Information Sciences Letters
An International Journal

http://dx.doi.org/10.18576/isl/121101

Predicting heart disease using modified GoogLeNet convolutional neural network architecture based on the heart sound

Abdullah R. Rashwan[1,*], *Laila El Fangary* [2] *and Safaa M. Azzam*[2]

[1]Department of Software Engineering, Faculty of Computers and Information, Kafrelsheikh University, Kafrelsheikh, Egypt
[2]Department of Information Systems, Faculty of Computers and Artificial Intelligence, Helwan University, Helwan, Egypt

Received: 5 Aug. 2023, Revised: 21 Sep. 2023, Accepted: 25 Oct. 2023
Published online: 1 Nov. 2023

Abstract: Based on the data of the world health organization (WHO), diagnosing heart disease is a great task, as heart disease (HD) is the most prevalent disease worldwide. We suggested a method based on heart sounds to deal with this difficult issue because the heart sound (HS) is an essential component for detecting heart conditions. A feature extraction technique and a classifier are used in the suggested strategy. We use the GoogLeNet convolutional neural network (CNN) architecture with some modifications to separate the most crucial attributes of HS, and the heart condition is classified as diseased or not diseased based on these attributes. The model is trained using the AdaBelief optimizer to tune the parameters of our modified GoogLeNet architecture. The model was trained and validated utilising various datasets from PhysioNet 2016. Additional training samples were provided by integrating the PASCAL dataset with the PhysioNet 2016 dataset. Additionally, the variety of samples from various sources enabled our system to learn about sounds from everyday life more accurately. Our results indicated that using a modified GoogLeNet architecture with the AdaBelief optimizer, the trained model obtained test accuracy of 100% and 99.9% on unseen HS recordings from PhysioNet and the merged datasets, respectively. By comparing our proposed model with the highest-scoring methods listed on the official PhysioNet website in these datasets, the results show significantly improved.

Keywords: Deep Learning ,GoogleNet , Feature extraction ,Heart disease prediction ,phonocardiogram.

1 Introduction

As stated by the WHO, HD is the top reason for global mortality. It is plausible to assert that cardiovascular disease accounts for around 30% of all fatalities. Heart disease imposes an enormous burden on the public health sector, causing 18 million deaths annually. As a result, the demand for early-stage heart disease diagnosis is unavoidable in order to avoid early death in persons at threat and to help clinicians make predictions [1–3]. Physicians are diagnosing heart disease using wearable monitors and a variety of tests, such as X-rays of the chest, stress checks, electrocardiograms, and coronary angiograms. However, mining suitable heart disease threat features from electric diagnostic valuations is hard as specialists make every effort to detect patients fast and accurately. These checks are time wasting, expensive, and have adverse consequences. Another way to detect the condition of the heart depends on the medical professional's knowledge and experience and is done physically using a stethoscope. Sometimes this way leads to miss diagnosis since heart sound has almost similar patterns. Although it may appear simple, heart sound auscultation requires extensive training because it needs the capability to separate among subtle changes in that sound. This is something that professionals can do. However, these specialists are not always accessible, mainly in rural places where the availability of specialists is limited. For the medical assessment of heart auscultation, the correctness of expectation by medical students is 20 to 40 percent, but it is around 80 percent for cardiologists who are professionals [4–8]. As a result, alternative methods of carrying out such a procedure are required to aid physicians in making accurate diagnoses.

We can rely on phonocardiograms (PCG) or electrocardiograms (ECG) for diagnosing heart diseases, where the ECG indicates the electrical activity and the PCG indicates the sounds produced by the human heartbeat. The auscultation

[*] Corresponding author e-mail: abdullahrashwan16@gmail.com

© 2023 NSP
Natural Sciences Publishing Cor.

method is now the most popular technique for predicting cardiovascular problems. This is accomplished through stethoscope-based monitoring and analysis of the cardiac sounds. So in this research, PCG is used to detect heart diseases because it provides essential information during one cycle of the heart's beat [9, 10].

When the heart sound was analyzed, it consisted of four main sounds called HS1, HS2, HS3, and HS4. HS1 and HS2 both refer to normal heart sounds, and these sounds are a result of the semilunar valve closing and the atrioventricular valve opening. While HS3 and HS4 refer to abnormal sounds when a wave of blood reaches the inside of the ventricles, HS3 shows up. As a result of atrial systole, HS4 appears before HS1. These components, which are segmented for the purpose of detecting cardiac problems, include important information about the heart system. Each sound is included in Table1 along with its frequency and duration [11].

Table 1: Different heart sounds' duration and frequency

The sound of the heart	The frequency	Period
HS1	from 30 to 45 Hz	from 0.1 to 0.12
HS2	from 50 to 70 Hz	from 0.08 to 0.14
HS3	less than 30 Hz	from 0.04 to 0.05
HS4	less than 20 Hz	from 0.04 to 0.05

Branches of artificial intelligence (AI) help people make decisions with high accuracy and work to reduce human errors. Currently, a variety of AI methods are being applied in medicine to precisely diagnose illnesses. The key element of data science that makes machines intelligent is AI, where every expert system needs to be capable of learning. AI employs several learning-based methods, like deep learning (DL), robotics, machine learning (ML), and others [12].

A few years ago, the DL approach saw tremendous success in a variety of fields and applications. On unstructured data, DL performs better than ML. DL enables the regular learning of features from samples at different layers. Due to the growing complexity of healthcare data, DL approaches have become widely used in healthcare applications. By using a massive amount of health records to uncover hidden patterns and mine important characteristics, DL techniques offer models for data analysis that are effective and efficient compared to traditional analytics, which cannot generate results quickly enough [13–17].

One of the most significant DL networks is considered to be CNN, as it has proven high efficiency in image recognition []. There are many DL networks, but CNN has proven to be highly effective in many applications that need to process pixel data. Where CNN provides a set of filters that extract the most significant attributes from the data in numerous image processing (IP) and computer vision (CV) applications. CNN, a deep learning model, has demonstrated good performance in these applications that require object recognition, such as self-driving cars and facial recognition []. It is a misconception that CNN deals with images only, but it deals with data that can be represented by numbers, such as images (pixels) and sounds (since sound is a frequency) [18–20].

There are many architectural examples of CNN with various layers that make up the CNN model, such as VGG-16, AlexNet, LeNet-5, GoogLeNet, and others. GoogLeNet is a deep CNN architecture proposed by a team at Google whose objective is to enhance how the network uses its computational resources during the classification and detection processes. Results improved by using computing resources more effectively, or in other words, by extracting more features with the same amount of processing. These were introduced with the idea of GoogLeNet's inception block that make the network wide rather than deep by having filters with several sizes that can run on the same level. To extract and integrate information from various scales, the GoogLeNet used groups of convolution kernels [21–23].

In Google's neural network, an "Inception block" is the fundamental convolutional block. As shown in diagram 2, there are four parallel branches in the block. The first three branches collect data from different spatial sizes using various layers of filter width: one-by-one, three-by-three, and five-by- five. One-by-one filters were applied in the middle of the block to minimise the channels and optimise the network. The fourth branch changes the number of channels using three-by-three max-pooling, accompanied by a one-by-one filter. The size of the input is preserved by using suitable padding throughout the four branches. As reflected in Figure 1, GoogLeNet employs a stack of nine inception blocks divided into three sets with global average pooling, max pooling, and filtering layers where the dimensionality is reduced by max-pooling between inception blocks [23, 24].

For the prediction of HD, several researchers (mentioned in 2) have proposed several methods based on HSs. However, it was noted that studies with high prediction accuracy relied on very small datasets, while studies that relied on large datasets did not achieve high accuracy. And because the prediction process is related to the field of health, the proposed method must be very accurate in order to rely on it with confidence and to overcome human errors in predicting the presence of disease or not. To achieve this, the model must be trained and tested on a large data set to verify results when high accuracy is achieved.

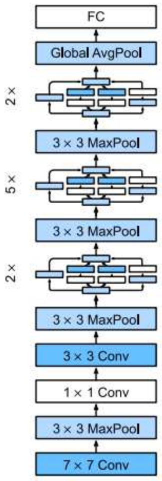


Fig. 1: The architecture of GoogLeNet

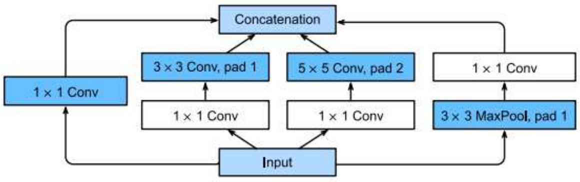


Fig. 2: Block structure for the inception

In this study[20], there is a disadvantage in determining the duration of the HS (only 30 seconds), because after 30 seconds there may be important information for determining the heart condition.

We created a novel DL model based on the GoogLeNet architecture with some modifications to extract significant attributes from HS and then use these attributes to predict whether the heart is diseased or not. Our network parameters were tuned using the AdaBelief optimizer. The following is a summary of this paper's major **contributions**:

- Overcome the drawbacks of training and testing the model on a small dataset in previous studies
- Modifications to the GoogLeNet and Inception block structures to match the heart sound data format
- Solve the limitations of using the limited duration of the HS in previous studies
- A reliable classification model capable of achieving high accuracy on a large dataset to validate results. Most of the previous studies that achieved high accuracy relied on a very small dataset, while most studies that relied on a large dataset did not achieve high accuracy. Therefore, the model must achieve very high accuracy on a large dataset because the problem is concentrated in health care
- The proposed architecture was evaluated using the PhysioNet dataset to compare the results with the highest scores listed on the official PhysioNet website. An evaluation was also carried out on the merged datasets (PhysioNet and PASCAL) to verify the results
- When we compare our proposed model to traditional methods in these two large datasets, the results are noticeably better

The following categories apply to this paper: An extensive summary of the relevant work is presented in Section 2. The suggested solution can be found in Section 3. The dataset, results, and comparison with others are described in 4, while the final Section 5 spells out the conclusion and the work to come in the future.

