

Article

Prediagnosis of Heart Failure (HF) Using Deep Learning and the Korotkoff Sound

Huanyu Zhang ¹, Ruwei Wang ¹, Hong Zhou ^{1,*}, Shudong Xia ^{2,*}, Sixiang Jia ² and Yiteng Wu ²¹ College of Biomedical Engineering & Instrument Science, Zhejiang University, Hangzhou 310013, China² The Fourth Affiliated Hospital of Zhejiang University School of Medicine, Yiwu 322000, China

* Correspondence: zhouhong_zju@126.com (H.Z.); shystone@zju.edu.cn (S.X.);

Tel.: +86-0571-8795-1833 (H.Z.)

Abstract: Heart failure (HF) is a devastating condition that impairs people's lives and health. Because of the high morbidity and mortality associated with HF, early detection is becoming increasingly critical. Many studies have focused on the field of heart disease diagnosis based on heart sound (HS), demonstrating the feasibility of sound signals in heart disease diagnosis. In this paper, we propose a non-invasive early diagnosis method for HF based on a deep learning (DL) network and the Korotkoff sound (KS). The accuracy of the KS-based HF prediagnosis was investigated utilizing continuous wavelet transform (CWT) features, Mel frequency cepstrum coefficient (MFCC) features, and signal segmentation. Fivefold cross-validation was applied to the four DL models: AlexNet, VGG19, ResNet50, and Xception, and the performance of each model was evaluated using accuracy (Acc), specificity (Sp), sensitivity (Se), area under curve (AUC), and time consumption (Tc). The results reveal that the performance of the four models on MFCC datasets is significantly improved when compared to CWT datasets, and each model performed considerably better on the non-segmented dataset than on the segmented dataset, indicating that KS signal segmentation and feature extraction had a significant impact on the KS-based CHF prediagnosis performance. Our method eventually achieves the prediagnosis results of Acc (96.0%), Se (97.5%), and Sp (93.8%) based on a comparative study of the model and the data set. The research demonstrates that the KS-based prediagnosis method proposed in this paper could accomplish accurate HF prediagnosis, which will offer new research approaches and a more convenient way to achieve early HF prevention.

Keywords: heart failure; Korotkoff sound; deep learning; MFCC; neural networks; prediagnosis

Citation: Zhang, H.; Wang, R.; Zhou, H.; Xia, S.; Jia, S.; Wu, Y. Prediagnosis of Heart Failure (HF) Using Deep Learning and the Korotkoff Sound. *Appl. Sci.* **2022**, *12*, 10322. <https://doi.org/10.3390/app122010322>

Academic Editors: Jiann-Der Lee and Jong-Chih Chien

Received: 7 September 2022

Accepted: 10 October 2022

Published: 13 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

According to the World Health Organization's 2022 statistics report [1], 33.2 million people worldwide died of non-communicable diseases in 2019, which was a 28% increase from 2000, and women in more than half of all countries and men in 3/4 of all countries face premature death from cardiovascular disease [2]. Although coronary computed tomography (CT) [3] and echocardiography [4] can be used to diagnose potential heart disease patients, these tests are time-consuming, costly, and require specialized testing equipment. More frustrating, the disruption in medical services caused by the COVID-19 pandemic may expose existing patients to increased risks of serious illness and death [5]. Given the dire situation, we urgently need to develop low-cost heart disease prediagnosis equipment and increase the popularity of early non-destructive heart disease testing, which will greatly promote heart disease diagnosis, treatment, and prevention while reducing the harm caused by heart disease, particularly heart failure (HF). In this paper, we attempt to develop a Korotkoff sound (KS)-based HF prediagnosis method based on deep learning (DL) by investigating the signal preprocessing and feature extraction techniques of KS data gathered.

As a long-term stable and accurate blood pressure monitoring signal, KS has been widely employed in the clinical diagnosis of hypertension and the daily monitoring of potential patients [6,7]. Researchers have been making continuous efforts in the field of KS-based disease detection, and the QKD interval [8], the ankle-brachial index (ABI) [9], and other indexes have been proposed and used to diagnose cardiovascular diseases. The QKD interval is the period between the beginning of depolarization on the electrocardiogram (Q) and the detection of the last Korotkoff sound (K) at the level of the brachial artery during cuff deflation, which corresponds to diastolic blood pressure (D). Although these approaches have limits in terms of accuracy, sensitivity, and specificity, they have proven successful in the early detection of cardiovascular disorders as a method of early prediagnosis. With the advancement of digital signal processing, big data, and artificial intelligence technology, we now have a more powerful tool to re-examine the role of KS signals in disease prediagnosis. To date, a large number of machine learning (ML) algorithms have been used to classify non-stationary time series signals [10], with impressive results. However, traditional ML approaches typically involve a large number of time-consuming preprocessing tasks, such as noise reduction [11,12], segmentation [13,14], and feature extraction [15]. Furthermore, the ML algorithm is heavily reliant on the accuracy of feature selection, and different feature sets have a significant impact on classification results. Deep convolutional neural (CNN) network avoids many of the problems associated with traditional ML, solves the problem of signal feature selection, and greatly improves the classifier's classification accuracy [16]. The CNN algorithm has become increasingly significant in processing non-stationary time series data as people's understanding of the CNN network has deepened [17,18].

In recent decades, heart sound (HS) [19,20], electrocardiograms (ECG) [21], and cardiac interbeat interval (RR interval) [22] have gradually attracted people's attention due to the benefits of convenience, non-invasion, and simple promotion [23]. Many scholars have developed a variety of ML and deep-learning computer-aided diagnosis systems for heart disease based on these signals [24,25]. These systems assist researchers in identifying key factors from large amounts of tedious data, reducing the burden on doctors and medical staff significantly. However, the popularity of the heart disease prediagnosis system in communities and families remains inadequate. According to some surveys, most people do not prioritize the early prevention of heart disease due to a lack of knowledge and attention [26].

A common method of blood pressure monitoring, the KS-based blood pressure test is widely recognized and expertly performed. KS is a group of short and rapid pulse sounds produced by brachial artery blood flow during the process of overcoming cuff pressure, which is closely related to brachial artery vibration and blood turbulence [27]. KS is only a pulse signal in comparison to heart sound, but its transient characteristics are obvious, and signal strength and characteristics are highly related to time and cuff pressure, posing some challenges to KS signal processing and analysis. However, if heart disease prediagnosis can be completed concurrently with blood pressure monitoring, it will improve many problems in current heart disease prevention and increase the likelihood of early detection of heart disease.

The primary objective of this study is to test whether we can develop a simple, non-invasive heart disease prediagnosis system for ordinary people to achieve a broader range of community and family promotion. Therefore, we reintroduce this traditional, minor source of concern: KS. It has a wide range of applications and a high penetration rate, and most people are very familiar with the blood pressure detection process, which simplifies KS signal acquisition and increases the popularity of related detection systems.

This paper proposes an innovative method for the prediagnosis of patients with HF using a DL network and KS. We extracted the continuous wavelet transform (CWT) and Mel Frequency Cepstrum Coefficient (MFCC) features of KS and sent the two feature maps to four common DL networks for training and classification, including AlexNet, VGG-19, ResNet50, and Xception. To confirm the best analysis model, the performance of

each model is evaluated using accuracy (Acc), specificity (Sp), sensitivity (Se), AUC criteria, and time consumption (Tc). The contributions in this paper may be summarized as follows:

- An innovative method for the prediagnosis of HF based on KS and DL networks has been proposed, analyzed, and validated in four types of DL networks.
- The impact of signal segmentation and feature extraction methods on KS-based HF prediagnosis is thoroughly investigated in this paper.
- The pre-trained neural network model is transferred to the KS-based HF prediagnosis task, which improved the model's training efficiency and ensured the reliability of this paper's conclusion.

The structure of this paper is as follows: Section 2 discusses related work in cardiovascular disease research; Section 3 describes the analysis method and materials used in this article; Section 4 describes the performance of our method in the prediagnosis of HF; Section 5 delves deeply into the experimental process and outcomes; and Section 6 presents the conclusions and future work.

2. Related Works

Unsteady time series signal noise reduction has long been a hot research topic, Mondal An et al. [28] proposed and demonstrated a reliable and robust HS denoising method based on wavelet packet transform (WPT) and SVD. Deng S W et al. [29] proposed an adaptive denoising algorithm that does not require a prior definition of the basis function, outperforming the traditional wavelet method in low noise levels. Lopac N et al. [30] proposed a noisy non-stationary time series signal classification method that combined deep CNN architectures with Cohen's class time-frequency representations (TFRs), resulting in higher performance index values and better classification performance.

With the continuous advancement of signal processing technology, ML and DL algorithms [31] for early diagnosis of heart disease have been developed in-depth. Noman F et al. [32] proposed the Markov-switching autoregressive with switched linear dynamic systems (MSAR-SLDS) algorithm, which achieved a classification accuracy of 86.1%. Nogueira D M et al. [33] improved the automatic classification of heart disease by combining time-domain and frequency-domain features, and achieved an accuracy of about 83.22%. Potes C et al. [34] used the AdaBoost-abstain classifier and CNN method to achieve 86.02% classification accuracy in PhysioNet/Computing in Computing Challenge 2016. Rubin J et al. [35] used image deep neural networks to classify HS and achieved an 84% classification accuracy. Using only raw data as input, Zhang W et al. [36] proposed a novel method for detecting abnormal HS without segmentation and achieved an analysis accuracy of 94.84%. He Y et al. [37] used the U-NET network to segment and classify HS, achieving a classification accuracy of 99.1%. Gjoreski M et al. [38] used a combination of machine learning and DL approaches to achieve a classification accuracy of 92.9% for HF patients.

