95.23	92.00	98.45
[41] 2021	SVM	Time-frequency	86.00	**93.31**	**78.59**
[8] 2022	ResNet	Spectrogram images	85.08	-	-
[12] 2022	KNN	Wavelet scattering	97.82	95.04	98.72
[37] 2023	YAMNet	Spectrogram images	92.23	**-**	**-**
Proposed Model 1	Ensemble	Gammatonegram Images	**99.51**	**99.34**	**99.67**
Proposed Model 2			99.12	99.29	98.96

6. Conclusion and limitations

Our work provides a strong example of how an ensemble method can be applied to effectively anticipate imbalanced HS signals from gammatonegram images. We have obtained a state-of-the-art performance benchmark on the PhysioNet dataset by putting the proposed approach into practice, demonstrating its importance in real-world clinical situations. The ensemble model that we present in this work provides an efficient way to deal with imbalanced data in automatic learning tasks by employing pretrained SqueezeNet, AlexNet, and VGG19. Selecting an effective ensemble model type, using data augmentation, and carefully allocating weights to various classes are all essential for maximizing performance when dealing with unbalanced data. Future efforts to further progress in this field of study can include extending our approach to other datasets and exploring CNN frameworks.

There are certain limitations associated with the proposed approach. Firstly, the inclusion of additional models may augment the complexity and computational overhead of the ensemble model, posing challenges, especially when confronted with extensive data. Furthermore, the ensemble model's efficacy depends on how well each of its basic models performs on a standalone basis. Using a wide range of models could lead to differences in performance, which could reduce the ensemble model's reliability. Lastly, a deep network founded on a deep learning algorithm demands extended training periods; however, once effectively trained, the model exhibits efficient inference capabilities.

Conflict of interests

There are no conflicts of interest in this study.

Acknowledgment

This paper is the outcome of research and development carried out under the rules and regulations of the Visvesvaraya PhD Scheme, which is managed by Digital India Corporation on behalf of the Government of India and regulated by the Ministry of Electronics and Information Technology (MeitY).

References

[1] Etchells E, Bell C, Robb K. Does this patient have an abnormal systolic murmur? JAMA 1997;277 (7):564-71. https://doi.org/10.1001/jama.1997.03540310062036

[2] Mangione S, Nieman LZ. Cardiac auscultatory skills of internal medicine and family practice trainees: a comparison of diagnostic proficiency. JAMA 1997 3;278(9):717-22. https://doi.org/10.1001/jama.1997.03550090041030

[3] Lam MZ, Lee TJ, Boey PY, Ng WF, Hey HW et al. Factors influencing cardiac auscultation proficiency in physician trainees. Singapore Medical Journal. 2005;46 (1):11.

[4] Leng S, Tan RS, Chai KT, Wang C, Ghista D et al. The electronic stethoscope. Biomedical Engineering Online. 2015;14 (1):1-37. https://doi.org/10.1186/s12938-015-0056-y

[5] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation. 2000; 101 (23):e215-20. https://doi.org/10.1161/01.CIR.101.23.e215

[6] Clifford GD, Liu C, Moody B, Springer D, Silva I et al. Classification of normal/abnormal heart sound recordings: The PhysioNet/Computing in Cardiology Challenge 2016. In: 2016 Computing in Cardiology Conference (CinC) 2016; 609-612. IEEE.

[7] Liu C, Springer D, Li Q, Moody B, Juan RA et al. An open access database for the evaluation of heart sound algorithms. Physiological Measurement. 2016; 37 (12):2181. https://doi.org/10.1088/0967-3334/37/12/2181

[8] Azam FB, Ansari MI, Nuhash SI, McLane I, Hasan T. Cardiac anomaly detection considering an additive noise and convolutional distortion model of heart sound recordings. Artificial Intelligence in Medicine. 2022; 133:102417. https://doi.org/10.1016/j.artmed.2022.102417

[9] Bozkurt B, Germanakis I, Stylianou Y. A study of time-frequency features for CNN-based automatic heart sound classification for pathology detection. Computers in Biology and Medicine. 2018; 100:132-43. https://doi.org/10.1016/j.compbiomed.2018.06.026