2 Related work

There are many studies on the use of AI methodologies in diagnosing heart diseases. This section contains a review of methodologies related to ML and DL in order to predict heart disease.

R. Valarmathi and T. Sheela [25] developed a sequential forward selection-based approach for predicting heart disease that would eliminate unimportant features. Two classifiers, Random Forest (RF) and XG Boost (XB), were relied upon. To tune the hyperparameter and enhance the functionality of the classifier models, the researchers applied set of algorithms: grid search, random search, and genetic programming. The system is evaluated using the Cleveland Cardiology (CHD) and the Z-Alizadeh Sani (ZS) datasets, where age, fasting blood glucose, etc. are prognostic features of cardiac status. The accuracy was 97.52% and 80.2% on the CHD and ZS datasets, respectively. However, it is possible to fall into the problem of over-fitting due to the smallness and the imbalance of the data between a patient and a normal person. Where the CHD dataset has 303 rows and the ZS dataset contains 303 examples; among them, 216 indicated the absence of heart disease, but 87 indicated the presence of heart disease. So the accuracy of this system can decrease in real time.

Kartik et al. [26] proposed a diagnostic system based on CHD examples. They used an XB classifier with an accuracy of 91.8% and a Bayesian algorithm to tune the hyperparameters for this classifier. This model should be evaluated on a different set of data to guarantee the reliability of the findings.

Similarly, [27] a new ML model for predicting heart disease was proposed on the basis of data mining methods. The RF, decision tree (DT), and hybrid algorithms (RF and DT) are used to implement the proposed model. The CHD is used to evaluate the proposed system. The highest accuracy is up to 88.7% with the hybrid model. However, prediction accuracy must be increased.

Furthermore, Deepika and Balaji [28] introduced a new HD prediction system with 97% accuracy based on a new optimised unsupervised method for feature selection and, for classification purposes, the Dragonfly algorithm with multi-layer perceptrons. The proposed system was evaluated using a dataset called ”UCI Clevel,” which contained 14 features with 305 examples. The limitation of this study is that they used a small dataset to assess the proposed system.

On the other hand, for the prediction of cardiovascular disease, Muhammad Nasiruddin and Rajib Kumar Halder [29] used a dataset available on Kaggle with 70,000 examples containing features like age, pressure, etc. For the purpose of selecting features, the suggested model uses the the Gain Attribute, Gain Percentage Attribute and Correlation Attribute Evaluators . For the purpose of classification, the suggested model uses the RF and Naive Bayesian algorithms. The suggested model was based on various ratios between the Kaggle heart disease dataset's training and testing sets (50:50 , 60:40, 70:30, 80:20, and 87.5:12.5) with accuracy of (98.88%, 99.53%, 99.98%, 98.36%, 96.66%, 97.77%, 99.56%, and 94.37%). However, the data must be scanned to find, collect associated features and to extract the hidden pattern to enable rapid learning and improve the detection accuracy in real time detection. This requires implementing a deep learning model.

Moreover, Rajkamal Rajendran and Anitha Karthi [3] combined an integrated set of datasets (Cleveland Cardiology ”CHDD”, Swiss, etc.) to build a confident classifier. This set required missing values to be computed based on continuous features. The feature engineering based on entropy was proposed. For the purpose of classification, they relied on ML algorithms such as K-nearest neighbours (KNN) and vector machines for support (SVM). The proposed approach obtained an accuracy of 92.7%. However, an improvement in accuracy is required as the issue is specifically related to health care.

Polipireddy Srinivas and Rahul Katarya [30] proposed an expert model (HyOPTX) using the XGBoost algorithm to classify heart conditions. They use OPTUNA (Ultra Parameter Optimization Technology) for hyper-parameter tuning. They accessed the CHD, HFP, and HDUCI datasets with an accuracy of 94.7%, 89.3%, and 88.5% , respectively. However, diagnostic accuracy should be increased to be able to rely on HyOPTX.

Jyoti Mishra et al. [31] identified key features for predicting heart disease using ML techniques. For the discovery of heart illnesses, a medical internet of things architecture based on recurrent neural networks (Rec-CONVnet) has been proposed. Clinical records are used to estimate heart disease risk based on age, pressure, and other features. Rec-CONVnet has an accuracy rate of 96.4%.

Yong Li et al.[32] suggested a new DL architecture (CraftNet) built on ECG signals and handcrafted features to predict HD. The proposed CraftNet network is influenced by the SVM idea. The designed loss function, named P-S, mixed with the architecture improves classification performance and generalisation ability. The MIT-BIH arrhythmia dataset is used to evaluate the proposed CraftNet. It contains 30 - 48 minutes recordings of an ECG logged at 360 Hz from 47 patients. The proposed CraftNet obtained an accuracy of 89.25%. The proposed CraftNet has the disadvantage that convergence was difficult during the training process due to the combination of P-S with the proposed CraftNet, which required great attention during the training process.

Haya al-Askar et al. [33] established a novel model that is built on HS to predict HD. For the purpose of obtaining attributes from HS, AlexNet was used. The SVM was used based on extracted features from AlexNet to classify the HS. They depended on the PhysioNet HS examples. The proposed system achieves an accuracy of up to 87%. However, the classification accuracy must be increased to rely on this model.

Aniruda Dutta et al.[34] developed a new methodology for predicting heart illness based on least absolute contraction and operator selection (lasso). To identify significant features from PCG signals, they relied on a feature weight estimate, which is monitored by co-voting feature selection. Significant features are smoothed using a fully connected layer, which is an important step before passing the layer's output to later convolutional stages. PhysioNet/Computing was used in the 2016 Cardiology Challenge dataset for model performance testing and achieved an overall accuracy of 87.31%, however, the linear nature of the estimator may cause a problem when using LASSO. LASSO is a penal regression technique in which the variance of variables is imposed by infinity.

Farahat Bint Azzam et al.[35] generated a method for detecting cardiac abnormalities based on the mathematical analysis of the influence of extra noise on the power-based features of the short-range slope filter bank and CNN layer. They proposed combining spectrogram-image features that are logarithmic and linear, which are then fed into a residual CNN network (ResNet) to classify PCG signals. The model has been evaluated in the Physionet dataset with 85.08% accuracy. The validation of the proposed method is empirical by nature, which is a major drawback. However, in order for us to rely on this model, efficiency must be increased because it is a health problem.

Omer Deperlioglu [36] developed a new methodology centered on the direct classification of energy values in the heart sound data. Classification was done using the directly employed energy of the resampled technique with the Stack AutoEncoder Network (SAEN). With an accuracy of 99.80%, the SAEN model used HS examples from a BPASCAL dataset [37]. The BPASCAL is a small dataset; it only contains 449 heart sound samples, so this dataset should be combined with another similar dataset to verify the results.

Utku Kose et al. [38] introduced a confident system for predicting heart illness using the Internet of Things for Health (IOH), based on secure data processing with a Tangle approach and multiple forms of authentication. For the purpose of classification, the suggested system uses Autoencoder Neural Networks (AEN). For the dataset, they relied on two sources: the first dataset was from PhysioNet, where Set A of the PhysioNet datasets was used, and the second source was PASCAL B. Normal heart sounds were diagnosed with an accuracy of 96.03%, while abnormal sounds were 90.11% and 91.91% for both murmurs and extrasystoles, respectively. By considering two separate datasets, the performance of embedded system diagnostics (together with AEN infrastructure) was evaluated. However, continuous training using the most recent data and examples may be required to make the AEN adaptable and prevent it from developing bias. There may still be some unexpected scenarios in terms of IOH architecture. Building materials, for example, may cause interruptions or problems with the wireless communication of signals.

Ali A. Samir et al. [20] established a new methodology that relies on the DL strategy to predict HD. They relied on CNN to obtain significant attributes from the HS and then classify the heart as diseased or not based on these attributes. To fine-tune CNN's hyperparameters, they used the JSO optimizer. The proposed CNN-JSO model relied on two sources of data: ”PhysioNet” and ”Kaggle.” with a test accuracy of 94.12% and a training accuracy of 97.76%. However, there are limits on the duration of heart sounds (only 30 seconds). Therefore, the accuracy of the test with the acceptance of the heart sound should be increased without limits in its duration, because after 30 seconds there may be important information for determining the state of the heart.

CardioXNet is developed by [39] for classification of heart disease using heart sound. They did not aim to overuse parameters and memory resources during training, so they relied on the learning strategies of representation and sequencing. For extracting main features from PCG signals, representation learning employs an adaptive feature extractor. For extracting temporal features, sequence residual learning is used. They relied on the PhysioNet dataset with an accuracy of 86.57%. However, the proposed model could be improved if multiple larger PCG datasets with different cardiology annotations are used. Several advanced networking strategies can be tried to achieve better PCG auto-classification performance when CardioXNet is integrated into wearable devices.

Haoran Kui et al.[40] developed a novel methodology that relies on HS to detect HD. The wavelet algorithm was used to clean the HS of any noise. The segmentation process of PCG signals was done by a model of hidden Markov. They used the dynamic frame length method to obtain significant attributes from the HS. Finally, these features were classified using CNN. They constructed two- and multi-classifiers with an accuracy of 93.89% and 86.25%, respectively. They relied on heart sounds collected from Kunming and Fuwei heart hospitals. This dataset contains 1800 cases of heart sounds, divided into 900 normal cases and 900 abnormal cases. However, this dataset is balanced, but more cases need to be collected to increase the accuracy of the proposed technique.