Machine learning methods have received little attention in the field of KS research. The majority of the work is devoted to traditional statistical analysis techniques. Gosse P et al. [39] invented a device based on QKD interval parameters to investigate the effect of blood pressure on arteriosclerosis and discovered a significant correlation between QKD slope and left ventricular mass. Constans J et al. [40] used univariate and multivariate analyses to demonstrate the efficacy of QKD in diagnosing atherosclerosis. Gosse P et al. [41] discovered that the equivalent KS interval (QKDh) was important in predicting cardiovascular complications and major events. El Tahlawi M et al. [42] used statistical analysis to show a significant positive correlation between oscillatory gap (OG) and coronary artery disease. Yamamoto T et al. [43] believed that the severity of coronary atherosclerosis was related to the difference in systolic blood pressure. Jia S et al. [44] demonstrated that there is a certain proportional relationship between brachial artery blood flow and cardiac output (about 1.23%), which provides a good foundation for our future work.

3. Materials and Methods

3.1. System Overview

The overall structure of this paper is depicted in Figure 1. The KS data were collected by the medical team of the Department of Cardiology, the fourth people's Hospital connected with Zhejiang University, using a 3M Littmann 3200 electronic stethoscope with a sampling frequency of 480 kHz. After data cleaning, all KS signals were downsampled to 4 kHz. We denoised the downsampled KS signals using the Wiener filtering approach to remove friction noise, ambient noise, electromagnetic noise, and human breath sounds.

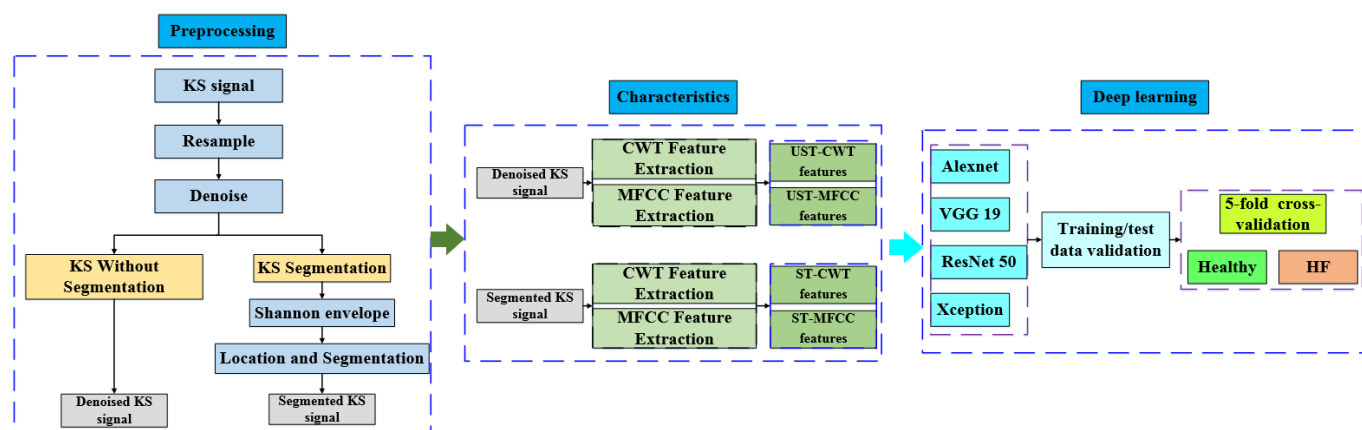


Figure 1. The overall structure of this paper.

The signal segmentation work refers to the heart sound segmentation algorithm combined with the Shannon energy envelope algorithm, signal locating, and segmentation method. KS locationing, unlike heart sounds, does not necessitate a fine distinction of effective signals; however, there will be over-localization during the positioning process, and the positioning results must be re-validated.

To compare the effect of KS signal segmentation on HF prediagnosis, we extracted CWT and MFCC features from segmented and unsegmented KS signals. The following datasets were obtained: the segmented-based CWT (ST-CWT) data set, the segmented-based MFCC (ST-MFCC) data set, the unsegmented-based CWT (UST-CWT) data set, and the unsegmented-based MFCC (UST-MFCC) data set. In CWT feature extraction, the Morse mother wavelet was employed. When performing MFCC feature extraction, 26 Mel filters were used, the frame length was set to 512, and the frame shift was set to 256; the obtained CWT and MFCC coefficients were converted into 224×224 images. After image augmentation, they were sent to the convolutional neural network for classification and performance evaluation.

3.2. Database Acquisition

This work was completed by the medical team of the Department of Cardiology, the fourth people's Hospital affiliated with Zhejiang University. All volunteers' heart conditions were evaluated in accordance with the ESC 2021 Guidelines [45]. Only HF patients with a left ventricular ejection fraction (LVEF) of less than 50% and healthy people with no cardiovascular disease or other underlying diseases were chosen as volunteers for this study. Each volunteer was thoroughly examined by experienced cardiologists prior to the test to ensure the sample's reliability.

As shown in Table 1, a total of 365 samples were collected, including 116 healthy people and 249 HF patients. The average age of the healthy subjects was 42 ± 22 years old, the average body mass index (BMI) was 24.8 ± 2.1 , the average LVEF was $65 \pm 7.5\%$, the average systolic blood pressure (SBP) and diastolic blood pressure (DBP) were 111 ± 15 and 73 ± 10 mm Hg, respectively; The patients' average age was 61 ± 17 , their average BMI

was 24.5 ± 3.2 , their average LVEF was $40 \pm 6.83\%$, and their average SBP and DBP were 129 ± 35 and 77 ± 20 mm Hg, respectively. NT-proBNP levels in selected patients ranged from 513 to 11,950 pg/mL, while NT-proBNP levels in healthy people were less than 125 pg/mL.

Table 1. Subjects' information.

Subjects	Female Ratio %	Ages Years	BMI Kg/m ²	LVEF %	SBP mm Hg	DBP mm Hg	NT-proBNP pg/mL	Numbers -
Healthy	50	42 \pm 22	24.8 \pm 2.1	65 \pm 7.5	111 \pm 15	73 \pm 10	513~11950	116
HF patients	40	61 \pm 17	24.5 \pm 3.2	40 \pm 6.83	129 \pm 35	77 \pm 20	<125	249

Note: SBP-systolic blood pressure, DBP-diastolic blood pressure, BMI-body mass index, LVEF- left ventricular ejection fraction, NT-proBNP-N-terminal pro-B-type natriuretic peptide.

When volunteers arrived at the echo lab, we advised them to sit for 10 min prior to the test in order to be as emotionally stable as possible, and then we took measurements in a temperature-controlled room (23–25 °C) [46]. The overall left ventricular systolic function of each patient was evaluated by two professional cardiac ultrasound technicians using a PHILIPS ultrasound instrument based on the patient's clinical status and established clinical guidelines [47]. We measured the left ventricular volume using the biplane method (improved Simpson) [48]. The corresponding LVEF and cardiac output was generated automatically by PHILIPS ultrasound machines.

The audio collection equipment was a 3M Littman 3200 electronic stethoscope (480 kHz effective sampling frequency), which has a powerful environmental noise suppression capability that ensures the capture of essential sounds while reducing background noise by 85%. As shown in Figure 2, the stethoscope was placed directly above the brachial artery during the test to eliminate the influence of other interference factors on the test results [49]. The stethoscope probe was placed about 5 cm from the lower edge of the cuff. According to the blood pressure detection procedure, the cuff was pressurized to 180 mm Hg, and the measurement began with the cuff step-down until the last KS pulse disappeared. The test duration ranged from 14 to 25 s. To ensure the consistency of the test results, each volunteer was measured three times. After the test, the data was transferred to the workstation via mobile storage media or Bluetooth for analysis and processing.

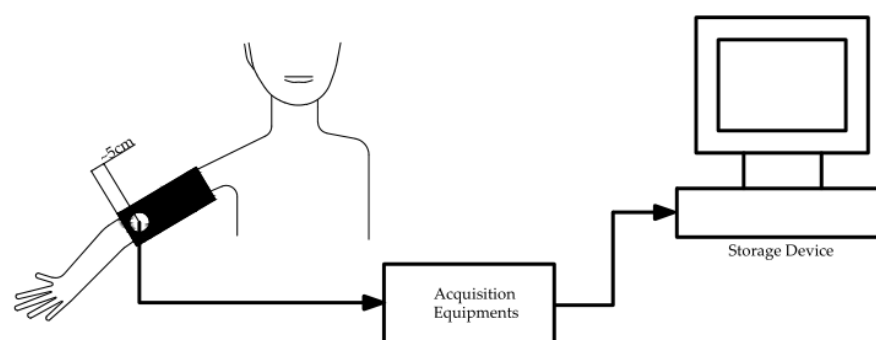


Figure 2. KS data acquisition process diagram.

It should be noted that the current study described in this paper only collected information on the hospitalization status of consecutively diagnosed patients with heart failure; the patient's course and the impact of pharmacological therapy on the prediagnosis are not considered. Furthermore, the impact of diabetes, hypertension, coronary heart disease, renal failure, anemia, and other comorbidities of HF patients on HF pre-diagnosis results, as well as the impact of separate body weight and breast analysis on KS records

based on BMI and gender, are very important. So far, we have not paid attention to these issues, which is a limitation of our current dataset.

All subjects signed the informed consent form after fully understanding the issues at hand in this test. The relevant work was approved by Zhejiang University's Ethics Committee.

3.3. Data Processing

To investigate the impact of signal segmentation on classification results, the features of segmented and non-segmented signals are extracted. The HS segmentation method is used for KS segmentation, but unlike HS, KS is a group of pulse signals, there are no obvious S1 or S2 characteristics, and KS has a period of silence during the acquisition process. In signal segmentation, we focus on the characteristics of KS signals in the audible range of the human ear, according to the basic principle of KS blood pressure detection [50]. We assumed that the silence period had no effect on HF prediagnosis and thus was considered an invalid sound in the KS segmentation. For non-segmented KS signals, only noise reduction and resampling were used to retain as much information as possible from the original data.

3.3.1. Wiener Filtering

The primary objective of the Wiener filter is to reduce the error between the signal's true value and the predicted result. The Wiener filter's basic operation is shown in Figure 3. The previous signal is 0.025 times the length of the KS signal, the window length is set to 80 ms, and the shift percentage is set to 60%.

