**Title of Your Document**

Author Name  
Institution Name

Deep Learning Models for Coronary Artery DiseaseClassification using PhonocardiogramsChenhe LiuThe British School of BeijingBeijing, ChinaJiasheng ZhangAnhui International Studies UniversityWuhu, ChinaNanomega Research InstituteBeijing, ChinaYibo ZhangNanomega Research InstituteBeijing, ChinaGezhi Future Research InstituteBeijing, ChinaSchool of Systems and Computing, UNSW AustraliaAbstractThis research paper presents a comparative analysisof three deep learning models for automatic heart sound classi-fication using spectrograms as features. The models investigatedinclude Custom CNN, Transfer Learning based on pretrainedCNN Architectures i.e., VGG16 and MobileNetV2, and PSO-optimized CNN. The performance of these models is evaluatedin terms of precision, recall, and F1-score for both normal andabnormal heart sounds. The results are compared with previousstudies that utilized different feature extraction techniques andclassification algorithms. The proposed work achieves promisingresults i.e., 98% of F1 score, demonstrating the effectivenessof using spectrograms and Transfer Learning for heart soundclassification.IndexTermsHeartsounds,classification,deeplearn-ing,transfer learning, optimizationI. INTRODUCTIONCardiovascular diseases (CVDs) pose a significant globalhealth challenge, accounting for a substantial number of deathsworldwide [1][3]. Among the various cardiovascular condi-tions, coronary artery disease (CAD) is particularly criticaldue to its high prevalence and potential life-threatening conse-quences. Early and accurate diagnosis of CAD plays a crucialrole in improving patient outcomes and reducing mortalityrates [4]. In recent years, the application of deep learning andmachine learning techniques to medical diagnostics has showngreat promise, and this paper focuses on the classificationof heart sounds or phonocardiograms (PCGs) using theseadvanced techniques.The use of PCGs in cardiac diagnosis has gained promi-nence due to their non-invasive nature and ability to providevaluable insights into cardiac abnormalities. PCGs capturethe acoustic signals generated by the heart during its normalfunctioning and can reveal abnormal patterns associated withvarious cardiac conditions [5][8]. Analyzing PCGs tradition-ally relies on expert auscultation, which is subjective anddemands specialized training. Therefore, the development ofautomated systems for PCG analysis using deep learningmethods holds tremendous potential in enhancing diagnosticaccuracy, reducing human error, and improving patient care.This research paper proposes a comprehensive comparativeanalysis of three different deep learning models for automaticPCG classification. The models considered in this study are:a custom convolutional neural network (CNN), a transfer-learning based CNN architectures, and a nature-inspired (Par-ticle Swarm Optimization - PSO) optimized CNN. The choiceof deep learning models is motivated by their ability to learnintricate patterns and features directly from data, making themsuitable for complex classification tasks.To extract meaningful features from PCGs, this study adoptsthe short-time Fourier transform (STFT) to generate spectro-grams. Spectrograms provide a visual representation of thefrequency content of the PCG signals over time, enablingthe models to capture important acoustic features related tocardiac abnormalities. By utilizing STFT-based spectrogramsas input features, the proposed approach aims to exploit therich information embedded in PCGs for accurate classification.The importance and novelty of this research lie in severalaspects. Firstly, by utilizing deep learning models, this study aimsto overcome the limitations of manual auscultation andprovide an automated and objective approach for PCG-based diagnosis. Secondly, the comparative analysis of four different mod-els allows for a comprehensive evaluation of their perfor-mance, highlighting their strengths and weaknesses in thecontext of CAD classification. Lastly, the incorporation of nature-inspired optimizationtechniques, such as PSO, into the deep learning frame-work adds a novel dimension to the research, potentiallyenhancing the efficiency and effectiveness of the classifi-cation process.The ultimate goal of this research paper is to contributeto the advancement of automated heart sounds classificationby providing a detailed comparative analysis of deep learningmodels for PCG classification. By evaluating the performanceof the proposed custom CNN, transfer-learning based CNN,and PSO-optimized CNN on STFT-based spectrograms, thisstudy aims to demonstrate the superiority of the proposedapproach in terms of accuracy, efficiency, and robustness. Thefindings of this research could have significant implications inthe field of cardiac diagnostics, paving the way for improvedclinical decision-making.