El Badlaoui et al. [41] proposed an approach to predict HD based on HS with an accuracy of 96% and 100% in two different datasets. They employed an analysis of principal components algorithm to reduce the amount of data. They depended on the SVM classifier and k-nearest neighbours (kNN) to classify the HS. However, this model has been archived with high accuracy, but it is not reliable because the two datasets they used were very small, the first containing 31 normal and 34 abnormal heart sounds (31 heart sounds for training and 34 heart sounds for testing), and the second containing 200 normal and 66 heart sounds (133 heart sounds for training and 133 heart sounds for testing). So they have to combine these two datasets with another big dataset.

More studies listed in Table 2 have made significant contributions to the prediction of HD based on HS. However, there are various obstacles to overcome in classifying a heart as diseased or not:

–Heart disease prediction utilising a new methodology based on heart sounds (PCG signals)
–Using the sound of the heart without specifying the duration
–Use appropriate feature extraction methods with large structural differences
–Use a dataset of appropriate size and balance (between an abnormal heart and a normal heart)

© 2023 NSP
Natural Sciences Publishing Cor.

Table 2: Detailed literature based on heart sound (PCG signals) for predicting heart disease

Ref	Year	Dataset	Approach	Accuracy
[42]	2020	PhysioNet, 2016	Feature maps from different clique blocks. 1D-CNN algorithm.	91%
[43]	2021	PhysioNet, 2016	The frequency domain features. CNN algorithm.	85%
[44]	2021	PhysioNet, 2016	Seesaw loss for Long-Tailed PCG signals with Res2Net.	87%
[45]	2021	PhysioNet, 2016	Wavelet scattering features. SVM algorithm.	92.23%
[46]	2021	PASCAL A and B, 2011	Discrete wavelet based features. Hidden Markov Models (HMM).	92.74%
[47]	2021	PhysioNet, 2016 and PASCAL	The time and frequency domain features. Light gradient boosting model.	92.88%
[48]	2021	PhysioNet 2016 and PASCAL 2011	Transforming HS into a recurring spectrogram based on patterns. Residual neural and CNN architecture networks: Xception, dense, mobile, and VGG16 based on CWT models.	89.04% (PhysioNet 2016, DenseNet model) 92.96% (PASCAL 2011, VGG model)
[49]	2021	PhysioNet, 2016	PCG-ECG signals based on time-frequency. kNN, SVM, and Ensemble.	93.14%
[50]	2021	PhysioNet, 2016	The time–frequency representation. Differential evolution with a straightforward genetic algorithm	93.60%
[51]	2021	Private	Time and frequency domain features. kNN algorithm.	94%
[52]	2022	PhysioNet, 2016	Carefully constructed processing pipeline to accurately identify individual heartbeats in PCG recordings. A bi-directional long-short-term memory structure then categorised these recordings.	68%
[7]	2022	the PASCAL CHSC	Features were acquired using a Spectrogram signal representation. VGG16 ,VGG19,MobileNet ,inceptionV3 models.	80.25%, 85.19%, 72.84% 54.32%,
[53]	2022	PhysioNet, 2016 and PASCAL CHSC	Features of the discrete wavelet transform and the mel-frequency cepstral coefficients (MFCCs). RF classifier	88.7% and 86.16% for the PhysioNet and Pascal CHSE datasets, respectively.
[10]	2022	PASCAL A & B, 2011	AlexNet to obtain attributes from the HS + SVM classifier.	90%
[54]	2022	Private	MFCCs, spectral and statistical features. kNN + SVM.	91.10%
[55]	2023	PhysioNet, 2016	The features that are retrieved from the segmented sound signal are put through a Bonferroni mean-based fuzzy K-nearest centroid neighbour classifier.	88.4%
[56]	2023	PhysioNet, 2016	Using conventional temporal-frequency visualisations as input attributes. Pre-trained CNNs (transfer learning)	92.23%
[57]	2023	PhysioNet, 2016	Attributes of PCG time and frequency signals. 2-D CNNs were created.	93.07%

–Maximize classification accuracy

So, given the limitations of current cardiac detection methods, we aim to develop a technique to enhance HD diagnosis.

3 Proposed model

3.1 Data preprocessing

In order to obtain significant attributes from the HS to determine whether the heart as diseased or not, we must first prepare the sounds to remove the noise and determine the duration of the sound to train the model. Figure 3 shows the steps involved in preparing HSs. In the first step, the first second of sound was cut off as noise from the recording device. Looking at Table 1, we find that the highest frequency of HSs is 70 Hz, so noises with a frequency of more than 70 Hz were removed. Then, after these steps, it is ensured that the length of the sound is not less than 2 seconds (s). If it is greater than 3.6 s, then this sound is divided into sounds, each equal to 3.6 s. If it is less than 3.6 s and greater than 2 s, the sound is repeated until it is equal to 3.6 s. A sampling rate (SR) of 22050 is used. This means that the sound has 22050 numeric values per second. To enhance the performance and training stability of the model, the average of each five values in the sound was taken to reduce the sound shape, and then normalisation was performed in the last step.

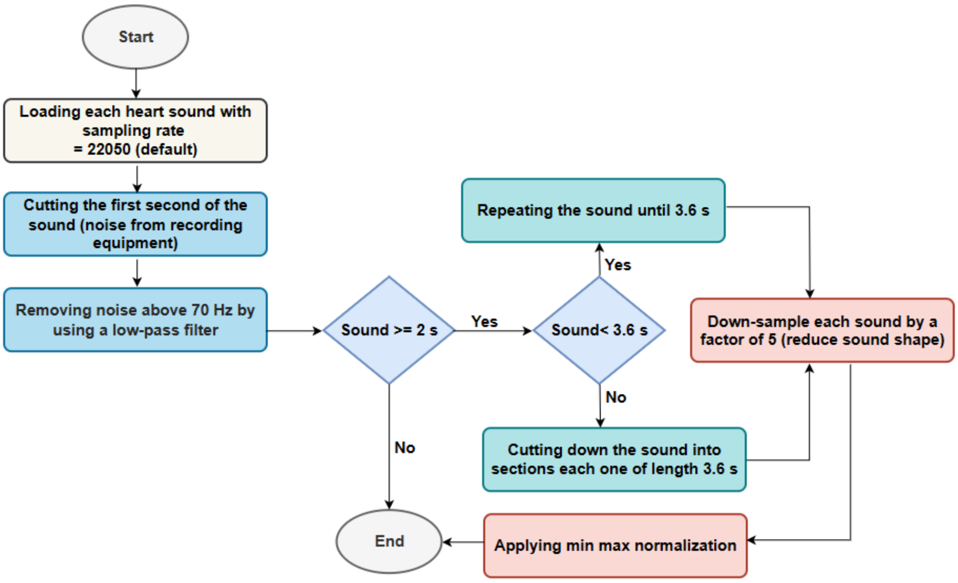

Start
Loading each heart sound with sampling rate = 22050 (default)
Cutting the first second of the sound (noise from recording equipment)
Removing noise above 70 Hz by using a low-pass filter
Sound >= 2 s
Yes
Sound< 3.6 s
Yes
Repeating the sound until 3.6 s
No
Cutting down the sound into sections each one of length 3.6 s
Down-sample each sound by a factor of 5 (reduce sound shape)
No
Applying min max normalization
End

Fig. 3: Data preprocessing

3.2 Modifications to the GoogLeNet Convolutional Neural Network architecture

Information on the various layers of the redesigned GoogLeNet and Inception block was shown in tables 11 and 12.

By comparing the GoogLeNet architecture as shown in Figures 1 and 2 with the proposed architecture shown in Figures 5 and 4, we can see the proposed modifications to the GoogLeNet architecture to extract the most important features from the PCG signals and then classify the heart as diseased or not.

–Changing the dimensions of the layers from 2D to 1D
–Changing the size of the kernels and pool size of the Maxpooling layers in the GoogleNet network
–Changing a convolution layer to a separable convolution layer
–Adding SpatialDropout1D layers to the architecture
–Adding Batch Normalization layers to the architecture

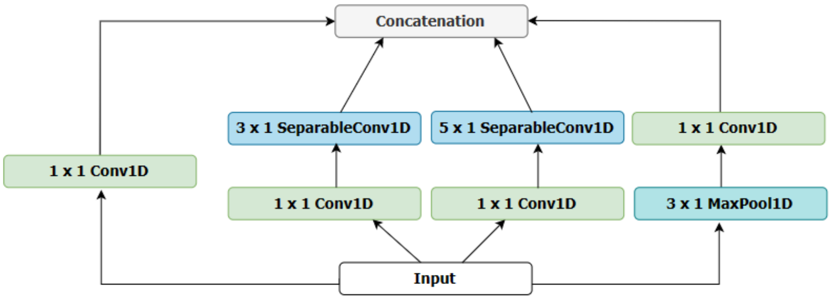


Fig. 4: Modified Inception Block

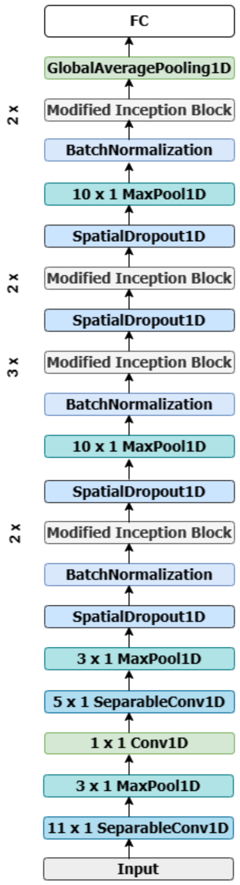


Fig. 5: Modified GoogLeNet Convolutional Neural Network architecture

The dimensions of the layers were changed from 2D to 1D to match the shape of the HS data, as the full sound was represented as a continuous number value in one row. The layers were also resized to fit the row size of the HS data, as the best fit size was reached to adequately extract the features during model training.