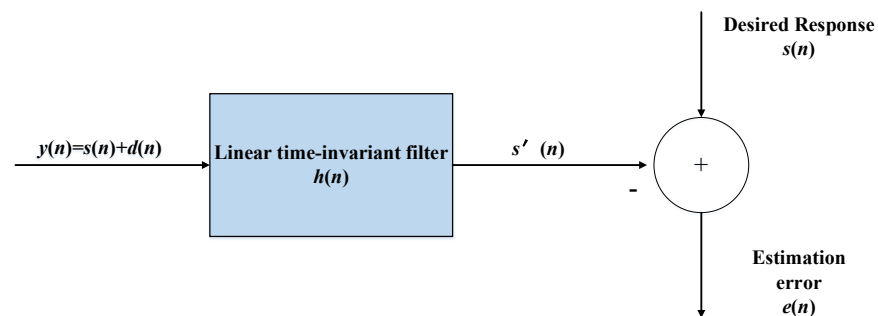


Figure 3. Flow chart of Wiener filtering.

Where $y(n)$ is the input signal, $s'(n)$ is the output signal, $s(n)$ is a pure speech signal, and $d(n)$ is a noise signal, as in the figure above, then the filter formula will be as follows:

$$s'(n) = y(n) \cdot h_n = \sum_{m=1}^{M-1} y(n-m) \cdot h_m \quad (1)$$

where, h_m is the filter coefficient and M is the number of filter taps. The system's filter coefficient can be calculated by minimizing the error:

$$e(n) = s(n) - s'(n) \quad (2)$$

3.3.2. Shannon Envelope

Compared with other segmentation techniques, the envelope method offers the benefits of convenient operation and high resilience. The Shannon energy envelope calculation technique has been widely employed in HS research, due to its improved sensitivity and specificity [51]. The second-order Shannon envelope formula is as follows [52]:

$$E_j = -\frac{1}{N} \sum_{i=1}^N Z_i^2 \times \log Z_i^2 \quad (3)$$

Where, N is the moving window length in samples, and Z_i is the normalized segmented signal, as given in Formula (4):

$$Z_i = x_i / \max |x_i| \quad (4)$$

The length of the segmented signal is set to 40 ms in this computation, and the length of the overlapping window is set to 50%. As shown in Formula (5), the normalizing procedure is utilized to reduce the detection's reliance on signal quality. After normalization is complete, all negative Shannon energy values are adjusted to 0, yielding a non-negative envelope signal. Figure 4B depicts the Shannon energy envelope effect diagram.

$$P_j = \frac{E_j - \text{mean}(E_j)}{\text{std}(E_j)} \quad (5)$$

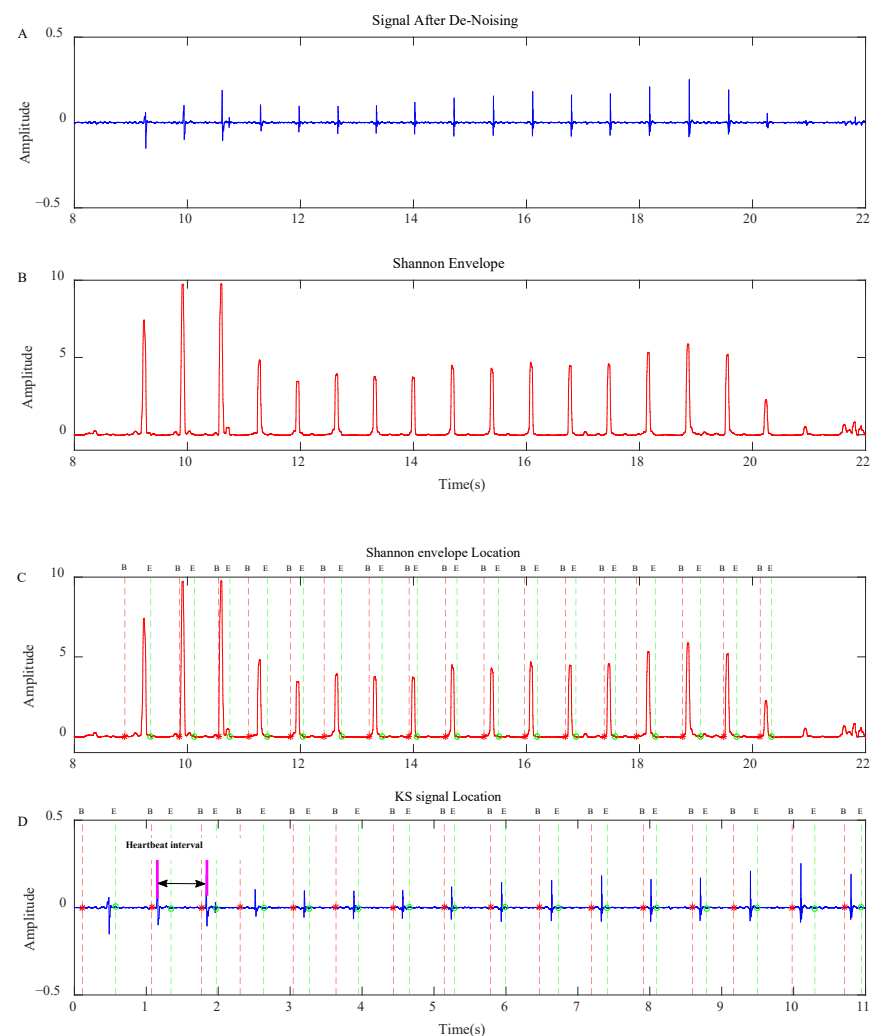


Figure 4. KS segmentation and location results: (A) de-noised KS signal; (B) Shannon envelope of the KS signal; (C) segmentation results of Shannon envelope; (D) segmentation results for KS. NOTE: The blue curve represents the KS signal's time domain signal, while the red curve represents the KS signal's Shannon envelope. B represents the start of KS, which is the same as the * added to the line, and E represents the end of KS, which is the same as the ° added to the line.

3.3.3. Location and Segmentation

Using the peak value as a benchmark, we determined the effective signal in KS segmentation; however, there are usually several peaks in a KS pulse signal, which may result in over-location. Furthermore, noise in the measurement procedure impedes the identification of effective signals. To reduce over-location and interference signals, anchor points with intervals of less than 100 ms and greater than 2000 ms were eliminated. The calibrated location point can be used for KS pulse signal segmentation and reorganization. Figure 4C,D show the results of KS signal localization and segmentation.

3.4. Feature Graph Generation

Because most CNN network architecture only supports two-dimensional signal input, we must convert one-dimensional raw data into two-dimensional image data. In this paper, CWT time-frequency analysis technology and MFCC feature extraction technology are used to transform raw data into 2D heat map.

3.4.1. CWT Time-Frequency Characteristics

CWT is an effective method for analyzing the frequency-time relationship [53,54]. A wavelet time-frequency scaling map can help to identify the signal's low-frequency characteristics and instantaneous components. To obtain the CWT coefficient [55], we constructed a CWT filter with the same parameters for each KS signal; the CWT filter banks use analytic Morse (3, 60) wavelets, and 48 wavelet bandpass filters are set in each octave. L1 standardization will be used in the CWT calculation to ensure that the constant amplitude oscillation components at different scales have the same intensity in the continuous wavelet transform.

After the wavelet transform is completed, each set of KS will generate a wavelet coefficient matrix of approximately $(600 \times 42,000)$. A set of standardized matrices was obtained by performing amplitude extraction on the wavelet coefficient matrix and scaling it to $(0, 1)$; the normalized matrix was then converted into an uint8 matrix between $(0-255)$. Based on this uint8 matrix, an image matrix in RGB format is generated and saved as a 224×224 image.

3.4.2. Mel Frequency Cepstrum Coefficient

Mel frequency is more suitable for simulating human auditory behavior than other audio-processing methods, and it is widely used in the field of human voice (speech recognition, speaker recognition) [56]. The following steps are involved in obtaining the Mel frequency cepstrum coefficient: pre-emphasis, framing, windowing, fast Fourier transform, Mel filter banks, and discrete cosine transform [33]. The formula for converting an ordinary frequency to a Mel frequency is as follows [57]:

$$mel(f) = 2595 \times \log_{10}(1 + f / 700) \quad (6)$$

The discrete cosine transform is used to convert the logarithmic Mel spectrum back to the time domain, and the result is known as MFCC, as shown in formula 7.

$$MFCC(n_1) = \sum_{m=1}^M \log[E(m, k)] \cdot \cos\left(\frac{(m-0.5)}{M} \cdot n_1 \cdot \pi\right), \quad n_1 = 1, 2, \dots, L \quad (7)$$

where, M is the number of filters, n_1 is the order, and $E(m, k)$ is the average energy of the k th frequency band.

The MFCC feature of the KS signal is extracted using 26 Mel filters, the frame length is set to 256, and the frame shift is set to 128; the signal is framed using the Hamming window function $W(n)$ to avoid sidelobe leakage [58]. The calculated MFCC coefficient matrix (167×26) is converted to the RGB format matrix and saved as an image of 224×224 using the same processing method as the CWT coefficient matrix.

3.5. Deep Learning (DL) Network

Hinton et al. [59] proposed an unsupervised greedy layer-by-layer training method and a supervised back-propagation algorithm based on deep belief networks (DBN) to solve the problem of gradient disappearance. The accuracy and speed of DL algorithms were consequently greatly improved by AlexNet, which was proposed by Krizhevsky et al. [60]. Their work has greatly increased academic interest in deep neural networks, and many deep neural network algorithms have been proposed. In this paper, some commonly used neural network structures such as AlexNet, VGG19, ResNet50, and Xception are introduced to explore the application of DL algorithms in KS-based HF prediagnosis.

3.5.1. AlexNet

AlexNet is a deep neural network structure that includes a convolution layer, a non-linear layer, a pooling layer, and a dropout layer [60]. The addition of nonlinear layers, pooling layers, and dropout layers maximizes the accuracy and generalization ability of the convolution network. The AlexNet network structure is depicted in Figure 5a below.