[10] Singh SA, Majumder S. Short unsegmented PCG classification based on ensemble classifier. Turkish Journal of Electrical Engineering and Computer Sciences. 2020; 28 (2):875-89. https://doi.org/10.3906/elk-1905-165

[11] Kay E, Agarwal A. DropConnected neural networks trained on time-frequency and inter-beat features for classifying heart sounds. Physiological Measurement. 2017; 38 (8):1645. https://doi.org/10.1088/1361-6579/aa6a3d

[12] Ajitkumar Singh S, Dinita Devi N, Majumder S. An improved unsegmented phonocardiogram classification using nonlinear time scattering features. The Computer Journal. 2023; 66 (6):1525-40. https://doi.org/10.1093/comjnl/bxac025

[13] Whitaker BM, Suresha PB, Liu C, Clifford GD, Anderson DV. Combining sparse coding and time-domain features for heart sound classification. Physiological Measurement. 2017 ;38 (8):1701. https://doi.org/10.1088/1361-6579/aa7623

[14] Xu W, Yu K, Ye J, Li H, Chen J et al. Automatic pediatric congenital heart disease classification based on heart sound signal. Artificial Intelligence in Medicine. 2022 ; 126:102257. https://doi.org/10.1016/j.artmed.2022.102257

[15] Eslamizadeh G, Barati R. Heart murmur detection based on wavelet transformation and a synergy between artificial neural network and modified neighbor annealing methods. Artificial Intelligence in Medicine. 2017; 78:23-40. https://doi.org/10.1016/j.artmed.2017.05.005

[16] Maknickas V, Maknickas A. Recognition of normal–abnormal phonocardiographic signals using deep convolutional neural networks and mel-frequency spectral coefficients. Physiological Measurement. 2017; 38 (8):1671. https://doi.org/10.1088/1361-6579/aa7841

[17] Beritelli F, Capizzi G, Lo Sciuto G, Napoli C, Scaglione F. Automatic heart activity diagnosis based on Gram polynomials and probabilistic neural networks. Biomedical Engineering Letters. 2018: 77-85. https://doi.org/10.1007/s13534-017-0046-z

[18] Humayun AI, Ghaffarzadegan S, Ansari MI, Feng Z, Hasan T. Towards domain invariant heart sound abnormality detection using learnable filterbanks. IEEE Journal of Biomedical and Health Informatics. 2020; 24 (8):2189-98. https://doi.org/10.1109/JBHI.2020.2970252

[19] Ge B, Yang H, Ma P, Guo T, Pan J et al. Detection of pulmonary hypertension associated with congenital heart disease based on time-frequency domain and deep learning features. Biomedical Signal Processing and Control. 2023; 81:104316. https://doi.org/10.1016/j.bspc.2022.104316

[20] Nia PS, Hesar HD. Abnormal heart sound detection using time-frequency analysis and machine learning techniques. Biomedical Signal Processing and Control. 2024; 90:105899. https://doi.org/10.1016/j.bspc.2023.105899

[21] Chen W, Zhou Z, Bao J, Wang C, Chen H et al. Classifying heart-sound signals based on CNN trained on MelSpectrum and Log-MelSpectrum features. Bioengineering. 2023; 10 (6):645. https://doi.org/10.3390/bioengineering10060645

[22] Ismail S, Ismail B. PCG signal classification using a hybrid multi round transfer learning classifier. Biocybernetics and Biomedical Engineering. 2023; 43 (1):313-34. https://doi.org/10.1016/j.bbe.2023.01.004

[23] Jamil S, Roy AM. An efficient and robust phonocardiography (pcg)-based valvular heart diseases (vhd) detection framework using vision transformer (vit). Computers in Biology and Medicine. 2023; 158:106734. https://doi.org/10.1016/j.compbiomed.2023.106734

[24] Zang J, Lian C, Xu B, Zhang Z, Su Y et al. AmtNet: Attentional multi-scale temporal network for phonocardiogram signal classification. Biomedical Signal Processing and Control. 2023; 85:104934. https://doi.org/10.1016/j.bspc.2023.104934