3.2.1 separable convolution layer

In order to explain the reasons for using a separable convolution layer instead of a traditional convolution layer, we first explain how the traditional CNN works.
As shown in this study [20], 1D CNN extracts features from a row of data by separating that row into smaller sets and

then performing some operations on these small subsets using the kernel. The set of values that make up the kernel are treated as parameters and chosen by training. These parameters are changed to obtain the optimal values during training that help us extract features from the data by multiplying these values with each element of the subset. Let’s use an example to clarify.

Data ”x”:

0.0	0.6	1.3	2.3	3.2	4.3	4.7	5.8	6.8

kernal values ”k”:

-1	0	1

Since the kernel ”k” has a size of three, the row of data ”x” is divided into smaller subsets, each subset ”d” containing three elements. The multiplication operation is performed by the following equation:

$$Y = \sum_{i=0}^{2} k_i d_i \tag{1}$$

First subset of data ”d” = [0.0, 0.6, 1.3].
Kernel values ”k” = [1, 0. 1].
By applying equation 1, the output:

$Y = (1 \times 0.0) + (0 \times 0.6) + (1 \times 1.3) = 0 + 0 + 1.3 = 1.3$
Depending on the number of steps, the next subset is selected. taken with a value of at least 1 from the first element of the preceding subset. This number of steps is called a stride ”S”. The following rows shown in tables, from 3 to 9, show subsets of data where k size = 3 and S = 1.

Table 3: A Subset (a)

0.0	0.6	1.3	2.3	3.2	4.3	4.7	5.8	6.8

Table 4: A Subset (b)

0.0	0.6	1.3	2.3	3.2	4.3	4.7	5.8	6.8

Table 5: A Subset (c)

0.0	0.6	1.3	2.3	3.2	4.3	4.7	5.8	6.8

Table 6: A Subset (d)

0.0	0.6	1.3	2.3	3.2	4.3	4.7	5.8	6.8

Table 10 presents the extracted features (final result) as a result of applying the kernel ”k” to the row of data ”x”. From equation 2, the shape of the result can be calculated, where W and Q represent the shape of the data, T and O represent the shape of the kernel, and S represents the stride.

$$\left(\left[\frac{W-T}{S} + 1 \right], \left[\frac{Q-O}{S} + 1 \right] \right) \tag{2}$$

© 2023 NSP
Natural Sciences Publishing Cor.

Table 7: A Subset (e)

0.0	0.6	1.3	2.3	3.2	4.3	4.7	5.8	6.8

Table 8: A Subset (f)

0.0	0.6	1.3	2.3	3.2	4.3	4.7	5.8	6.8

Table 9: A Subset (g)

0.0	0.6	1.3	2.3	3.2	4.3	4.7	5.8	6.8

Table 10: The end result of the data after applying the kernel

1.3	1.7	1.9	2	1.5	1.5	2.1

Then the shape of the result in the previous example is as follows:

$$\left(\left[\frac{1-1}{1}+1\right],\left[\frac{9-3}{1}+1\right]\right)=(1,7).$$

To extract more than one feature, more than one kernel is used. These kernels are lined up together, and the number of kernels is called a channel. So the dimension of the previous example is expressed as one row, seven columns, and one kernel (1,7,1). If three kernels are used (three channels), the shape of the result would be (1,7,3). Each kernel extracts one feature from the same and whole data through the operations described above. All kernels have the same size but different values.

From the comparison illustrated by figures 6 and 7, the 1D separable convolution divides the process of the original 1D convolution into two sub-processes: the first is called "spatial convolution," and the second is called "channel convolution." Spatial convolution extracts the middle features of the input signal, and then the final output (features) is extracted from these middle features using point-by-point convolution. The output of the separable and original convolutions is the same, but separable convolution decreases the number of calculations and parameters [58], which enhances system efficiency and speeds up model training compared to traditional convolution as a result of dividing the original convolution process into two sub-processes. For these reasons, a separable convolution was used in the proposed modified GoogLeNet architecture.

3.2.2 Spatial Dropout 1D layers

It is an excellent tool for reducing overfitting in a model, so spatial dropout 1D layers were added to the proposed architecture to avoid overfitting during training with heart-sound data, as the first set of training examples influences learning. This, in turn, prevents the learning of features that only appear in subsequent samples or groups. Due to the fact that neighbouring frames in feature maps that are tightly connected (as is frequently the case in early convolution layers) will reduce the effective learning rate, the spatial dropout 1D layer drops entire 1D feature maps during training to create independence between feature maps.

3.2.3 Batch Normalization layers

Although we have normalised the data in pre-processing, the input layer in our architecture is not the only input layer. For the network, the output of layer R-1 acts as the input of layer R. The input value distribution for each layer in the network is influenced by different ranges as a result of operations in the layers before it during the network training. Varying the values slows down the training speed of the model because we have to use a small learning rate (it takes a long time to converge). To solve this problem, batch normalisation was used, which resolves the variance of the values and thus speeds up the training of the model. This is accomplished by subtracting the mean from each of the input features and dividing the result by the standard deviation [59].

The input shape in the first layer is (1, 16000) and the kernel size is (1, 11, 32) with a stride equal to 1, so the output shape of the first layer is (1, 15990, 32). This output is the input to the next layer, and it is big for training, so the max-pooling technique was used in our architecture to reduce the input shape by dividing the input raw data into small subsets according to a pool size and a stride value, and then each subset is converted to the maximum value in this subset. Let's use an example to clarify; we'll use max-pooling with a pool size and stride of 3.

0.0	0.6	1.3	2.3	4.3	3.2	4.7	5.8	6.8

Max of (0.0 & 0.6 & 1.3) = 1.3
Max of (2.3 & 4.3 & 3.2) = 4.3
Max of (4.7 & 5.8 & 6.8) = 6.8

Final output:

1.3	4.3	6.8

A ReLU activation function has been implemented after each separable layer and a 1 x 1 conv for a non-linearlity implementation where non-linear layers expand the capabilities of the model. This is to avoid the problems of linear equations in training and to reach accurate results. ReLU also helps avoid the vanishing gradient problem (VGP) [60].

3.2.4 Modified Inception Block

From figures 2 and 4,

- The dimension of max-pooling layer in the fourth branch changed from 2D to 1D.
- The 3 x 3 and 5 x 5 convolutional layers were converted to 1D separable convolution (3 x 1 and 5 x 1) layers.

The last layer before classification is the global average booling (GAB) to convert the feature maps from the previous layers into a single vector, and then this vector is fed into fully connected network (FC) with the softmax layer to classify it as normal or abnormal. This is done by averaging all the values in each feature map and representing this as a single element in the vector. So feature maps can easily be inferred as species guarantee maps. No optimization parameter is needed in this layer. Moreover, the GAB totals out the spatial information, so it is very robust to spatial translations of inputs. One of the reasons for using this layer is to solve the limitations of using the limited duration of the heart sound in previous studies. Where GAB converts any input dimension (height and width) to 1 x 1 x C, if you enter any length, it will convert to 1 x 1 x C, where C is the number of channels, hence it gives no error.

To accomplish this, a fully connected neural network (FC) with 2 neurons (dense layer) followed by the SoftMax activation function to perform the final prediction. Softmax returns an array of length equal to the number of classes, in our case 2 (normal or abnormal), with probabilities distributed over them. The softmax ensures that the sum of all our output probabilities will be equal to 1. Although our problem is a binary classification (normal or abnormal), we use softmax instead of the sigmoid activation function as usual to avoid the problem of VGP [61]. The following is the softmax function's equation 3:

$$Softmax(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \tag{3}$$

- $\vec{z}$: The softmax function's input vector, composed of $(z_0, ...z_K)$
- All of the z_i values are elements of the softmax function's input vector, and they can take any real value, positive, zero, or negative. A neural network, for example, could output a vector like (-0.52, 7.12, 2.23), which is not a valid probability distribution, which is why the softmax is required
- e^{z_j}: Every element of the input vector is subjected to the standard exponential function. This results in a positive value that is greater than zero, which will be very little if the input was negative and very huge if the input was large. It is still not fixed in the range (0, 1), which is what a probability must be.
- $\sum_{j=1}^{K} e^{z_j}$: This term ensures that all of the function's output values sum to 1 and are in the range (0, 1), resulting in a valid probability distribution (normalization)
- K: Refer to the number of classes

The model's output (nromal or abnormal) is combined with the data labels to calculate the loss function; the purpose of the training process is to decrease the loss function value. The equation of the loss function is expressed as:

$$l = y\log(y') + (1-y)\log(1-y') \tag{4}$$

$$L = \sum_{i=0}^{n} l_i \tag{5}$$

© 2023 NSP

- L stands for the overall dataset loss,
- l stands for the loss of a single input,
- y stands for the input label,
- y' displays the predicted label for this input,
- n stands for the total number of inputs in the dataset.

Our proposed technique is illustrated in Figure 8.