II. LITERATURE REVIEWXiao et al in [9], introduces a novel deep learning methodfor classifying heart sounds to predict cardiovascular diseases.The approach consists of pre-processing steps and a convolu-tional neural network (CNN) architecture. Rather than using2-D time-frequency representations, the method directly uses1-D raw waveform phonocardiograms (PCGs). Sliding windowsegmentation is applied to split PCG recordings into fixed-length patches. The proposed CNN architecture includes cliqueblocks and transition blocks, enabling spatial and channelattention. Features from each block are concatenated andcompressed before being fed into global pooling. The squeezedfeatures are merged and passed through a fully-connectedlayer for classification. Experimental results demonstrate su-perior classification performance compared to state-of-the-artmethods while utilizing fewer parameters. Similarly, Ren etal in [10], investigates the use of pretrained ConvolutionalNeural Networks (CNNs) for classifying Phonocardiogram(PCG) signals. The PCG files are segmented and transformedinto scalogram images using wavelet transformation. Twoapproaches are explored: 1) employing a pretrained CNN orfine-tuning it on heart sound data, and 2) using an end-to-end CNN through transfer learning. Deep PCG representationsare extracted from fully connected layers, and linear SVMis used for classification. Experimental results demonstratethat the deep PCG representations obtained from a fine-tunedCNN achieve the highest mean accuracy of 56.2% for heartsound classification, outperforming conventional methods. Thestudy also explores adapting the parameters of VGG16 usingtransfer learning, showing promising results for efficient CNN-based classification. The findings highlight the effectiveness ofpretrained CNNs in capturing meaningful features from PCGsignals, with significant improvements in accuracy comparedto traditional approaches. This research contributes to theadvancement of PCG-based classification methods for cardio-vascular disease diagnosis.Safara et al introduces multi-level basis selection (MLBS) in[11] as a method to extract informative features from waveletpacket transform (WPT) for heart sound classification. MLBSapplies exclusion criteria based on frequency range, noise fre-quency, and energy threshold to preserve the most informativebases of the WPT decomposition tree. The proposed MLBSachieves an accuracy of 97.56% in classifying normal heartsound and different heart valve disorders. The preprocessingstep involves normalization and segmentation of the PCGsignals, and feature extraction is performed by selecting nodesfrom the bottom levels of the WPT tree. The candidate setconsists of nodes from levels 6, 7, and 8, totaling 448 nodes,which are then pruned based on the exclusion criteria. Simi-larly, Nguyen et al [12] proposes two deep learning models forclassifying heart sound signals based on log-mel spectrogramfeatures. The dataset consists of five classes, including onenormal class and four anomalous classes. The models, namelyLong Short-Term Memory (LSTM) and Convolutional NeuralNetwork (CNN), are utilized for heartbeat sound classification.The heart sound signals are framed to a consistent length, andlog-mel spectrogram features are extracted. The LSTM modelconsists of two LSTM layers and three fully connected layers,while the CNN model includes three convolutional layers andtwo fully connected layers. The analysis demonstrates highclassification performance, achieving an overall accuracy ofapproximately 99.67%. The results also indicate improvedperformance compared to previous studies in this field.The remainder of this paper is organized as follows: insection III,the proposed methodology is discussed in detail.In section IV, the achieved results are discussed and analyzed.Finally, the overall work is concluded and future directions areprovided in section V.III. METHODOLOGYA. DatasetThe 2016 PhysioNet/CinC Challenge [13] focused on de-veloping algorithms to classify heart sound recordings anddetermine whether further expert diagnosis was required. Thechallenge provided a collection of heart sound recordings fromvarious clinical and nonclinical environments. The recordingswere obtained from different locations on the body, includingthe aortic area, pulmonic area, tricuspid area, and mitral area.The recordings consisted of normal and abnormal heart sounds,with the abnormal ones coming from patients with cardiacdiagnoses such as