[25] Riccio D, Brancati N, Sannino G, Verde L, Frucci M. CNN-based classification of phonocardiograms using fractal techniques. Biomedical Signal Processing and Control. 2023; 86:105186. https://doi.org/10.1016/j.bspc.2023.105186

[26] Shuvo SB, Alam SS, Ayman SU, Chakma A, Barua PD et al. NRC-Net: Automated noise robust cardio net for detecting valvular cardiac diseases using optimum transformation method with heart sound signals. Biomedical Signal Processing and Control. 2023; 86:105272. https://doi.org/10.1016/j.bspc.2023.105272

[27] Bao X, Xu Y, Lam HK, Trabelsi M, Chihi I et al. Time-frequency distributions of heart sound signals: a comparative study using convolutional neural networks. Biomedical Engineering Advances. 2023 Jun 1;5:100093. https://doi.org/10.1016/j.bea.2023.100093

[28] Valero X, Alias F. Gammatone cepstral coefficients: Biologically inspired features for non-speech audio classification. IEEE Transactions on Multimedia. 2012; 14 (6):1684-9. https://doi.org/10.1109/TMM.2012.2199972

[29] Schmidt SE, Toft E, Holst-Hansen C, Graff C, Struijk JJ. Segmentation of heart sound recordings from an electronic stethoscope by a duration dependent Hidden-Markov model. In: 2008 Computers in Cardiology 2008; 345-348. IEEE.

[30] Van Dyk DA, Meng XL. The art of data augmentation. Journal of Computational and Graphical Statistics. 2001; 10 (1):1-50. https://doi.org/10.1198/10618600152418584

[31] Khan AA, Chaudhari O, Chandra R. A review of ensemble learning and data augmentation models for class imbalanced problems: combination, implementation and evaluation. Expert Systems with Applications. 2023: 122778. https://doi.org/10.1016/j.eswa.2023.122778

[32] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems. 2012;25.

[33] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S et al. Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2015 (pp. 1-9).

[34] Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. arXiv preprint arXiv:1602.07360. 2016 Feb 24.

[35] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. 2014 Sep 4.

[36] Ghiasi M, Wang Z, Mehrandezh M, Jalilian S, Ghadimi N. Evolution of smart grids towards the Internet of energy: Concept and essential components for deep decarbonisation. IET Smart Grid. 2023;6 (1):86-102. https://doi.org/10.1049/stg2.12095

[37] Maity A, Pathak A, Saha G. Transfer learning based heart valve disease classification from Phonocardiogram signal. Biomedical Signal Processing and Control. 2023; 85:104805. https://doi.org/10.1016/j.bspc.2023.104805

[38] Zhang W, Han J, Deng S. Abnormal heart sound detection using temporal quasi-periodic features and long short-term memory without segmentation. Biomedical Signal Processing and Control. 2019; 53:101560. https://doi.org/10.1016/j.bspc.2019.101560

[39] Sawant NK, Patidar S, Nesaragi N, Acharya UR. Automated detection of abnormal heart sound signals using Fano-factor constrained tunable quality wavelet transform. Biocybernetics and Biomedical Engineering. 2021; 41 (1):111-26. https://doi.org/10.1016/j.bbe.2020.12.007

[40] Chen Y, Wei S, Zhang Y. Classification of heart sounds based on the combination of the modified frequency wavelet transform and convolutional neural network. Medical and Biological Engineering and Computing. 2020 ;58:2039-47. https://doi.org/10.1007/s11517-020-02218-5

[41] Hazeri H, Zarjam P, Azemi G. Classification of normal/abnormal PCG recordings using a time–frequency approach. Analog Integrated Circuits and Signal Processing. 2021; 109 (2):459-65. https://doi.org/10.1007/s10470-021-01867-2

[42] Zeng W, Yuan J, Yuan C, Wang Q, Liu F et al. A new approach for the detection of abnormal heart sound signals using TQWT, VMD and neural networks. Artificial Intelligence Review. 2021; 54 (3):1613-47. https://doi.org/10.1007/s10462-020-09875-w