Table 11: Information on the various layers of the redesigned GoogLeNet architecture

Layer	Size	Stride	Activation
Input	16000 x 1		
SeparableConv1D	11 x 1 x 32		ReLU
MaxPool1D	3 x 1	1	
Conv1D	1 x 1 x 32		ReLU
SeparableConv1D	5 x 1 x 32		ReLU
MaxPool1D	3 x 1	2	
SpatialDropout1D			
BatchNormalization			
Inception block (2a)			
SpatialDropout1D			
MaxPool1D	10 x 1	5	
BatchNormalization			
Inception block (3b)			
SpatialDropout1D			
Inception block (2c)			
SpatialDropout1D			
MaxPool1D	10 x 1	5	
BatchNormalization			
Inception block (2d)			
GlobalAveragePooling1D			
Fully connected network			Softmax

3.3 AdaBelief optimization algorithm

In this study, we relied on the AdaBelief algorithm to tune the parameters of our architecture. This is due to the fact that this algorithm simultaneously achieves three objectives: training stability in challenging environments, strong generalisation similar to stochastic gradient descent(SGD), and quick convergence similar to adaptive approaches. Deep neural network parameters are iteratively updated by AdaBelief algorithm by completely utilizing "belief." The gradient's prediction correctness is what determines the dependability of the "belief," and the choice of the smoothing parameter "β_1" is what determines this prediction accuracy. The calculation of the totally squared difference between the observed gradient "gt" and the predicted gradient "mt" for the loss function is the core of AdaBelief. Let "$gt = \triangledown ft(\theta_t)$"be the observed

Table 12: Information on the various layers of the redesigned Inception block

Branch number	Layer name	Size				Stride	Activation
		a	b	c	d		
First	Conv1D	1 x 1 x 32	1 x 1 x 64	1 x 1 x 64	1 x 1 x 128		ReLU
Second	MaxPool1D	3 x 1	3 x 1	3 x 1	3 x 1	1	
Second	Conv1D	1 x 1 x 32	1 x 1 x 64	1 x 1 x 64	1 x 1 x 128		ReLU
Third	Conv1D	1 x 1 x 32	1 x 1 x 64	1 x 1 x 64	1 x 1 x 128		ReLU
Third	SeparableConv1D	5 x 1 x 32	5 x 1 x 64	5 x 1 x 64	5 x 1 x 128		ReLU
Fourth	Conv1D	1 x 1 x 32	1 x 1 x 64	1 x 1 x 64	1 x 1 x 128		ReLU
Fourth	SeparableConv1D	3 x 1 x 32	3 x 1 x 64	3 x 1 x 64	3 x 1 x 128		ReLU

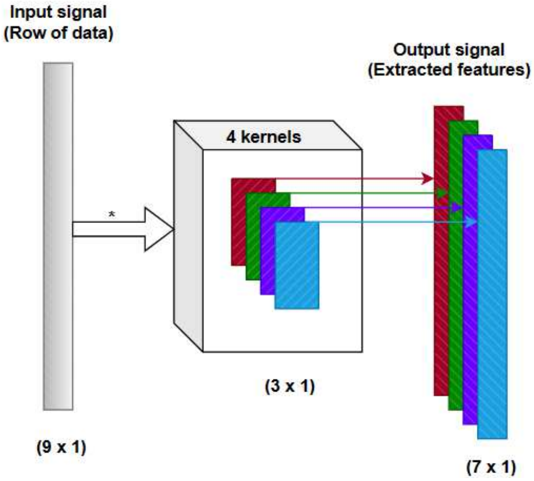


Fig. 6: Original 1D convolution

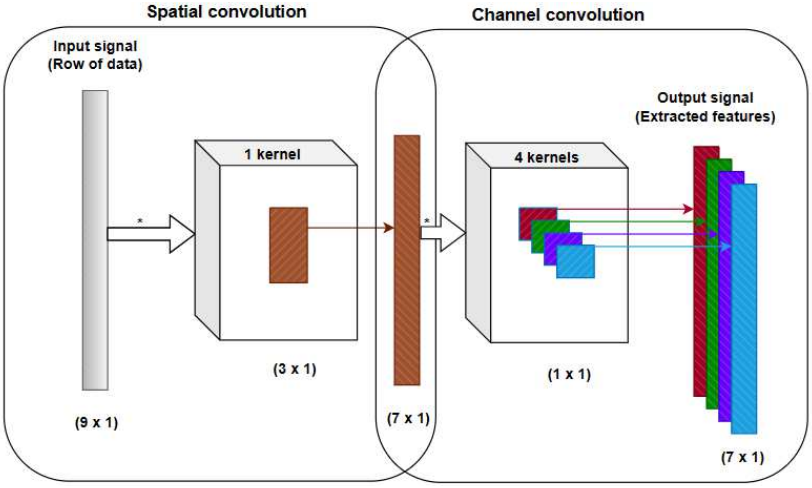


Fig. 7: Separable 1D convolution

gradient, and "$ft(\theta_t)$" denote the loss function of a Deep neural network. The AdaBelief algorithm suggested in [62] can be expressed as:

$$\hat{m}_t = \frac{\beta_1 \hat{m}_{t\text{-}1} + (1-\beta_1)g_t}{1-\beta_1^t} \tag{6}$$

$$\hat{s}_t = \frac{\beta_2 \hat{s}_{t\text{-}1} + (1-\beta_2)(g_t - \hat{m}_t)^2}{1-\beta_2^t} \tag{7}$$

$$\theta_t = \theta_{t-1} - \frac{\alpha \hat{m}_t}{\sqrt{\hat{s}_t} + \varepsilon} \tag{8}$$

β_1 and β_2 denote the degrees of the moving average decay, and the learning rate is denoted by α.

Notations used in the AdaBelief algorithm, as shown in algorithm 1

© 2023 NSP
Natural Sciences Publishing Cor.

Algorithm 1 AdaBelief algorithm

```
**Initialize** $\theta_0, m_0 \leftarrow 0, s_0 \leftarrow 0, t \leftarrow 0$
**while** $\theta_t$ not converged **do**
    $t \leftarrow t+1$
    $g_t \leftarrow \nabla_\theta f_t(\theta_{t-1})$
    $m_t \leftarrow \beta_1 m_{t-1} + (1-\beta_1) g_t$
    $s_t \leftarrow \beta_2 s_{t-1} + (1-\beta_2)(g_t - m_t)^2 + \varepsilon$
    **Bias Correction**
        $\widehat{m_t} \leftarrow \frac{m_t}{1-\beta_1^t}, \widehat{s_t} \leftarrow \frac{s_t}{1-\beta_2^t}$
    **Update**
        $\theta_t \leftarrow \Pi_{\mathcal{F}}, \sqrt{\widehat{s_t}}\left(\theta_{t-1} - \frac{\alpha \widehat{m_t}}{\sqrt{\widehat{s_t}}+\varepsilon}\right)$
**end while**
```

–$f(\theta) \in \mathbb{R}, \theta \in \mathbb{R}^d$: where θ is the parameter in $\mathbb{R}^d$ and f is the loss function to decrease,
–$\Pi_{\mathcal{F},M}(y) = \text{argmin}_{x \in \mathcal{F}} \|\mathrm{M}^{1/2}(x-y)\|$: Prediction of y on a convex possible set $\mathcal{F}$
–g_t : The step t of the gradient
–m_t: The exponential moving average of g_t
–s_t: The exponential moving average of$(g_t - m_t)^2$
–ε, α : ε is a low number, usually set to 10^{-8} and α is the learning rate, with a default value of 10^{-3}
–β_1 and β_2 : smoothing parameters, with usual values of 0.9 and 0.999, respectively
–β_{1t} is commonly set to be constant (for example, $\beta_{1t} = \beta_1, \forall t \in \{1,2,...T\}$) and represents the momentum for m_t at step t.

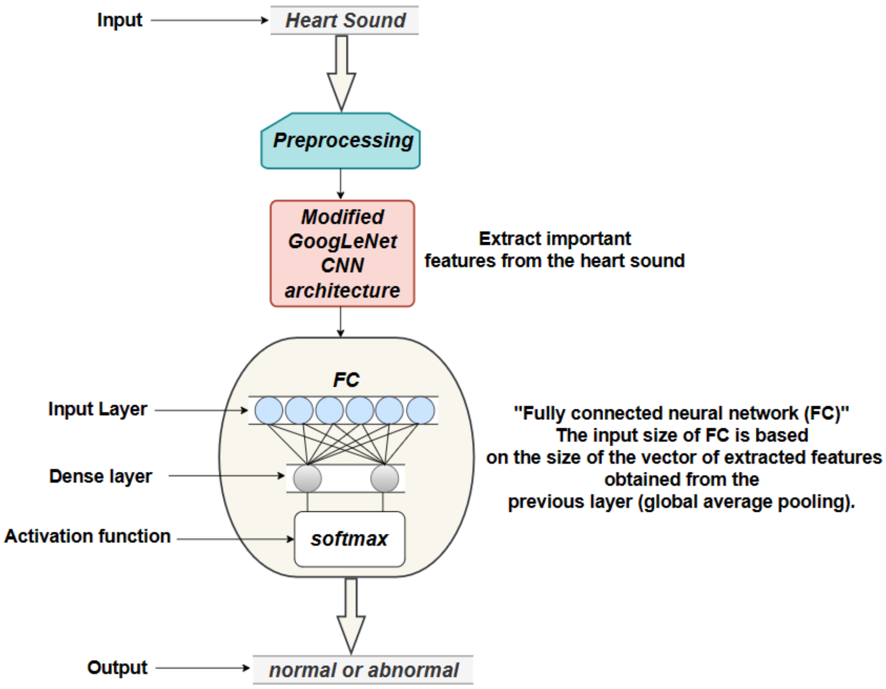

Input
Heart Sound
Preprocessing
Modified GoogLeNet CNN architecture
Extract important features from the heart sound
FC
Input Layer
Dense layer
Activation function
softmax
"Fully connected neural network (FC)" The input size of FC is based on the size of the vector of extracted features obtained from the previous layer (global average pooling).
Output
normal or abnormal

Fig. 8: proposed technique

© 2023 NSP
Natural Sciences Publishing Cor.