heart valve defects and coronary arterydisease.The training set provided for the challenge consisted of fivedatabases containing a total of 3,126 heart sound recordings,ranging from 5 seconds to over 120 seconds in duration.The recordings were in WAV format and had been resampledto 2,000 Hz. Each recording represented a single precordiallocation.B. Pre-processingIn the conducted study, a third-order Butterworth band-pass filter [14] was utilized during the pre-processing stage toextract the desired frequencies in the PCG signals while elimi-nating unwanted frequencies. The cut-off frequencies of 25 Hzand 200 Hz were selected to capture the relevant informationin the recordings. This pre-processing step was performed toenhance the subsequent analysis and classification of the heartsound signals.C. SpectrogramsThe Short-Time Fourier Transform (STFT) is a techniquethat analyzes the frequency content of non-stationary signalsover time [15][17]. It involves dividing the signal into shortsegments, applying a window function, and computing theFourier Transform for each segment. This yields a time-frequency representation known as the spectrogram. The spec-trogram displays the power or intensity of frequency compo-nents over time. To enhance the spectrogram, the magnitudevalues are often logarithmically compressed, resulting in alog-spectrogram. Additionally, a mel-scale transformation is

Fig. 1: Proposed MethodologyFig. 2: Phonocardiogramapplied to the log-spectrogram to approximate the humanauditory systems response to different frequencies.The STFT of a signal x[n] is defined as:X[m, ] =Xn=x[n]w[n m]ejn,(1)where X[m, ] represents the complex-valued STFT coef-ficients at time frame m and frequency . w[n] denotes awindow function applied to the signal.To obtain the magnitude spectrogram S[m, ], we computethe magnitude of the STFT coefficients:S[m, ] = |X[m, ]|.(2)The phase spectrogram [m, ] can be obtained by extract-ing the phase information from the STFT coefficients:[m, ] = arg(X[m, ]).(3)The spectrogram S(t, f) is a visual representation of themagnitude spectrogram, where t represents the time and frepresents the frequency. It is typically displayed as a 2D plot,with time on the x-axis, frequency on the y-axis, and coloror intensity representing the magnitude. The spectrograms fornormal and abnormal OCG are shown in Figure 3.(a)(b)Fig. 3: (a) Spectrogram of Normal PCG (b) Spectrogram ofAbnormal PCGD. ClassificationThree approaches are implemented and described in thissection, also shown in 4.1) Custom CNN: The custom CNN architecture utilized inthis work consists of five convolutional layers, each with aspecific number of neurons: 64, 64, 32, 32, and 16, respec-tively. These convolutional layers are responsible for extractinghierarchical features from the input spectrograms [18]. Atypical blocj diagram for CNN is shown in Figure 4.Convolution: Y = (W X + b)(4)Input image: X =x11x12 x1nx21x22 x2n............xm1xm2 xmn(5)

Weight matrix: W =w11w12 w1kw21w22 w2k............wk1wk2 wkk(6)Bias term: b =b1b2 bk(7)The rectified linear unit (ReLU) activation function is appliedafter each convolutional layer. ReLU introduces non-linearityto the model by replacing negative values with zeros, allowingthe network to capture complex patterns and enhance its abilityto learn and discriminate between different features.ReLU: Y = max(0, X)(8)Following each convolutional layer, max pooling layers witha pool size of 2 are employed. Max pooling reduces thespatial dimensions of the feature maps, retaining the mostsalient features and discarding irrelevant or less importantinformation. This downsampling operation helps to decreasecomputational complexity and spatial resolution.To prevent overfitting and enhance generalization, a dropoutlayer is inserted before the fully connected layer. With adropout rate of 25% or 0.25, this layer randomly sets afraction of the input units to zero during each training iteration.By doing so, dropout encourages the network to learn morerobust and diverse representations, reducing the risk of over-reliance on specific features and improving the models abilityto generalize to unseen data.The fully connected layer following the dropout layer com-prises 80 neurons. This layer is responsible for aggregatingthe learned features and capturing high-level representations,enabling the model to capture complex relationships and makemore informed predictions based on the extracted features.The output layer of the custom