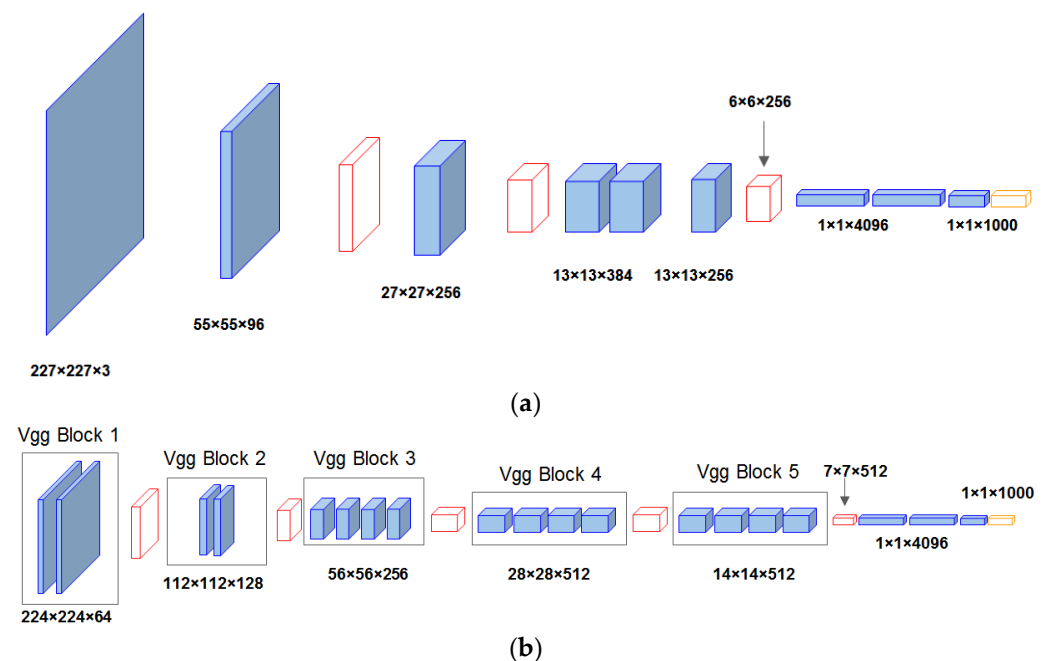


Figure 5. Schematic diagram of (a) AlexNet and (b) VGG19 architecture.

3.5.2. VGG19

As shown in Figure 5b, the VGG19 network is made up of 16 convolution layers and 3 full connection layers. The 3×3 matrix is widely used in the network to replace the 11×11 matrix in AlexNet to reduce the computing scale caused by network depth increase. The VGG19 network increases the depth of the network, and the networks of each layer are stacked on top of each other, resulting in more calculation parameters and a slower calculation speed. However, as the first scheme to expand the scale of CNN, VGG19 can achieve better accuracy in most tasks, and it is a network structure that has been widely referenced [61].

3.5.3. ResNet

With the deepening of the network structure, it was discovered that the simple superposition of the convolution layer and pooled layer cannot effectively improve network performance, but instead led to network degradation and other problems. The residual network proposed by Kaiming et al. [62] solved this problem and greatly encouraged the

use of DL networks. Resnet reduces the number of network parameters by introducing skip or residual connections and allows for further network expansion by simply repeating VGG blocks. The Resnet50 network will be used in this article, and its structure is depicted in Figure 6 below.

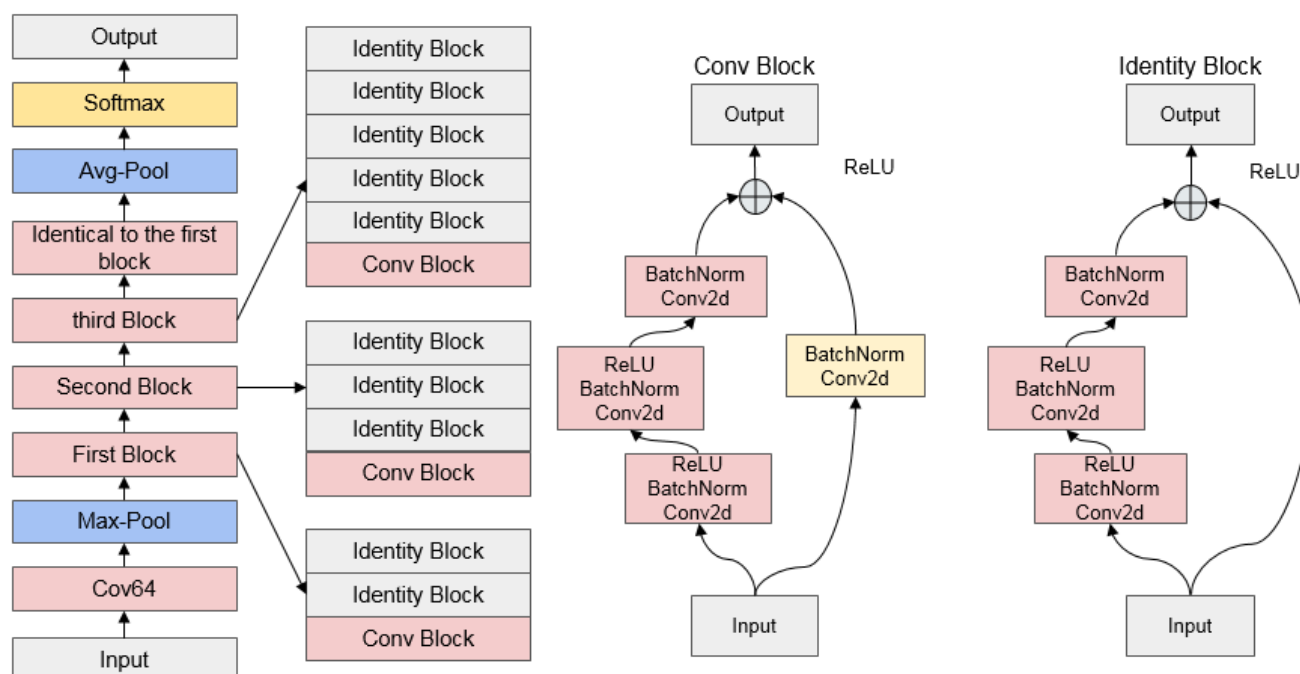


Figure 6. Schematic diagram of ResNet50 architecture.

3.5.4. Xception

Google's Xception is yet another improvement to Inception-v3. As shown in Figure 7, it employs a 36-layer convolution neural network structure, with the core idea being to use depth separable convolution to reduce network parameters, allowing it to achieve relatively high network accuracy with fewer parameters [63]. Depth separable convolution includes spatial convolution (3×3) and dot convolution (1×1). Simultaneously, Xception introduces the residual network structure to improve the network's editability and efficiency.

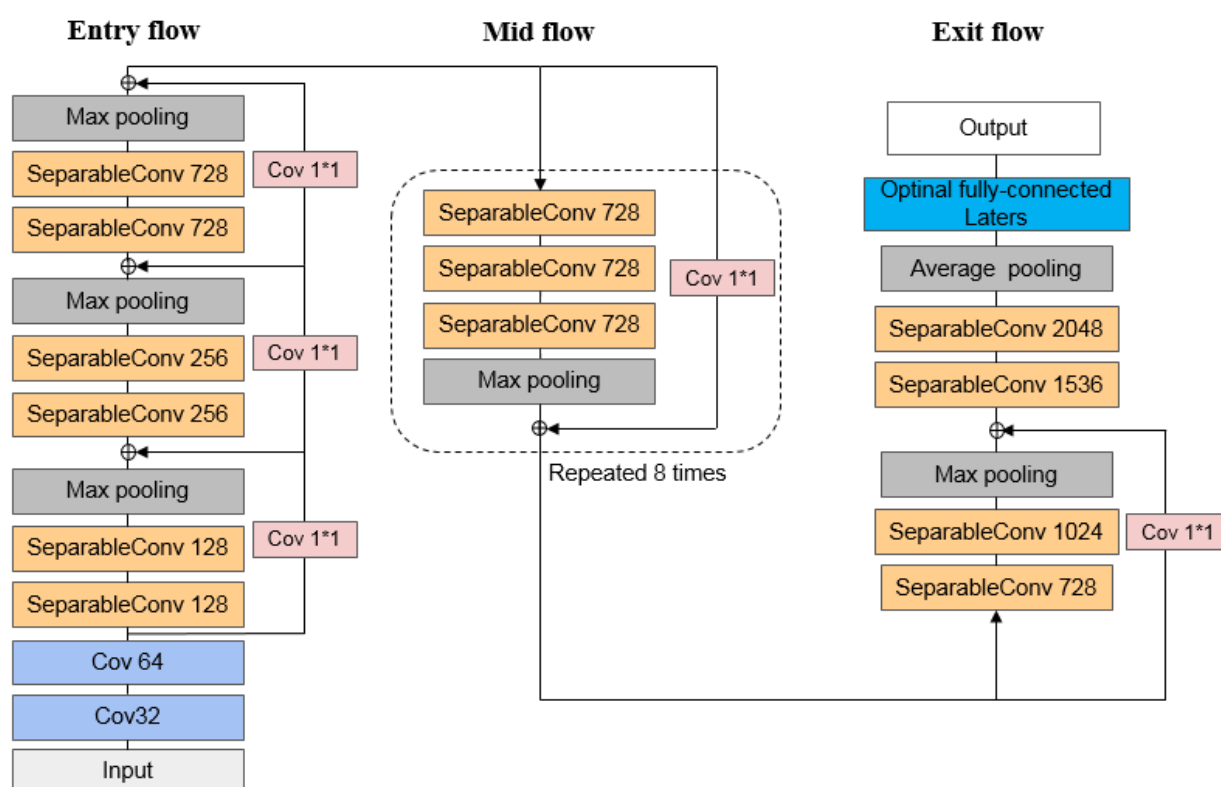


Figure 7. Schematic diagram of Xception architecture.

3.5.5. Data Augmentation

Deep neural networks are data-greedy learning algorithms, which means that the network's training effect is proportional to the size of the data. Data augmentation technology is an important technical support for neural network algorithms, as it can increase the number of samples and prevent the algorithm from overfitting [64]. In our experiments, we applied the following augmentation strategies to the original feature image.