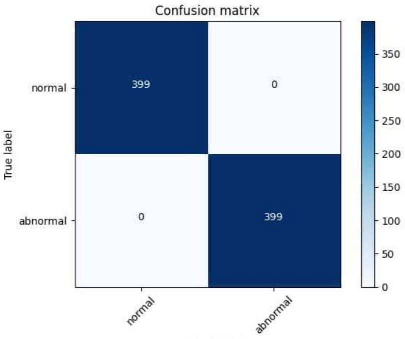


Fig. 9: Testing PhysioNet dataset

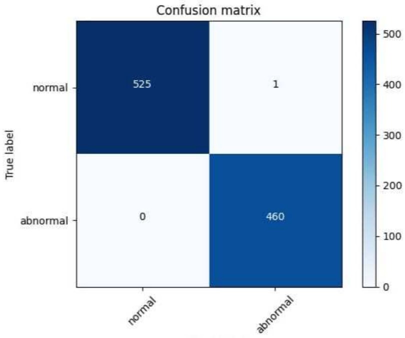


Fig. 10: Testing merged dataset

4 Results of experiments and comparisons

4.1 Dataset description

Table 13: Explanation of the combined dataset

Items	Explanation
First part, from 0 to 16000	refers to the full sound, which is represented by a float type (continuous value) in the range from -0.8 to 0.8.
Second part 16001	Sampling Rate (SR) is used to load audio and is represented as an int type with a value of 22050.
Third part 16002	Its value is 1 or 0, to know the status of each heart sound in the dataset (normal or abnormal).

© 2023 NSP
Natural Sciences Publishing Cor.

4.1.1 Kaggle dataset (PASCAL)

This dataset is very important, although it is small in size. It contains HSs collected from different sources along with noise sounds. This dataset is divided into two groups. From the first group, we use 34 murmur sounds and 31 normal sounds, and from the second group, we use 29 noisy murmur sounds, 120 noisy normal sounds, 66 murmur sounds, and 200 normal sounds [37].

4.1.2 Physionet dataset

This dataset was launched in 2016 to challenge the prediction of HD based on HSs. It consists of eight sets of HS data with 4,430 recordings from 1,072 individuals for a total of 233,512 HSs collected from both normal persons and patients [63].

4.1.3 Merged dataset

We address the limitations of the previous proposed methods for training and testing the model on small PCG datasets by integrating the two Kaggle and Physionet HS datasets. After cleaning the data by preprocessing the HSs, we get 5800 normal HSs and 5148 abnormal HSs. Each sound, as shown in Table 13, consists of 16002 features, which are split into three parts: the first 16000 features are the entire audio, and the final two features are the audio SR and label, respectively.

4.2 Performance metrics

Table 14: Explanation of Performance metrics values

Abbreviation	Description	Prediction
TN	The true negative is that the heart sound is actually normal, and it has a normal classification.	correctly
FP	The false positive is that the heart sound is normal, but it has an abnormal classification.	incorrectly
FN	The false negative is that the heart sound is abnormal, but it has a normal classification.	incorrectly
TP	The true positive is that the heart sound is actually abnormal, and it has an abnormal classification.	correctly

Various performance measures were used to evaluate the proposed system based on the confusion matrix values as shown in Table 14, including:

4.2.1 Accuracy

Refers to the percentage of true predictors (including true positives and true negatives) among all cases evaluated. From another perspective, the fraction of correct predictions of normal and abnormal heart sounds.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{9}$$

© 2023 NSP

4.2.2 Precision

This refers to the accuracy calculated only for positive persons. So this metric informs us how often we are right when we classify a person as positive (abnormal heart sound). This is done by calculating the fraction of true positive predictions out of all positive predictions. From another perspective, ”Of all the sounds that are classified as abnormal, how many predictions are correct?”.

$$Precision = \frac{TP}{TP+FP} \tag{10}$$

4.2.3 Recall

Also called sensitivity, this shows if the classifier is able to detect any abnormal sound at all. This is done by calculating the fraction of true positive predictions from all positive samples in the dataset. From another perspective, ”the number of sounds classified as abnormal relative to all abnormal sounds in the dataset”.

$$Recall = \frac{TP}{TP+FN} \tag{11}$$

4.2.4 specificity

$$specificity = \frac{TN}{TN+FP} \tag{12}$$

4.2.5 F1 score

The F1-score integrates the recall and precision into a lone metric by taking their harmonic mean, so we can’t develop a high F1 if either one is very low. F1-score only cares about sounds that the model expects to be positive (abnormal heart sounds) and sounds that are really positive (abnormal heart sounds).

$$F1\ score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{13}$$

4.2.6 Balanced Accuracy

The F1 score is not concerned with how many negative sounds (normal heart sounds) are in the dataset or how many correctly classified sounds. So the balanced accuracy metric was used to pay attention to negative sounds.

$$Balanced\ Accuracy = 0.5 * \left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP}\right) \tag{14}$$

4.2.7 Geometric Mean

Refers to the square root of the product of Precision and Recall.

$$Geometric\ Mean = \sqrt{Precision * Recall} \tag{15}$$

Table 15 shows the results of our proposed technique.

Table 15: Experimental Results

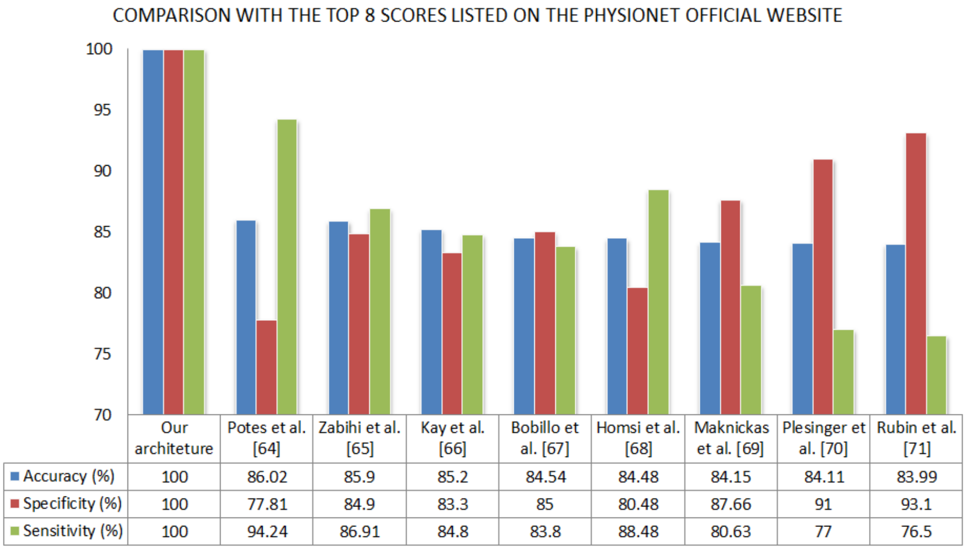
Dataset	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1-score (%)	Balanced Accuracy (%)	Geometric Mean (%)
PhysioNet	100	100	100	100	100	100	100
Merged	99.9	99.8	100	99.8	99.9	99.9	99.9

Table 16: comparison with the top 11 scores listed on the Physionet official website.

Dataset	Year	Score (%)	Reference
PhysioNet	**2023**	**100**	**our architecture**
PhysioNet	2016	86.02	[64]
PhysioNet	2016	85.90	[65]
PhysioNet	2016	85.20	[66]
PhysioNet	2016	84.54	[67]
PhysioNet	2016	84.48	[68]
PhysioNet	2017	84.15	[69]
PhysioNet	2016	84.11	[70]
PhysioNet	2016	83.99	[71]
PhysioNet	2016	82.82	[72]
PhysioNet	2016	82.63	[73]
PhysioNet	2016	81.85	[72]

Table 17: Comparison with the scores in the last five years

Dataset	Year	Score (%)	Reference
PhysioNet	2017	89.26	[74]
PhysioNet	2018	91.50	[75]
PhysioNet	2019	92.90	[76]
PhysioNet	2020	91.00	[42]
PhysioNet	2021	95.50	[77]
Merged	2021	94.11	[20]
PhysioNet	**2023**	**100**	**our architecture**
Merged	**2023**	**99.9**	**our architecture**



	Our architeture	Potes et al. [64]	Zabihi et al. [65]	Kay et al. [66]	Bobillo et al. [67]	Homsi et al. [68]	Maknickas et al. [69]	Plesinger et al. [70]	Rubin et al. [71]
🟦 Accuracy (%)	100	86.02	85.9	85.2	84.54	84.48	84.15	84.11	83.99
🟥 Specificity (%)	100	77.81	84.9	83.3	85	80.48	87.66	91	93.1
🟩 Sensitivity (%)	100	94.24	86.91	84.8	83.8	88.48	80.63	77	76.5

Fig. 11: Comparing the accuracy, sensitivity, and specificity of all competitors in a PhysioNet challenge

4.3 Comparative analysis

The PhysioNet (2016) competition [78] intends to promote the creation of techniques for classifying cardiac status based on HS recordings. Table 16 and figure 11 show the comparison with each competitor in this competition.
Table 17 shows comparisons with other competitors over the past five years.

5 Conclusion and future work

Several studies, as shown in Table 2 have made significant contributions to the prediction of heart disease based on PCG signals. However, most of the previous high-scoring research relied on a very small dataset, while most of the research with a large dataset did not score close to 99%. Because the problem is centred in healthcare, we still need a new and reliable classification model capable of achieving high accuracy on a large dataset to validate the results. Therefore, from a review of the previous literature presented in this study (Section 2), we proposed a new method to classify PCG signals extracted from normal and abnormal hearts. We relied on a dataset available in PhysioNet 2016. To expand the dataset, further training samples were made available by integrating the PASCAL dataset with PhysioNet 2016. By incorporating a new PCG dataset, we address the limitations of training and testing the model on the small PCG datasets of the previous proposed methods. Our approach is based on the GoogLeNet Convolutional Neural Network architecture with some modifications (described in section 3) to extract the most important features from the PCG signals and then classify the heart sound as normal or abnormal. The parameters of the modified GoogLeNet architecture were tuned by the AdaBelief optimizer. Our algorithm was tested on a combined dataset and also tested on PhysioNet 2016 to compare results with the highest scores listed on the official PhysioNet website. When compared, our proposed model achieved a significant improvement with a test accuracy of 100% and 99.9% for PhysioNet and the merged datasets, respectively. Therefore, high accuracy was achieved on a large dataset, and thus the defects of previous studies were overcome. The limitations of using the limited duration of the heart sound in previous studies have also been resolved. **In the future**, we will collect a dataset containing a description of the type of heart disease, not only whether the heart sound is normal or abnormal, so in the case of abnormalities, the type of disease is determined.