CNN architecture employsa sigmoid activation function and consists of a single neuron.With this setup, the model can provide a probability estimatefor the input belonging to the normal or abnormal class. Thesigmoid activation function ensures that the output value fallswithin the range of 0 to 1, representing the probability of theinput being classified as abnormal or normal, respectively.Sigmoid: Y =11 + eX(9)The input shape of the custom CNN is defined as224x224x3, reflecting the dimensions of the spectrogram im-ages. The 3 corresponds to the three color channels (RGB) ofthe spectrograms, allowing the model to learn from the multi-channel input and capture color information if present.Overall, the custom CNN architecture, with its specific con-figuration of convolutional layers, activation functions, poolinglayers, dropout layer, fully connected layer, and output layer, isdesigned to extract discriminative features from spectrogramsand enable accurate classification of heart sounds into thenormal and abnormal classes.2) Transfer Learning: Transfer learning is a technique thatutilizes pre-trained models on large-scale datasets to improvethe performance of a model on a related task [19]. In this work,the VGG16 model is used as a pre-trained model for heartsound classification. VGG16 is a deep convolutional neuralnetwork trained on the ImageNet dataset [20].The pre-trained VGG16 [21] model is used as a featureextractor for heart sound spectrograms. The lower layers ofVGG16 capture low-level features like edges and textures,while the higher layers capture more abstract features. Byleveraging these learned filters, meaningful features can beextracted from the heart sound spectrograms.MobileNetV2 is a CNN architecture that is based on theinverted residual structure. The inverted residual structure isa way of organizing CNN layers that allows the networkto achieve high accuracy with fewer parameters and fewercomputational resources.Similarly, the MobileNetV2 architecture [22] consists of aseries of inverted residual blocks. Each inverted residual blockconsists of three layers:A 1x1 convolution layer that reduces the number of chan-nels. A depthwise separable convolution layer that performsconvolutions on each channel independently. A 1x1 convolu-tion layer that increases the number of channels back to theoriginal value. The inverted residual blocks are arranged in aseries of stages. Each stage consists of a number of invertedresidual blocks, and the number of inverted residual blocks ineach stage increases as the network goes deeper.The fully connected layers of both architectures are replacedwith new layers for heart sound classification. These new lay-ers include a global average pooling layer and fully connectedlayers specific to the classification task.During training, the weights of the pre-trained VGG16 andMobilenetV2 layers are kept frozen, and only the weights ofthe newly added layers are updated. This allows the model tolearn task-specific representations while retaining the knowl-edge from the pre-trained model.The Adam optimizer is used to train the model. Adam com-bines the Adaptive Moment Estimation and Root Mean SquarePropagation methods, dynamically adapting the learning ratefor each parameter and improving the training processs con-vergence speed and robustness.By employing transfer learning with the VGG16 and Mo-bileNetV2 models and fine-tuning the models on heart soundspectrograms using the Adam optimizer, this approach aimsto benefit from the learned representations and generalizationcapabilities of pre-trained models to improve the heart soundclassification performance.3) PSO-Optimized CNN:Particle Swarm Optimization(PSO) is a nature-inspired optimization algorithm that is usedto find optimal solutions in complex search spaces [23]. Itis inspired by the social behavior of bird flocking or fishschooling, where individuals collectively work towards findingthe best solution.In PSO, a population of particles represents potential solu-tions in the search space. Each particles position corresponds