Flipping: The patient's and healthy person's feature images are flipped 180 degrees horizontally.

Add Gaussian noise: The feature maps of patients and healthy people are augmented with Gaussian noise. The variance of Gaussian noise is set at 0.02 and the mean value is set at 0.2.

Contrast enhancement: During the patient's feature map processing, we generated enhanced images by linear mapping of color values greater than 0.3×255 . In the feature image processing of healthy people, we performed three types of contrast enhancement: the first is to set the color values of less than 0.3×255 to 0, and the enhanced image is generated by linear mapping; the second is to generate the image by nonlinear mapping of the original image, with the nonlinear coefficient $\gamma = 0.6$. The third step is to non-linearly map the 0.3×255 –255 color value of the original image into the 0.3×255 –255 color image, $\gamma = 0.9$.

After data augmentation, the KS image feature set of HF patients increased threefold, from 249 to 996; and the image feature set of healthy people increased fivefold, from 116 to 696.

3.5.6. Transfer Learning

Deep neural networks require a large amount of data to ensure model analysis accuracy and resistance to overfitting. Unfortunately, most researchers struggle to collect enough data during the data collection process, limiting the performance of DL models [65]. The emergence of transfer learning lends strong support to the advancement of DL

algorithms. It allows the transfer of knowledge learned from a domain pattern to a new domain, and performs classification [66]. The ImageNet database, which contains 14 million images classified into 1000 categories, has been used to train our deep neural network pre-training model. To match the classification task presented in this paper, we reduced the number of fully connected layer outputs from 1000 to 2 and set the fully connected layer's initial weight and bias to 10.

3.6. Evaluation Metrics

To assess the performance of each neural network classifier, we use accuracy (Acc), sensitivity (Se), specificity (Sp), area under curve (AUC), and time consumption (Tc). Acc, Se, and Sp are crucial clinical parameters that serve as a reference for us when we evaluate model performance. Acc is defined as the proportion of correctly classified samples to all samples. Se represents the classifier's ability to identify patients. Sp denotes the classifier's ability to identify healthy individuals. AUC is used to evaluate the model's performance with unbalanced data, which is one of our primary metrics for assessing model performance. Tc is used to evaluate how much computer resources each neural network consumes.

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{TN} + \text{FP}} \quad (8)$$

$$\text{Se} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (9)$$

$$\text{Sp} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (10)$$

where, TP is a true positive signal, FN is a false negative signal, TN is a true negative signal, and FP is a false positive signal.

4. Results

We obtained four datasets from the preprocessing and feature analysis of the KS signal, which are the segmented-based CWT (ST-CWT) data set, the segmented-based MFCC (ST-MFCC) data set, the unsegmented-based CWT (UST-CWT) data set, and the unsegmented-based MFCC (UST-MFCC) data set. After data augmentation, each dataset contained 996 healthy KS images and 696 patient KS images. The pre-trained AlexNet, VGG19, ResNet50, and Xception networks are used for HF prediagnosis studies based on the four data sets. Model validation was accomplished using a fivefold cross-validation method, and model performance was measured using average Acc, Se, Sp, and Tc.

The model's optimization function is SGDM in the calculation; because performance across model structures is inconsistent, the learning rate, batchSize setting, and max epoch were re-optimized and matched. Following our optimization, the parameters of each neural network were as follows:

AlexNet: learning rate 2×10^{-4} , batch size 128, and maximum epoch 20.

VGG19: learning rate 1×10^{-4} , batch size 32, and maximum epoch 15.

ResNet50: learning rate 4×10^{-3} , batchSize 64, and maximum epoch 10.

Xception: learning rate 5×10^{-4} , batch size 32, and maximum epoch 15.

The computing resources are small workstations equipped with an Intel XEON Platinum 8260L CPU and an NVIDIA RTX4000 graphics card.

Table 2 shows a comparison of classification results for each neural network model on the ST-CWT and ST-MFCC datasets. It can be seen that Xception had the best performance in the ST-CWT datasets, with AUC, Acc, Se, and Sp of 0.979, 93.4%, 94.3%, and 92.2%, respectively. It was followed by VGG19, with AUC (0.965), Acc (89.0%), Se (91.5%), and Sp (85.5%). The ResNet50 came third, and its AUC, Acc, Se, and Sp were 0.962, 89.8%, 91.1%, and 87.9%, respectively. AlexNet came last, with AUC (0.947), Acc (87.0%), Se

(91.8%), and Sp (80.2%). AlexNET, VGG19, ResNet50, and Xception had Tcs of 3 min, 9 min, 6 min, and 55 min, respectively.

Table 2. Performance of four types of neural networks in the ST-CWT datasets.

Classifier	Acc (%)	Se (%)	Sp (%)	AUC	Tc (min)
AlexNet	87.0	91.8	80.2	0.947	3
VGG19	89.0	91.5	85.5	0.965	9
ResNet50	89.8	91.1	87.9	0.962	6
Xception	93.4	94.3	92.2	0.979	55

As shown in Table 3, when using the ST-MFCC datasets for HF prediagnosis, Xception performed the best, with AUC (0.988), Acc (95.0%), Se (98.6%), and Sp (89.9%). It was followed by Resnet50, with AUC (0.977), Acc (95.0%), Se (91.1%), and Sp (97.7%). The VGG19 achieved AUC (0.976), Acc (93.0%), Se (94.8%), and Sp (90.5%), and AlexNet obtained AUC (0.951), Acc (90.4%), Se (94.4%), and Sp (84.8%). The average Tc for AlexNET, VGG19, ResNet50, and Xception was 3 min, 9 min, 6 min, and 55 min, respectively.

Table 3. Performance of four types of neural networks in the ST-MFCC datasets.

Classifier	Acc (%)	Se (%)	Sp (%)	AUC	Tc (min)
AlexNet	90.4	94.4	84.8	0.951	3
VGG19	93.0	94.8	90.5	0.976	9
ResNet50	95.0	91.1	97.7	0.983	6
Xception	95.0	98.6	89.9	0.988	55

The ROC curve, which is based on the confusion matrix, is not only a comprehensive index that reflects the continuous variables of sensitivity and specificity, but also an important reference for evaluating the prediction ability of the binary classification model. Figure 8a shows the receiver operating characteristic (ROC) curves of the four neural networks when classifying HF using the ST-CWT and ST-MFCC data sets. When using the ST-CWT dataset, the four classifiers ranked based on the ROC curve are Xception, VGG19, ResNet50, and AlexNet, with AUC values of 0.979, 0.965, 0.962, and 0.947, respectively. When the ST-MFCC dataset is used, the four classifiers are ranked in the following order: Xception, ResNet50, VGG19, and AlexNet, with AUC values of 0.988, 0.983, 0.976, and 0.951, respectively.

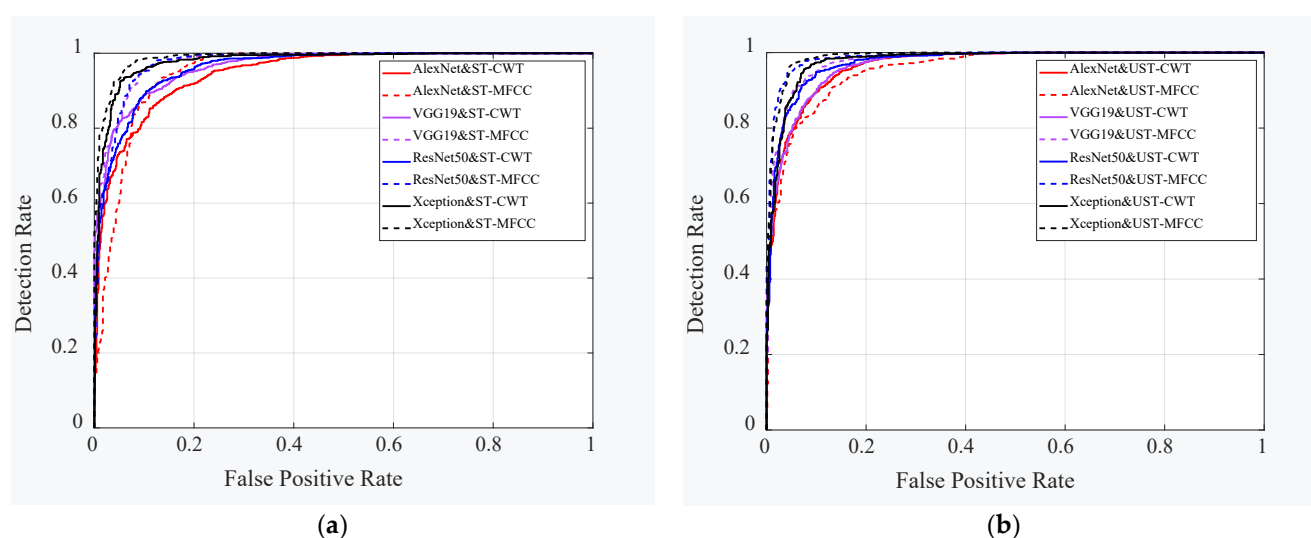


Figure 8. Performance comparison of different neural network models. (a) ST-CWT dataset, (b) ST-MFCC dataset.

The effect of unsegmented KS signal characteristics on HF prediagnosis was investigated. Tables 4 and 5 show the performance of the four neural networks on the UST-CWT and UST-MFCC datasets. As shown in Table 4, when the UST-CWT dataset is used for classification, Xception got AUC (0.979), Acc (94.3%), Se (95.7%), and Sp (92.4%). Resnet50 came in second, with an AUC of 0.974, Acc of 92.2%, Se of 90.5%, and Sp of 93.4%. The AUC, Acc, Se, and Sp obtained by VGG19 were 0.968, 91.1%, 96.4%, and 83.6%, respectively, and the AlexNet got AUC (0.966), Acc (90.9%), Se (94.9%), and Sp (85.2%). The average Tc for each model was as follows: Xception (55 min), ResNet50 (6 min), VGG19 (9 min), and AlexNet (3 min).

Table 4. Performance of four types of neural networks on the UST-CWT datasets.