References

[1] World Health Organization (WHO). Overview and statistics on cardiovascular diseases (cvds). https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds), 2021. Online; accessed 20 January 2023.

[2] Hager Ahmed, Eman MG Younis, Abdeltawab Hendawi, and Abdelmgeid A Ali. Heart disease identification from patients' social posts, machine learning solution on spark. *Future Generation Computer Systems*, 111:714–722, 2020.

[3] Rajkamal Rajendran and Anitha Karthi. Heart disease prediction using entropy based feature engineering and ensembling of machine learning classifiers. *Expert Systems with Applications*, 207:117882, 2022.

[4] Hisaki Makimoto. Artificial intelligence in medicine (aim) in cardiovascular disorders. In *Artificial Intelligence in Medicine*, pages 813–823. Springer, New York, 2022.

[5] Maryam Hamidi, Hassan Ghassemian, and Maryam Imani. Classification of heart sound signal using curve fitting and fractal dimension. *Biomedical Signal Processing and Control*, 39:351–359, 2018.

[6] S Rajkumar, K Sathesh, and Neeraj Kumar Goyal. Neural network-based design and evaluation of performance metrics using adaptive line enhancer with adaptive algorithms for auscultation analysis. *Neural Computing and Applications*, 32(18):15131–15153, 2020.

[7] Omair Rashed Abdulwareth Almanifi, Mohd Azraai Mohd Razman, Rabiu Muazu Musa, Muhammad Yusri Ismail, Anwar PP Abdul Majeed, et al. The classification of heartbeat pcg signals via transfer learning. In *Recent Trends in Mechatronics Towards Industry 4.0*, pages 49–59. Springer, New York, 2022.

[8] Salvatore Mangione. Cardiac auscultatory skills of physicians-in-training: a comparison of three english-speaking countries. *The American journal of medicine*, 110(3):210–216, 2001.

[9] Sinam Ajitkumar Singh, Takhellambam Gautam Meitei, and Swanirbhar Majumder. Short pcg classification based on deep learning. In *Deep Learning Techniques for Biomedical and Health Informatics*, pages 141–164. Elsevier, Amsterdam, 2020.

[10] Shahid Ismail, Basit Ismail, Imran Siddiqi, and Usman Akram. Pcg classification through spectrogram using transfer learning. *Biomedical Signal Processing and Control*, 79:104075, 2023.

[11] Ashok Mondal, Parthasarathi Bhattacharya, and Goutam Saha. An automated tool for localization of heart sound components s1, s2, s3 and s4 in pulmonary sounds using hilbert transform and heron's formula. *SpringerPlus*, 2(1):1–14, 2013.

© 2023 NSP
Natural Sciences Publishing Cor.

[12] Simarjeet Kaur, Jimmy Singla, Lewis Nkenyereye, Sudan Jha, Deepak Prashar, Gyanendra Prasad Joshi, Shaker El-Sappagh, Md Saiful Islam, and SM Riazul Islam. Medical diagnostic systems using artificial intelligence (ai) algorithms: Principles and perspectives. *IEEE Access*, 8:228049–228069, 2020.

[13] Iqbal H Sarker. Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. *SN Computer Science*, 2(6):1–20, 2021.

[14] Christian Janiesch, Patrick Zschech, and Kai Heinrich. Machine learning and deep learning. *Electronic Markets*, 31(3):685–695, 2021.

[15] Amitha Mathew, P Amudha, and S Sivakumari. Deep learning techniques: an overview. In *International conference on advanced machine learning technologies and applications*, pages 599–608. Springer, 2020.

[16] Son Phung, Ashnil Kumar, and Jinman Kim. A deep learning technique for imputing missing healthcare data. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 6513–6516. IEEE, 2019.

[17] Shahab Shamshirband, Mahdis Fathi, Abdollah Dehzangi, Anthony Theodore Chronopoulos, and Hamid Alinejad-Rokny. A review on deep learning approaches in healthcare systems: Taxonomies, challenges, and open issues. *Journal of Biomedical Informatics*, 113:103627, 2021.

[18] Alessandro Achille and Stefano Soatto. Information dropout: Learning optimal representations through noisy computation. *IEEE transactions on pattern analysis and machine intelligence*, 40(12):2897–2905, 2018.

[19] PM Lavanya and E Sasikala. Deep learning techniques on text classification using natural language processing (nlp) in social healthcare network: A comprehensive survey. In *2021 3rd International Conference on Signal Processing and Communication (ICPSC)*, pages 603–609. IEEE, 2021.

[20] Ali A Samir, Abdullah R Rashwan, Karam M Sallam, Ripon K Chakrabortty, Michael J Ryan, and Amr A Abohany. Evolutionary algorithm-based convolutional neural network for predicting heart diseases. *Computers & Industrial Engineering*, 161:107651, 2021.

[21] Laith Alzubaidi, Jinglan Zhang, Amjad J Humaidi, Ayad Al-Dujaili, Ye Duan, Omran Al-Shamma, José Santamaría, Mohammed A Fadhel, Muthana Al-Amidie, and Laith Farhan. Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions. *Journal of big Data*, 8(1):1–74, 2021.

[22] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.

[23] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[24] Satawat Singprayoon and Siriporn Supratid. Effects of number and position of auxiliary networks used in inception convolutional neural network on object recognition. In *2021 9th International Electrical Engineering Congress (iEECON)*, pages 452–455. IEEE, 2021.

[25] R Valarmathi and T Sheela. Heart disease prediction using hyper parameter optimization (hpo) tuning. *Biomedical Signal Processing and Control*, 70:103033, 2021.

[26] Kartik Budholiya, Shailendra Kumar Shrivastava, and Vivek Sharma. An optimized xgboost based diagnostic system for effective prediction of heart disease. *Journal of King Saud University-Computer and Information Sciences*, 2020.

[27] Yu-Sheng Su, Ting-Jou Ding, and Mu-Yen Chen. Deep learning methods in internet of medical things for valvular heart disease screening system. *IEEE Internet of Things Journal*, 8(23):16921–16932, 2021.

[28] D Deepika and N Balaji. Effective heart disease prediction using novel mlp-ebmda approach. *Biomedical Signal Processing and Control*, 72:103318, 2022.

[29] Mohammed Nasir Uddin and Rajib Kumar Halder. An ensemble method based multilayer dynamic system to predict cardiovascular disease using machine learning approach. *Informatics in Medicine Unlocked*, 24:100584, 2021.

[30] Polipireddy Srinivas and Rahul Katarya. hyoptxg: Optuna hyper-parameter optimization framework for predicting cardiovascular disease using xgboost. *Biomedical Signal Processing and Control*, 73:103456, 2022.

[31] Jyoti Mishra, Mahendra Tiwari, Sanjay T Singh, and Sanjay Goswami. Detection of heart disease employing recurrent convoluted neural networks (rec-convnet) for effectual classification process in smart medical application. In *2021 4th International Conference on Recent Trends in Computer Science and Technology (ICRTCST)*, pages 389–394. IEEE, 2022.

[32] Yong Li, Zihang He, Heng Wang, Bohan Li, Fengnan Li, Ying Gao, and Xiang Ye. Craftnet: a deep learning ensemble to diagnose cardiovascular diseases. *Biomedical Signal Processing and Control*, 62:102091, 2020.

[33] Haya Alaskar, Nada Alzhrani, Abir Hussain, and Fatma Almarshed. The implementation of pretrained alexnet on pcg classification. In *International conference on intelligent computing*, pages 784–794. Springer, 2019.

[34] Aniruddha Dutta, Tamal Batabyal, Meheli Basu, and Scott T Acton. An efficient convolutional neural network for coronary heart disease prediction. *Expert Systems with Applications*, 159:113408, 2020.

[35] Farhat Binte Azam, Md Istiaq Ansari, Shoyad Ibn Sabur Khan Nuhash, Ian McLane, and Taufiq Hasan. Cardiac anomaly detection considering an additive noise and convolutional distortion model of heart sound recordings. *Artificial Intelligence in Medicine*, 133:102417, 2022.

[36] Omer Deperlioglu. Heart sound classification with signal instant energy and stacked autoencoder network. *Biomedical Signal Processing and Control*, 64:102211, 2021.

[37] Bentley Peter, Nordehn Glenn, M. Coimbra, and S. Mannor. The pascal classifying heart sounds challenge 2011 (chsc2011). http://www.peterjbentley.com/heartchallenge/, 2011. Online; accessed 5 Novamber 2022.

[38] Omer Deperlioglu, Utku Kose, Deepak Gupta, Ashish Khanna, and Arun Kumar Sangaiah. Diagnosis of heart diseases by a secure internet of health things system based on autoencoder deep neural network. *Computer Communications*, 162:31–50, 2020.

[39] Samiul Based Shuvo, Shams Nafisa Ali, Soham Irtiza Swapnil, Mabrook S Al-Rakhami, and Abdu Gumaei. Cardioxnet: A novel lightweight deep learning framework for cardiovascular disease classification using heart sound recordings. *IEEE Access*, 9:36955–36967, 2021.

[40] Haoran Kui, Jiahua Pan, Rong Zong, Hongbo Yang, and Weilian Wang. Heart sound classification based on log mel-frequency spectral coefficients features and convolutional neural networks. *Biomedical Signal Processing and Control*, 69:102893, 2021.