Fig. 4: Typical CNN block diagramsto a potential solution, and its velocity represents the directionand magnitude of movement in the search space. The particlescollectively explore the search space by updating their posi-tions and velocities based on their own experiences and theexperiences of the best-performing particles in the population.The update process in PSO consists of two main com-ponents: cognitive component and social component. Thecognitive component represents a particles tendency to movetowards its own best solution found so far, while the socialcomponent represents its tendency to move towards the bestsolution discovered by any particle in the population.Mathematically, the position and velocity of a particle areupdated as follows:Velocity update:V (t+1) = wV (t)+c1rand()(PbestX(t))+c2rand()(GbestX(t))where V(t) is the current velocity, w is the inertia weightthat controls the impact of the previous velocity, c1 and c2 arethe acceleration coefficients, Pbest represents the personal bestposition of the particle, X(t) is the current position, and Gbestis the best position found among all particles in the population.Position update:X(t + 1) = X(t) + V (t + 1)) The inertia weight w is usually decreased over iterationsto gradually reduce the impact of the previous velocity andallow more exploration in the early stages and exploitation inthe later stages of optimization. The acceleration coefficientsc1 and c2 control the influence of the personal and global bestpositions, respectively. They determine the balance betweenexploration and exploitation.The optimization process continues for a predefined numberof iterations or until a termination criterion is met, such asreaching a satisfactory solution or a maximum number ofiterations.IV. RESULTS AND DISCUSSIONThe research paper focuses on the automatic classification ofheart sounds or phonocardiograms (PCG) using deep learningtechniques. Three different models are investigated and com-pared in terms of their performance: Custom CNN, TransferLearning, and PSO-optimized CNN. The features used forthe models are spectrograms generated through the Short-Time Fourier Transform (STFT). The first model, CustomFig. 5: Learning Curve for Custom CNNCNN, achieved high precision (0.98) for identifying normalheart sounds, indicating a low false positive rate. However,the precision for abnormal heart sounds was lower (0.74),suggesting a higher false positive rate. In terms of recall, themodel performed well for abnormal heart sounds (0.98) butrelatively poorly for normal heart sounds (0.67), indicating ahigher false negative rate for normal heart sounds. The F1-scores for both classes ranged from 0.80 to 0.85, suggesting a

trade-off between precision and recall for normal heart sounds.The average F1-score for the Custom CNN model was 0.82.Fig. 6: Results for Custom CNNFig. 7: Obtained ResultsThe second model, Transfer Learning (VGG16), demon-strated high precision for both normal (0.95) and abnormal(0.97) heart sounds, indicating a low false positive rate. Sim-ilarly, the recall values were high for both normal (0.97) andabnormal (0.95) heart sounds, suggesting a low false negativerate. The F1-scores for both classes were 0.96, indicating awell-balanced performance between precision and recall. TheTransfer Learning model achieved the highest average F1-scoreof 0.96 among the three models. Similarly, The MobileNetV2model is a smaller and more efficient version of VGG16, andit is better suited for mobile devices. This approach achievedan average F1-score of 0.98, which is the highest score of allthe methods. The precision and recall for both classes are alsovery high, which suggests that this method is the most accurateat classifying PCGs.The third model, PSO-optimized CNN, achieved perfectprecision (1.00) for normal heart sounds, indicating no falsepositives. However, the precision for abnormal heart soundswas lower (0.65), suggesting a higher false positive rate. Therecall for abnormal heart sounds was high (1.00), indicatinga low false negative rate, but it was significantly lower fornormal heart sounds (0.48), suggesting a higher false negativerate. The F1-scores were 0.64 for normal heart sounds and 0.79for abnormal heart sounds, indicating an imbalance betweenprecision and recall. The PSO-optimized CNN model had thelowest average F1-score of 0.71 among the three models.The results of this study suggest that transfer learningis a promising approach for classifying PCGs. The transferlearning methods (VGG16 and MobileNetV2) achieved thebest results, with an average F1-score of 0.96 and 0.98,respectively. The custom CNN-based method also achieved agood score, with an average F1-score of 0.82. However, thePSO-optimized CNN method achieved the lowest score, withan average F1-score of 0.71.The high F1-scores for the transfer learning methods suggestthat these methods are able to achieve a