Classifier	Acc (%)	Se (%)	Sp (%)	AUC	Tc (min)
AlexNet	90.9	94.9	85.2	0.966	3
VGG19	91.1	96.4	83.6	0.968	9
ResNet50	92.2	90.5	93.4	0.974	6
Xception	94.3	95.7	92.4	0.979	55

Table 5. Performance of four types of neural networks in the UST-MFCC datasets.

Classifier	Acc (%)	Se (%)	Sp (%)	AUC	Tc (min)
AlexNet	89.3	93.9	82.8	0.956	3
VGG19	93.0	97.5	86.5	0.979	9
ResNet50	95.4	92.0	97.8	0.988	6
Xception	96.0	97.5	93.8	0.989	55

Table 5 demonstrates that, when the UST-MFCC dataset is applied for HF prediagnosis, Xception obtained AUC (0.989), Acc (96.0%), Se (97.5%), and Sp (93.8%). ResNet50 got AUC (0.988), Acc (95.4%), Se (92.0%), and Sp (97.8%). The classification performance of VGG19 ranked third, with an AUC of 0.979, Acc of 93.0%, Se of 97.5%, and Sp of 86.5%. The AUC, Acc, Se, and Sp obtained by AlexNet were 0.956, 89.3%, 93.9%, and 82.8%. The average Tc for AlexNET, VGG19, ResNet50, and Xception was 3 min, 9 min, 6 min, and 55 min, respectively.

Figure 8b shows the ROC curves of four neural networks when classifying HF using UST-CWT and UST-MFCC data sets. Using the UST-CWT dataset, the four classifiers ranked as follows: Xception, ResNet50, VGG19, and AlexNet, with AUC values of 0.979, 0.974, 0.968, and 0.966, respectively. When the UST-MFCC dataset was used, the four classifiers performed in the following order: Xception, ResNet50, VGG19, and AlexNet, with AUC values of 0.989, 0.988, 0.979, and 0.956, respectively.

5. Discussion

The convenience and non-invasiveness of KS acquisition formed the foundation of our work. On this basis, we employed a DL network to conduct HF prediagnosis research with KS as the research object, which is a significant effort in the field of KS research. Four feature sets were used to study the application of KS signal in HF prediagnosis in detail, including the ST-CWT dataset, ST-MFCC dataset, UST-CWT dataset, and UST-MFCC dataset. DL architectures AlexNet, VGG19, ResNet50, and Xception were used to classify these four datasets. A fivefold cross-validation was used for each model validation, and the performance of each DL classifier was evaluated using Acc, Se, Sp, and AUC. The effects of KS signal segmentation and different feature extraction methods on HF prediagnosis were investigated.

According to our research, the recognition effect of neural networks on MFCC features is better than that of CWT features in the KS-based HF prediagnosis, as shown in Tables 2 and 3. When the ST-CWT dataset was used, Xception achieved the highest AUC

and ACC of all classifiers (0.979 and 93.4%), followed by VGG19, which had an AUC of 0.965, and Acc of 89.0%. Using the ST-MFCC dataset, Xception achieved an AUC of 0.988, and an Acc of 95.0%. ResNet50 got an AUC of 0.983 and an Acc of 95.0%. Simultaneously, VGG19 and AlexNet performed significantly better in classification in the ST-MFCC dataset than that in the ST-CWT dataset. When the unsegmented signal processing method was used, the neural network classifier demonstrated the same trend in the CWT and MFCC datasets, as shown in Tables 4 and 5. In the UST-MFCC dataset, Xception had an AUC of 0.989, and an Acc of 95.09%, while ResNet50 obtained an AUC of 0.988 and an Acc of 95.4%. In the UST-CWT dataset, Xception's AUC and ACC were reduced to 0.979, and 94.3%, respectively; the AUC and ACC of ResNet50 were reduced to 0.974 and 92.2%, respectively.

The average Acc comparison of the classifiers on the four datasets is shown in Figure 9. As we can see, the classification performance of the four classifiers in the unsegmented CWT and the MFCC datasets is better than that obtained by the segmentation method. Therefore, it is reasonable to conclude that although the silent period characteristics of KS signals are very weak, they are significant components of KS, and this component has a non-negligible effect in HF prediagnosis tasks. Based on the classification accuracy of the four classifiers across all data sets, we find that Xception performs exceptionally well, and its overall performance outperforms the other three classifiers. The ResNet50 network, which exhibits outstanding classification performance on these four data sets, is likewise regarded as a potent classifier in KS-based HF pre-diagnosis. Combining the ROC curve comparison results in Figure 8, we believe that Xception and ResNet50 are two excellent deep neural network algorithms in the KS-based HF prediagnosis task.

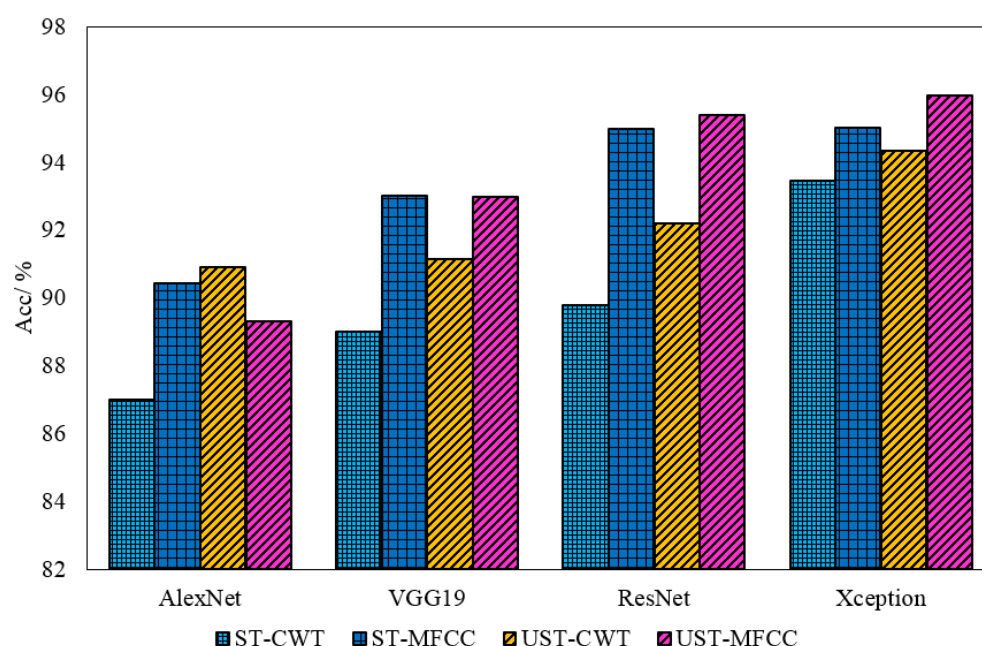


Figure 9. Comparison of the classification accuracy of four neural networks.

Furthermore, we believe that, in addition to the AUC and ACC, the computational scale and Tc are important parameters for evaluating the model's performance. According to the comparison of the four models (Tables 2–5), while Xception has the highest AUC and ACC, it consumes significantly more computing time than the other three models, which is obviously not conducive to the model's deployment and promotion in portable devices. As a result, we think that ResNet50 may be a better choice in terms of practical application.

Scholars have been widely concerned about the early diagnosis of HF, and many researchers have made outstanding contributions to the early diagnosis of HF. These

research findings not only served as a useful reference for our work but also provided us with a great deal of inspiration and motivation. As shown in Table 6, we chose a typical acoustic-based HF prediagnosis algorithm and compared their results to our algorithm. We can see that the accuracy of the proposed algorithm and the HS-based HF algorithm is basically maintained at the same level, which means our efforts have yielded largely positive results. The current results give us more confidence in continuing to work in this area.

Table 6. Comparison of HF classification algorithms based on acoustics.

Authors	Data Set	Number of Subjects	Method	Performance
Zheng et al. (2015) [67]	Collected by HS acquisition system	88 healthy volunteers and 64 CHF patients	LS-SVM	Acc 95.39% Se 96.59% Sp 93.75%
Potes et al. (2016) [34]	Physionet databases	2575 normal signals and 665 abnormal signals	AdaBoost and CNN	Acc 86.0% Se 94.2% Sp 77.8%
Gjoreski et al. (2020) [38]	six (A to F) PhysioNet Challenge datasets & measured HS by digital stethoscope	3153 signals from PhysioNet Challenge datasets and 110 healthy people, 51 CHF recorded by digital stethoscope	Machine-Learning and end-to-end DeepLearning	Acc 92.9% Se 82.3% Sp 96.2%
Yang et al. (2021) [68]	Acquired from the First Affiliated Hospital of Chongqing Medical University	41 healthy volunteers and 30 left ventricular diastolic dysfunction patients	VGG-16, VGG-19, ResNet-18, ResNet-50, DenseNet-121, and AlexNet	Acc 98.7% Se 98.6% Sp 98.8%
Zheng et al. (2022) [69]	Dataset from First Affiliated Hospital and the University-Town Hospital of Chongqing Medical University	51 healthy volunteers and 224 CHF patients	LS-SVM	Acc 82% Se 82.1% Sp 95.5%
Our method	Dataset of measured KS from the Fourth People's Hospital of Zhejiang University	116 healthy subjects and 249 CHF patients	DeepLearning(DL)	Acc 96.0% Se 97.5% Sp 93.8%

However, it should be noted that the work in this paper is an exploratory study, and we also recognize that the KS-based HF prediagnosis algorithm is still in its infancy, with many flaws in the construction of the KS database, the KS generation mechanism, and the association mechanism between KS and HF, which will be the focus of our future research.

6. Conclusions

In this paper, we propose an innovative HF prediagnosis method to investigate the use of the KS signal in the early diagnosis of HF. We investigated the effect of signal segmentation algorithms and different signal features on HF prediagnosis by creating ST-CWT, ST-MFCC, UST-CWT, and UST-MFCC datasets. Alexnet, VGG19, ResNet50, and Xception were the four DL networks used for training and classification. All the networks were validated using a fivefold cross-validation approach, and evaluated using the Acc, Se, Sp, AUC and Tc.