[41] O El Badlaoui, A Benba, and A Hammouch. Novel pcg analysis method for discriminating between abnormal and normal heart sounds. *Irbm*, 41(4):223–228, 2020.

[42] Bin Xiao, Yunqiu Xu, Xiuli Bi, Junhui Zhang, and Xu Ma. Heart sounds classification using a novel 1-d convolutional neural network with extremely low parameter consumption. *Neurocomputing*, 392:153–159, 2020.

[43] Tao Li, Yibo Yin, Kainan Ma, Sitao Zhang, and Ming Liu. Lightweight end-to-end neural network model for automatic heart sound classification. *Information*, 12(2):54, 2021.

[44] Guangyang Tian, Cheng Lian, and Zhigang Zeng. Integrated res2net combined with seesaw loss for long-tailed pcg signal classification. In *2021 11th International Conference on Intelligent Control and Information Processing (ICICIP)*, pages 53–58. IEEE, 2021.

[45] Na Mei, Hongxia Wang, Yatao Zhang, Feifei Liu, Xinge Jiang, and Shoushui Wei. Classification of heart sounds based on quality assessment and wavelet scattering transform. *Computers in Biology and Medicine*, 137:104814, 2021.

[46] Rima Touahria, Abdenour Hacine-Gharbi, and Philippe Ravier. Discrete wavelet based features for pcg signal classification using hidden markov models. In *ICPRAM*, pages 334–340, 2021.

[47] Nidhi Kalidas Sawant, Shivnarayan Patidar, Naimahmed Nesaragi, and U Rajendra Acharya. Automated detection of abnormal heart sound signals using fano-factor constrained tunable quality wavelet transform. *Biocybernetics and Biomedical Engineering*, 41(1):111–126, 2021.

[48] Vinay Arora, Karun Verma, Rohan Singh Leekha, Kyungroul Lee, Chang Choi, Takshi Gupta, and Kashish Bhatia. Transfer learning model to indicate heart health status using phonocardiogram. *Computers, Materials and Continua*, 2021.

[49] Sinam Ajitkumar Singh, Sinam Ashinikumar Singh, Ningthoujam Dinita Devi, and Swanirbhar Majumder. Heart abnormality classification using pcg and ecg recordings. *Computación y Sistemas*, 25(2):381–391, 2021.

[50] Miguel A Alonso-Arévalo, Alejandro Cruz-Gutiérrez, Roilhi F Ibarra-Hernández, Eloísa García-Canseco, and Roberto Conte-Galván. Robust heart sound segmentation based on spectral change detection and genetic algorithms. *Biomedical Signal Processing and Control*, 63:102208, 2021.

[51] Umair Riaz, Sumair Aziz, Muhammad Umar Khan, Syed Azhar Ali Zaidi, Muhammad Ukasha, and Aamir Rashid. A novel embedded system design for the detection and classification of cardiac disorders. *Computational Intelligence*, 37(4):1844–1864, 2021.

[52] Philip Gemke, Nicolai Spicher, and Tim Kacprowski. An lstm-based listener for early detection of heart disease. In *2022 Computing in Cardiology (CinC)*, volume 498, pages 1–4. IEEE, 2022.

[53] Adyasha Rath, Debahuti Mishra, Ganapati Panda, and Madhumita Pal. Development and assessment of machine learning based heart disease detection using imbalanced heart sound signal. *Biomedical Signal Processing and Control*, 76:103730, 2022.

[54] Muhammad Umar Khan, Sumair Aziz, Khushbakht Iqtidar, Galila Faisal Zaher, Shareefa Alghamdi, and Munazza Gull. A two-stage classification model integrating feature fusion for coronary artery disease detection and classification. *Multimedia Tools and Applications*, 81(10):13661–13690, 2022.

[55] Yashwanth Gadde and T Kishore Kumar. Prediction of heart abnormality using heart sound signals. In *Machine Intelligence Techniques for Data Analysis and Signal Processing: Proceedings of the 4th International Conference MISP 2022, Volume 1*, pages 669–680. Springer, 2023.

[56] Arnab Maity, Akanksha Pathak, and Goutam Saha. Transfer learning based heart valve disease classification from phonocardiogram signal. *Biomedical Signal Processing and Control*, 85:104805, 2023.

[57] Anandita Bhardwaj, Sandeep Singh, and Deepak Joshi. Explainable deep convolutional neural network for valvular heart diseases classification using pcg signals. *IEEE Transactions on Instrumentation and Measurement*, 2023.

[58] Xueyi Li, Jialin Li, Chengying Zhao, Yongzhi Qu, and David He. Gear pitting fault diagnosis with mixed operating conditions based on adaptive 1d separable convolution with residual connection. *Mechanical Systems and Signal Processing*, 142:106740, 2020.

[59] Christian Garbin, Xingquan Zhu, and Oge Marques. Dropout vs. batch normalization: an empirical study of their impact to deep learning. *Multimedia Tools and Applications*, 79(19):12777–12815, 2020.

[60] Hidenori Ide and Takio Kurita. Improvement of learning for cnn with relu activation by sparse regularization. In *2017 international joint conference on neural networks (IJCNN)*, pages 2684–2691. IEEE, 2017.

[61] Matías Roodschild, Jorge Gotay Sardiñas, and Adrián Will. A new approach for the vanishing gradient problem on sigmoid activation. *Progress in Artificial Intelligence*, 9(4):351–360, 2020.

[62] Juntang Zhuang, Tommy Tang, Yifan Ding, Sekhar C Tatikonda, Nicha Dvornek, Xenophon Papademetris, and James Duncan. Adabelief optimizer: Adapting stepsizes by the belief in observed gradients. *Advances in neural information processing systems*, 33:18795–18806, 2020.

[63] Chengyu Liu, David Springer, Qiao Li, Benjamin Moody, Ricardo Abad Juan, Francisco J Chorro, Francisco Castells, José Millet Roig, Ikaro Silva, Alistair EW Johnson, et al. An open access database for the evaluation of heart sound algorithms. *Physiological measurement*, 37(12):2181, 2016.

[64] Cristhian Potes, Saman Parvaneh, Asif Rahman, and Bryan Conroy. Ensemble of feature-based and deep learning-based classifiers for detection of abnormal heart sounds. In *2016 computing in cardiology conference (CinC)*, pages 621–624. IEEE, 2016.

[65] Morteza Zabihi, Ali Bahrami Rad, Serkan Kiranyaz, Moncef Gabbouj, and Aggelos K Katsaggelos. Heart sound anomaly and quality detection using ensemble of neural networks without segmentation. In *2016 computing in cardiology conference (CinC)*, pages 613–616. IEEE, 2016.

[66] Edmund Kay and Anurag Agarwal. Dropconnected neural network trained with diverse features for classifying heart sounds. In *2016 Computing in Cardiology Conference (CinC)*, pages 617–620. IEEE, 2016.

[67] Ignacio J Diaz Bobillo. A tensor approach to heart sound classification. In *2016 Computing in Cardiology Conference (CinC)*, pages 629–632. IEEE, 2016.

[68] Masun Nabhan Homsi, Natasha Medina, Miguel Hernandez, Natacha Quintero, Gilberto Perpiñan, Andrea Quintana, and Philip Warrick. Automatic heart sound recording classification using a nested set of ensemble algorithms. In *2016 Computing in Cardiology Conference (CinC)*, pages 817–820. IEEE, 2016.

[69] Vykintas Maknickas and Algirdas Maknickas. Recognition of normal–abnormal phonocardiographic signals using deep convolutional neural networks and mel-frequency spectral coefficients. *Physiological measurement*, 38(8):1671, 2017.

[70] Filip Plesinger, Juraj Jurco, Pavel Jurak, and Josef Halamek. Discrimination of normal and abnormal heart sounds using probability assessment. In *2016 Computing in Cardiology Conference (CinC)*, pages 801–804. IEEE, 2016.

[71] Jonathan Rubin, Rui Abreu, Anurag Ganguli, Saigopal Nelaturi, Ion Matei, and Kumar Sricharan. Classifying heart sound recordings using deep convolutional neural networks and mel-frequency cepstral coefficients. In *2016 Computing in cardiology conference (CinC)*, pages 813–816. IEEE, 2016.

[72] Hong Tang, Huaming Chen, Ting Li, and Mingjun Zhong. Classification of normal/abnormal heart sound recordings based on multi-domain features and back propagation neural network. In *2016 Computing in Cardiology Conference (CinC)*, pages 593–596. IEEE, 2016.

[73] Mostafa Abdollahpur, Shadi Ghiasi, Mohammad Javad Mollakazemi, and Ali Ghaffari. Cycle selection and neuro-voting system for classifying heart sound recordings. In *2016 Computing in Cardiology Conference (CinC)*, pages 1–4. IEEE, 2016.

[74] Bradley M Whitaker, Pradyumna B Suresha, Chengyu Liu, Gari D Clifford, and David V Anderson. Combining sparse coding and time-domain features for heart sound classification. *Physiological measurement*, 38(8):1701, 2017.

[75] Wei Han, Zuyuan Yang, Jun Lu, and Shengli Xie. Supervised threshold-based heart sound classification algorithm. *Physiological Measurement*, 39(11):115011, 2018.

[76] Vinay Arora, Rohan Leekha, Raman Singh, and Inderveer Chana. Heart sound classification using machine learning and phonocardiogram. *Modern Physics Letters B*, 33(26):1950321, 2019.

[77] Suyi Li, Feng Li, Shijie Tang, and Fan Luo. Heart sounds classification based on feature fusion using lightweight neural networks. *IEEE Transactions on Instrumentation and Measurement*, 70:1–9, 2021.

[78] Moukadem A and Dieterlen A. Classification of heart sound recordings - the physionet computing in cardiology challenge 2016. https://physionet.org/content/challenge-2016/1.0.0/, 2016. Online; accessed 1 Novamber 2022.