good balance betweenprecision and recall. Precision is the ability to correctly identifythe positive class, while recall is the ability to correctly identifyall of the positive instances. A good balance between precisionand recall is important for medical applications, as it isimportant to avoid both false positives (incorrectly classifyinga normal PCG as abnormal) and false negatives (incorrectlyclassifying an abnormal PCG as normal).The high precision for the normal class in the PSO-optimized CNN method suggests that this method is very goodat avoiding false positives. However, the low recall for theabnormal class suggests that this method is more likely tomiss abnormal PCGs.Overall, the results of this study suggest that transferlearning is a promising approach for classifying PCGs. Theconfusion matrix for each techniqus are illsutrated in Figures11, 9 and 10.Fig. 8: Confusion Matrix for Custom CNNFor each optimization technique, we evaluated the perfor-mance of the optimized models using several performance

Fig. 9: Confusion Matrix for VGG16Fig. 10: Confusion Matrix for MobileNetV2metrics, including accuracy, overall score, sensitivity, preci-sion, and specificity. These metrics provided insights into themodels ability to classify heart sounds accurately and its bal-ance between correctly identifying positive cases (sensitivity)and negative cases (specificity) using Eq. 10,11,12 and 13,where TP is used for true positives, FP is used for falsepositives, and FN is used for false negative numbers.Accuracy =TP + TNTP + TN + FP + FN(10)Precision =TPTP + FP(11)Recall =TPTP + FN(12)F1 = 2 Precision RecallPrecision + Recall(13).Fig. 11: BenchmarkingV. CONCLUSIONIn this study, we explored the application of deep learn-ing models for automatic heart sound classification. Thecomparative analysis of Custom CNN, Transfer Learning(VGG16+mobileNetV2), and PSO-optimized CNN revealedvarying levels of performance. The Transfer Learning modelwith MobileNetV2 architecture, utilizing spectrograms as fea-tures, demonstrated the highest average F1-score (0.98) amongthe four models, indicating a well-balanced performance forboth normal and abnormal heart sounds. It outperformed theCustom CNN model, which showed a trade-off between pre-cision and recall for normal heart sounds. The PSO-optimizedCNN model exhibited a significant imbalance in performancebetween normal and abnormal heart sounds.Furthermore, the proposed work, when compared with pre-vious studies, achieved a higher accuracy (0.98) and balancedsensitivity and specificity (0.98) for heart sound classification.These results highlight the effectiveness of using spectrogramsand Transfer Learning in this context.To further enhance the classification performance of heartsounds, future work can explore the integration of VisionTransformers and Convolutional Autoencoders. Vision Trans-formers have shown promising results in various image clas-sification tasks, and their application to spectrogram analysisfor heart sound classification could provide valuable insights.Additionally, Convolutional Autoencoders can be utilized forunsupervised feature learning, enabling the models to capturemore informative representations of heart sounds. Integratingthese advanced techniques into the existing models couldpotentially improve the accuracy and robustness of heart soundclassification systems.

AuthorDatasetMethodAccuracySensitivitySpecificityM. Zabihi et al [24]Liu et al [13]MFCC+LPC+Ensemble ANN-86.9184.90J. Rubin et al [25]Liu et al [13]MFCC+ CNN-76.5093.1Khan et al [26]Liu et al [13]MFCC+ LSTM80.6883.2499.55Proposed workLiu et al [13]Spectrograms+MobileNetV298.098.098.0TABLE I: Benchmark with previous studiesREFERENCES[1] George A Mensah and David W Brown. An overview of cardiovasculardisease burden in the united states. Health affairs, 26(1):3848, 2007.[2] Thomas Gaziano, K Srinath Reddy, Fred Paccaud, Sue Horton, andVivek Chaturvedi. Cardiovascular disease. Disease Control Priorities inDeveloping Countries. 2nd edition, 2006.[3] Christopher C Imes and Frances Marcus Lewis.Family history ofcardiovascular disease (cvd), perceived cvd risk, and health-relatedbehavior: A review of the literature.The Journal of cardiovascularnursing, 29(2):108, 2014.[4] Karen Okrainec, Devi K Banerjee, and Mark J Eisenberg.Coronaryartery disease in the developing world.American heart journal,148(1):715, 2004.[5] Louis-Gilles Durand and Philippe Pibarot. Digital signal processing ofthe phonocardiogram: review of the most recent advancements. CriticalReviewsin biomedical engineering, 23(3-4), 1995.