According to our research, the deep neural network model performed better on unsegmented datasets than that on segmented datasets. By comparing the differences between the two data sets, we believe that the KS quiet period signal characteristics discarded by signal segmentation is the main cause of this phenomenon. Our study demonstrates that the KS silent period played a crucial role in KS-based HF prediagnosis, although the human ear is not sensitive to that. Furthermore, in both segmented datasets

(ST-CWT, ST-MFCC) and unsegmented datasets (UST-CWT, UST-MFCC), the performance of the neural network based on the MFCC feature outperforms the neural network based on the CWT feature, illustrating the advantage of the MFCC feature extraction method in the audio recognition field. Our research shows the potential of KS-based non-invasive HF prediagnosis technology in early warning of HF potential risk groups and prognosis monitoring of HF patients, implying that KS-based methods can be an important auxiliary tool for HF detection. Reasonable signal processing, effective feature extraction, and efficient DL network selection, in our opinion, are the keys to achieving these goals.

This is a preliminary study with a small sample size. In the future, we will conduct more in-depth research on the mechanism of KS generation, the mechanism of correlation between KS and HF, and the mechanism of correlation between drugs, basic diseases, and KS. Prospective testing and blind verification in the test population will also be conducted. Simultaneously, we will select an efficient DL model and perform optimization and large-scale database verification based on the existing research foundation to achieve embedded device migration and community promotion.

Author Contributions: Conceptualization, H.Z. (Hong Zhou) and S.X.; methodology, H.Z. (Huanyu Zhang); software, R.W.; validation, H.Z. (Hong Zhou) and S.X.; formal analysis, H.Z. (Hong Zhou) and S.X.; investigation, H.Z. (Huanyu Zhang), R.W., S.J. and Wu, Y.; resources, S.J. and Y.W.; data curation, H.Z. (Huanyu Zhang), S.J. and Y.W.; writing—original draft preparation, H.Z. (Huanyu Zhang); writing—review and editing, H.Z. (Huanyu Zhang), R.W.; visualization, H.Z. (Huanyu Zhang) and R.W.; supervision, H.Z. (Hong Zhou) and S.X.; project administration, H.Z. (Hong Zhou) and S.X.; funding acquisition, H.Z. (Hong Zhou) and S.X. All authors have read and agreed to the published version of the manuscript.

Funding: The paper was supported by the Key R&D Program of Zhejiang Province of China (No. 2021C03030).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available from The Fourth Affiliated Hospital Zhejiang University School of Medicine, Zhejiang University, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. World Health Statistics. 2022. Available online: <https://www.who.int/data/gho/publications/world-health-statistics> (accessed on May 20, 2022).
2. Bennett, J.E.; Kontis, V.; Mathers, C.D.; Guillot, M.; Rehm, J.; Chalkidou, K.; Kengne, A.P.; Carrillo-Larco, R.M.; Bawah, A.A.; Dain, K.; et al. NCD Countdown 2030: Pathways to achieving Sustainable Development Goal target 3.4. *Lancet* **2020**, *396*, 918–934.
3. Ohta, Y.; Kitao, S.; Yunaga, H.; Fujii, S.; Mukai, N.; Yamamoto, K.; Ogawa, T. Myocardial delayed enhancement CT for the evaluation of heart failure: Comparison to MRI. *Radiology* **2018**, *288*, 682–691.
4. Sanna, G.D.; Canonico, M.E.; Santoro, C.; Esposito, R.; Masia, S.L.; Galderisi, M.; Parodi, G.; Nihoyannopoulos, P. Echocardiographic longitudinal strain analysis in heart failure: Real usefulness for clinical management beyond diagnostic value and prognostic correlations? A comprehensive review. *Curr. Heart Fail. Rep.* **2021**, *18*, 290–303.
5. Lippi, G.; Henry, B.M.; Sanchis-Gomar, F. Physical inactivity and cardiovascular disease at the time of coronavirus disease 2019 (COVID-19). *Eur. J. Prev. Cardiol.* **2020**, *27*, 906–908.
6. Eeftink Schattenkerk, D.W.; Van Lieshout, J.J.; Van Den Meiracker, A.H.; Wesseling, K.R.; Blanc, S.; Wieling, W.; Van Montfrans, G.A.; Settels, J.J.; Wesseling, K.H.; Westerhof, B.E.; et al. Nexfin noninvasive continuous blood pressure validated against Riva-Rocci/Korotkoff. *Am. J. Hypertens.* **2009**, *22*, 378–383.
7. Padwal, R.; Campbell, N.R.; Schutte, A.E.; Olsen, M.H.; Delles, C.; Etyang, A.; Cruickshank, J.K.; Stergiou, G.; Rakotz, M.K.; Wozniak, G.; et al. Optimizing observer performance of clinic blood pressure measurement: A position statement from the Lancet Commission on Hypertension Group. *J. Hypertens.* **2019**, *37*, 1737.

8. Gosse, P.; Durandet, P.; Roudaut, R. The QKD interval as an easy assessment of pulse wave velocity: Validation of the method. *J. Hypertens.* **1991**, *9*, S447.
9. Takahashi, O.; Shimbo, T.; Rahman, M.; Musa, R.; Kurokawa, W.; Yoshinaka, T.; Fukui, T. Validation of the auscultatory method for diagnosing peripheral arterial disease. *Fam. Pract.* **2006**, *23*, 10–14.
10. Hussain, L.; Awan, I.A.; Aziz, W.; Saeed, S.; Ali, A.; Zeeshan, F.; Kwak, K.S. Detecting congestive heart failure by extracting multimodal features and employing machine learning techniques. *BioMed Res. Int.* **2020**, *2020*, 4281243.
11. Kabir, M.A.; Shahnaz, C. Denoising of ECG signals based on noise reduction algorithms in EMD and wavelet domains. *Biomed. Signal Process. Control* **2012**, *7*, 481–489.
12. Misal, A.; Sinha, G.R. Denoising of PCG signal by using wavelet transforms. *Adv. Comput. Res.* **2012**, *4*, 46–49.
13. Liang, H.; Lukkarienen, S.; Hartimo, I. Heart sound segmentation algorithm based on heart sound envelopogram. In Proceedings of the Computers in Cardiology 1997, Lund, Sweden, 7–10 September 1997; IEEE: Piscataway, NJ, USA, 1997; pp. 105–108.
14. Springer, D.B.; Tarassenko, L.; Clifford, G.D. Logistic regression-HSMM-based heart sound segmentation. *IEEE Trans. Biomed. Eng.* **2015**, *63*, 822–832.
15. Papadaniil, C.D.; Hadjileontiadis, L. Efficient heart sound segmentation and extraction using ensemble empirical mode decomposition and kurtosis features. *IEEE J. Biomed. Health Inform.* **2013**, *18*, 1138–1152.
16. Chen, W.; Sun, Q.; Chen, X.; Xie, G.; Wu, H.; Xu, C. Deep learning methods for heart sounds classification: A systematic review. *Entropy* **2021**, *23*, 667.
17. Khare, S.K.; Bajaj, V. Time–frequency representation and convolutional neural network-based emotion recognition. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 2901–2909.
18. Arias-Vergara, T.; Klumpp, P.; Vasquez-Correa, J.C.; Nöth, E.; Orozco-Arroyave, J.R.; Schuster, M. Multi-channel spectrograms for speech processing applications using deep learning methods. *Pattern Anal. Appl.* **2021**, *24*, 423–431.
19. El-Segaier, M.; Lilja, O.; Lukkarienen, S.; Sörnmo, L.; Sepponen, R.; Pesonen, E. Computer-based detection and analysis of heart sound and murmur. *Ann. Biomed. Eng.* **2005**, *33*, 937–942.
20. Varghees, V.N.; Ramachandran, K.I. A novel heart sound activity detection framework for automated heart sound analysis. *Biomed. Signal Process. Control* **2014**, *13*, 174–188.
21. Acharya U R, Fujita H, Adam M; et al. Automated characterization and classification of coronary artery disease and myocardial infarction by decomposition of ECG signals: A comparative study. *Inf. Sci.* **2017**, *377*, 17–29.
22. Narin, A.; Isler, Y.; Ozer, M. Investigating the performance improvement of HRV Indices in CHF using feature selection methods based on backward elimination and statistical significance. *Comput. Biol. Med.* **2014**, *45*, 72–79.
23. Watrous, R.L. Computer-aided auscultation of the heart: From anatomy and physiology to diagnostic decision support. In Proceedings of the 2006 International Conference of the IEEE Engineering in Medicine and Biology Society, New York, NY, USA, 30 August–3 September 2006; IEEE: Piscataway, NJ, USA, 2006; pp. 140–143.
24. Shuvo, S.B.; Ali, S.N.; Swapnil, S.L.; Al-Rakhami, M.S.; Gumaei, A. CardioXNet: A novel lightweight deep learning framework for cardiovascular disease classification using heart sound recordings. *IEEE Access* **2021**, *9*, 36955–36967.
25. Son, G.Y.; Kwon, S. Classification of heart sound signal using multiple features. *Appl. Sci.* **2018**, *8*, 2344.
26. Jin, K.; Neubeck, L.; Koo, F.; Ding, D.; Gullick, J. Understanding prevention and management of coronary heart disease among Chinese immigrants and their family carers: A socioecological approach. *J. Transcult. Nurs.* **2020**, *31*, 257–266.
27. Benmira, A.; Perez-Martin, A.; Schuster, I.; Veye, F.; Triboulet, J.; Berron, N.; Aichoun, I.; Coudray, S.; Laurent, J.; Bereksi-Reguig, F.; et al. An ultrasound look at Korotkoff sounds: The role of pulse wave velocity and flow turbulence. *Blood Press. Monit.* **2017**, *22*, 86–94.
28. Mondal, A.; Saxena, I.; Tang, H.; Banerjee, P. A noise reduction technique based on nonlinear kernel function for heart sound analysis. *IEEE J. Biomed. Health Inform.* **2017**, *22*, 775–784.
29. Deng, S.W.; Han, J.Q. Adaptive overlapping-group sparse denoising for heart sound signals. *Biomed. Signal Process. Control* **2018**, *40*, 49–57.
30. Lopac, N.; Hrzić, F.; Vuksanović, I.P.; Lerga, J. Detection of Non-Stationary GW Signals in High Noise From Cohen’s Class of Time–Frequency Representations Using Deep Learning. *IEEE Access* **2021**, *10*, 2408–2428.
31. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444.
32. Noman, F.; Salleh, S.H.; Ting, C.M.; Samdin, S.B.; Ombao, H.; Hussain, H. A Markov-switching model approach to heart sound segmentation and classification. *IEEE J. Biomed. Health Inform.* **2019**, *24*, 705–716.
33. Nogueira, D.M.; Ferreira, C.A.; Gomes, E.F.; Jorge, A. Classifying heart sounds using images of motifs, MFCC and temporal features. *J. Med. Syst.* **2019**, *43*, 1–13.
34. Potes, C.; Parvaneh, S.; Rahman, A.; Rahman, B. Ensemble of feature-based and deep learning-based classifiers for detection of abnormal heart sounds. In Proceedings of the 2016 Computing in Cardiology Conference (CinC), Vancouver, BC, Canada, 11–14 September 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 621–624.
35. Rubin, J.; Abreu, R.; Ganguli, A.; Nelaturi, S.; Matei, I.; Sricharan, K. Classifying heart sound recordings using deep convolutional neural networks and mel-frequency cepstral coefficients. In Proceedings of the 2016 Computing in Cardiology Conference (CinC), Vancouver, BC, Canada, 11–14 September 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 813–816.
36. Zhang, W.; Han, J.; Deng, S. Abnormal heart sound detection using temporal quasi-periodic features and long short-term memory without segmentation. *Biomed. Signal Process. Control* **2019**, *53*, 101560.