[6] SM Debbal and Fethi Bereksi-Reguig.Computerized heart soundsanalysis. Computers in biology and medicine, 38(2):263280, 2008.[7] Rangraj M Rangayyan and Richard J Lehner.Phonocardiogram sig-nal analysis: a review.Critical reviews in biomedical engineering,15(3):211236, 1987.[8] MS Obaidat.Phonocardiogram signal analysis: techniques and per-formance comparison. Journal of medical engineering & technology,17(6):221227, 1993.[9] Bin Xiao, Yunqiu Xu, Xiuli Bi, Junhui Zhang, and Xu Ma.Heartsounds classification using a novel 1-d convolutional neural network withextremely low parameter consumption. Neurocomputing, 392:153159,2020.[10] Zhao Ren, Nicholas Cummins, Vedhas Pandit, Jing Han, Kun Qian, andBjorn Schuller. Learning image-based representations for heart soundclassification. In Proceedings of the 2018 international conference ondigital health, pages 143147, 2018.[11] Fatemeh Safara, Shyamala Doraisamy, Azreen Azman, Azrul Jantan, andAsri Ranga Abdullah Ramaiah. Multi-level basis selection of waveletpacket decomposition tree for heart sound classification. Computers inbiology and medicine, 43(10):14071414, 2013.[12] Minh Tuan Nguyen, Wei Wen Lin, and Jin H Huang. Heart sound clas-sification using deep learning techniques based on log-mel spectrogram.Circuits, Systems, and Signal Processing, 42(1):344360, 2023.[13] Chengyu Liu, David Springer, Qiao Li, Benjamin Moody, Ricardo AbadJuan, Francisco J Chorro, Francisco Castells, Jose Millet Roig, IkaroSilva, Alistair EW Johnson, et al.An open access database for theevaluation of heart sound algorithms.Physiological measurement,37(12):2181, 2016.[14] SS Daud and R Sudirman. Butterworth bandpass and stationary wavelettransform filter comparison for electroencephalography signal. In 20156th international conference on intelligent systems, modelling and sim-ulation, pages 123126. IEEE, 2015.[15] Jonathan Le Roux, Hirokazu Kameoka, Nobutaka Ono, and ShigekiSagayama. Fast signal reconstruction from magnitude stft spectrogrambased on spectrogram consistency. In Proc. DAFx, volume 10, pages397403, 2010.[16] Abdelghani Djebbari and F Bereksi Reguig. Short-time fourier transformanalysis of the phonocardiogram signal.In ICECS 2000. 7th IEEEInternational Conference on Electronics, Circuits and Systems (Cat. No.00EX445), volume 2, pages 844847. IEEE, 2000.[17] Wen-kai Lu and Qiang Zhang. Deconvolutive short-time fourier trans-form spectrogram.IEEE Signal Processing Letters, 16(7):576579,2009.[18] Saad Albawi, Tareq Abed Mohammed, and Saad Al-Zawi. Understand-ing of a convolutional neural network. In 2017 international conferenceon engineering and technology (ICET), pages 16. Ieee, 2017.[19] Lisa Torrey and Jude Shavlik.Transfer learning.In Handbook ofresearch on machine learning applications and trends: algorithms,methods, and techniques, pages 242264. IGI global, 2010.[20] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In 2009 IEEEconference on computer vision and pattern recognition, pages 248255.Ieee, 2009.[21] Hussam Qassim, Abhishek Verma, and David Feinzimer. Compressedresidual-vgg16 cnn model for big data places image recognition.In2018 IEEE 8th annual computing and communication workshop andconference (CCWC), pages 169175. IEEE, 2018.[22] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov,and Liang-Chieh Chen.Mobilenetv2: Inverted residuals and linearbottlenecks. In Proceedings of the IEEE conference on computer visionand pattern recognition, pages 45104520, 2018.[23] Federico Marini and Beata Walczak. Particle swarm optimization (pso).a tutorial. Chemometrics and Intelligent Laboratory Systems, 149:153165, 2015.[24] Morteza Zabihi, Ali Bahrami Rad, Serkan Kiranyaz, Moncef Gabbouj,and Aggelos K Katsaggelos. Heart sound anomaly and quality detectionusing ensemble of neural networks without segmentation.In 2016computing in cardiology conference (CinC), pages 613616. IEEE, 2016.[25] Jonathan Rubin, Rui Abreu, Anurag Ganguli, Saigopal Nelaturi, IonMatei, and Kumar Sricharan.Classifying heart sound recordingsusing deep convolutional neural networks and mel-frequency cepstralcoefficients. In 2016 Computing in cardiology conference (CinC), pages813816. IEEE, 2016.[26] Faiq Ahmad Khan, Anam Abid, and Muhammad Salman Khan. Au-tomatic heart sound classification from segmented/unsegmented phono-cardiogram signals using time and frequency features.Physiologicalmeasurement, 41(5):055006, 2020.