37. He, Y.; Li, W.; Zhang, W.; Zhang, S.; Pi, X.; Liu, H. Research on segmentation and classification of heart sound signals based on deep learning. *Appl. Sci.* **2021**, *11*, 651.
38. Gjoreski, M.; Gradišek, A.; Budna, B.; Gams, M.; Poglajen, G. Machine learning and end-to-end deep learning for the detection of chronic heart failure from heart sounds. *IEEE Access* **2020**, *8*, 20313–20324.
39. Gosse, P.; Guillo, P.; Ascher, G.; Clementy, J. Assessment of arterial distensibility by monitoring the timing of Korotkoff sounds. *Am. J. Hypertens.* **1994**, *7*, 228–233.
40. Constans, J.; Germain, C.; Gosse, P.; Taillard, J.; Tiev, K.; Delevaux, I.; Mouthon, L.; Schmidt, C.; Granel, F.; Soria, P.; et al. Arterial stiffness predicts severe progression in systemic sclerosis: The ERAMS study. *J. Hypertens.* **2007**, *25*, 1900–1906.
41. Gosse, P.; Cremer, A.; Papaioannou, G.; Yeim, S. Arterial stiffness from monitoring of timing of Korotkoff sounds predicts the occurrence of cardiovascular events independently of left ventricular mass in hypertensive patients. *Hypertension* **2013**, *62*, 161–167.
42. El Tahlawi, M.; Abdelbaset, M.; Gouda, M.; Hussein, I. Can we predict the presence of coronary lesions from blood pressure measurement? A new clinical method. *Hypertens. Res.* **2015**, *38*, 260–263.
43. Yamamoto, T.; Miura, S.; Suematsu, Y.; Kuwano, T.; Sugihara, M.; Ike, A.; Iwata, A.; Nishikawa, H.; Saku, K. A relative difference in systolic blood pressure between arms by synchronal measurement and conventional cardiovascular risk factors are associated with the severity of coronary atherosclerosis. *Heart Vessel.* **2016**, *31*, 863–870.
44. Jia, S.; Wu, Y.; Wang, W.; Lin, W.; Chen, Y.; Zhang, H.; Xia, S.; Zhou, H. An Exploratory Study on the Relationship between Brachial Arterial Blood Flow and Cardiac Output. *J. Healthc. Eng.* **2021**, *2021*, 1251199.
45. McDonagh, T.A.; Metra, M.; Adamo, M.; Gardner, R.S.; Baumbach, A.; Böhm, M.; Burri, H.; Butler, J.; Čelutkienė, J.; Chioncel, O.; et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: Developed by the Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) With the special contribution of the Heart Failure Association (HFA) of the ESC. *Eur. Heart J.* **2021**, *42*, 3599–3726.
46. Suzuki, K.; Washio, T.; Tsukamoto, S.; Kato, K.; Iwamoto, E.; Ogoh, S. Habitual cigarette smoking attenuates shear-mediated dilation in the brachial artery but not in the carotid artery in young adults. *Physiol. Rep.* **2020**, *8*, e14369.
47. Yao, G.H.; Deng, Y.; Liu, Y.; Xu, M.J.; Zhang, C.; Deng, Y.B.; Ren, W.D.; Li, Z.A.; Tang, H.; Zhang, Q.B.; et al. Echocardiographic measurements in normal Chinese adults focusing on cardiac chambers and great arteries: A prospective, nationwide, and multicenter study. *J. Am. Soc. Echocardiogr.* **2015**, *28*, 570–579.
48. Lang, R.M.; Badano, L.P.; Mor-Avi, V.; Afilalo, J.; Armstrong, A.; Ernande, L.; Flachskampf, F.A.; Foster, E.; Goldstein, S.A.; Kuznetsova, T.; et al. Recommendations for cardiac chamber quantification by echocardiography in adults: An update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. *Eur. Heart J-Cardiovasc. Imaging* **2015**, *16*, 233–271.
49. Pan, F.; Zheng, D.; He, P.; Murray, A. Does the position or contact pressure of the stethoscope make any difference to clinical blood pressure measurements: An observational study. *Medicine* **2014**, *93*, e301.
50. Pan, F.; Chen, F.; Liu, C.; Yang, Z.; Liu, Z.; Zheng, D. Quantitative Comparison of Korotkoff Sound Waveform Characteristics: Effects of Static Cuff Pressures and Stethoscope Positions. *Ann. Biomed. Eng.* **2018**, *46*, 1736–1744.
51. Saini, M. Proposed algorithm for implementation of Shannon energy envelope for heart sound analysis. *Int. J. Electron. Commun. Technol.* **2016**, *7*, 15–19.
52. Giordano, N.; Knaflitz, M. A novel method for measuring the timing of heart sound components through digital phonocardiography. *Sensors* **2019**, *19*, 1868.
53. Rajani Kumari, L.V.; Padma Sai, Y.; Balaji, N. R-peak identification in ECG signals using pattern-adapted wavelet technique. *IETE J. Res.* **2021**, 1–10. <https://doi.org/10.1080/03772063.2021.1893229>.
54. Rajput, J.S.; Sharma, M.; Kumar, T.S.; Acharya, U.R. Automated Detection of Hypertension Using Continuous Wavelet Transform and a Deep Neural Network with Ballistocardiography Signals. *Int. J. Environ. Res. Public Health* **2022**, *19*, 4014.
55. Sharma, M.; Dhiman, H.S.; Acharya, U.R. Automatic identification of insomnia using optimal antisymmetric biorthogonal wavelet filter bank with ECG signals. *Comput. Biol. Med.* **2021**, *131*, 104246.
56. Lalitha, S.; Geyasruti, D.; Narayanan, R.; Shravani, M. Emotion detection using MFCC and cepstrum features. *Procedia Comput. Sci.* **2015**, *70*, 29–35.
57. Tiwari, V. MFCC and its applications in speaker recognition. *Int. J. Emerg. Technol.* **2010**, *1*, 19–22.
58. Astuti, W.; Sediono, W.; Aibinu, A.M.; Akmeliawati, R.; Salami, M.J.E. Adaptive Short Time Fourier Transform (STFT) Analysis of seismic electric signal (SES): A comparison of Hamming and rectangular window. In Proceedings of the 2012 IEEE Symposium on Industrial Electronics and Applications, Bandung, Indonesia, 23–26 September 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 372–377.
59. Hinton, G.E.; Salakhutdinov, R.R. Reducing the dimensionality of data with neural networks. *Science* **2006**, *313*, 504–507.
60. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90.
61. Bansal, M.; Kumar, M.; Sachdeva, M.; Mittal, A. Transfer learning for image classification using VGG19: Caltech-101 image data set. *J. Ambient Intell. Humaniz. Comput.* **2021**, 1–12. <https://doi.org/10.1007/s12652-021-03488-z>.
62. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

63. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
64. Touvron, H.; Cord, M.; Douze, M.; Massa, F.; Sablayrolles, A.; Jégou, H. Training data-efficient image transformers & distillation through attention. In Proceedings of the International Conference on Machine Learning, PMLR 2021, Virtual Event, 18–24 July 2021; pp. 10347–10357.
65. Gajendran, M.K.; Khan, M.Z.; Khattak MA, K. ECG Classification using Deep Transfer Learning. In Proceedings of the 2021 4th International Conference on Information and Computer Technologies (ICICT), Kahului, HI, USA, 11–14 March 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1–5.
66. Olivas, E.S.; Guerrero JD, M.; Martinez-Sober, M.; Magdalena-Benedito, J.R.; Serrano, L. (Eds.) *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques: Algorithms, Methods, and Techniques*; IGI Global: Hershey, PA, USA, 2009.
67. Zheng, Y.; Guo, X.; Qin, J.; Xiao, S. Computer-assisted diagnosis for chronic heart failure by the analysis of their cardiac reserve and heart sound characteristics. *Comput. Methods Programs Biomed.* **2015**, *122*, 372–383.
68. Yang, Y.; Guo, X.M.; Wang, H.; Zheng, Y.N. Deep Learning-Based Heart Sound Analysis for Left Ventricular Diastolic Dysfunction Diagnosis. *Diagnostics* **2021**, *11*, 2349.
69. Zheng, Y.; Guo, X.; Wang, Y.; Qin, J.; Lv, F. A multi-scale and multi-domain heart sound feature-based machine learning model for ACC/AHA heart failure stage classification. *Physiol. Meas.* **2022**, *43*, 065002.