International Journal of Computer Assisted Radiology and Surgery (2022) 17:413–425
https://doi.org/10.1007/s11548-021-02542-7

ORIGINAL ARTICLE

Automatic morphological classification of mitral valve diseases in echocardiographic images based on explainable deep learning methods

Majid Vafaeezadeh[1] · Hamid Behnam[1] · Ali Hosseinsabet[2] · Parisa Gifani[3]

Received: 1 June 2021 / Accepted: 30 November 2021 / Published online: 12 December 2021

Abstract

Purpose Carpentier's functional classification is a guide to explain the types of mitral valve regurgitation based on morphological features. There are four types of pathological morphologies, regardless of the presence or absence of mitral regurgitation: Type I, normal; Type II, mitral valve prolapse; Type IIIa, mitral valve stenosis; and Type IIIb, restricted mitral leaflet motion. The aim of this study was to automatically classify mitral valves using echocardiographic images.

Methods In our procedure, after the classification of apical 4-chamber (A4C) and parasternal long-axis (PLA) views, we extracted the systolic/diastolic phase of the cardiac cycle by calculating the left ventricular area. Six typical pre-trained models were fine-tuned with a 4-class model for the PLA and a 3-class model for the A4C views. As an additional contribution, to provide explainability, we applied the Gradient-weighted Class Activation Mapping (Grad-CAM) algorithm to visualize areas of echocardiographic images where the different models generated a prediction.

Results This approach conferred a proper understanding of where various networks "look" into echocardiographic images to predict the four types of pathological mitral valve morphologies. Considering the accuracy metric and Grad-CAM maps and by applying the Inception-ResNet-v2 architecture to classify Type II in the PLA view and ResNeXt50 architecture to classify the other three classes in the A4C view, we achieved an 80% rate of model accuracy in the test data set.

Conclusions We suggest an explainable, fully automated, and rule-based procedure to classify the four types of mitral valve morphologies based on Carpentier's functional classification using deep learning on transthoracic echocardiographic images. Our study results infer the feasibility of the use of deep learning models to prepare quick and precise assessments of mitral valve morphologies in echocardiograms. According to our knowledge, our study is the first one that provides a public data set regarding the Carpentier classification of MV pathologies.

Keywords DCNN · Echocardiography · Mitral valve morphology · Carpentier classification

Introduction

Among the most prevalent valvular heart diseases are those of the mitral valve (MV) [1]. If not treated, MV diseases result in cardiac dysfunction and even heart failure or death, in severe cases. Echocardiography is a frequently used test in the evaluation of MV abnormalities and their follow-up. In many situations, there may be a need for nonspecialist physicians in emergency departments, hospital wards, and intensive care units to make decisions based on echocardiographic findings.

Carpentier's functional classification was primarily developed for the classification of the mechanism of mitral valve regurgitation (MR) [2]. Nonetheless, it is reasonable to suppose that it could be applied to the morphological classification of the MV. Changes in the MV morphology due to pathological processes can be different. These differences can be prominent during the systolic or diastolic phase of the cardiac cycle.

According to Carpentier's functional classification, MR is divided into four types based on the leaflet motions of

✉ Hamid Behnam
behnam@iust.ac.ir; behnamh22@gmail.com

1 Biomedical Engineering Department, Iran University of Science and Technology, Tehran, Iran

2 Cardiology Department, Tehran Heart Center, Tehran University of Medical Sciences, Tehran, Iran

3 Medical Sciences and Technologies Department, Science and Research Branch, Islamic Azad University, Tehran, Iran

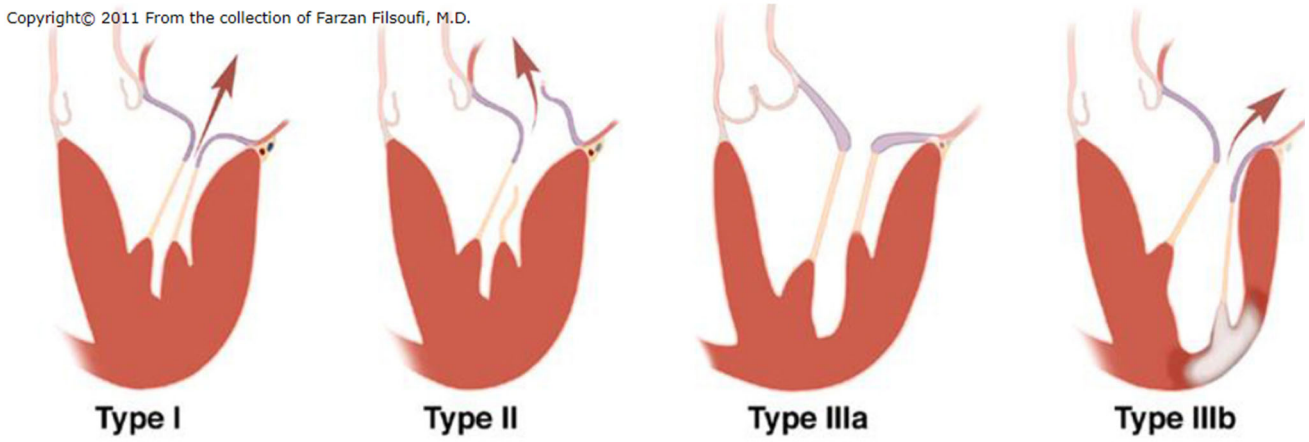

Fig. 1 The image depicts Carpentier's functional classification of mitral valve regurgitation [3]

the MV (Fig. 1). In Type I, the leaflet motions are normal, and MR occurs when there is annular dilatation. Perforated leaflets and cleft leaflets also fall into this classification. In Type II, the leaflet motions are excessive. This type, therefore, encompasses mitral valve prolapse (MVP). In Type III, the leaflet motions are restricted whether during both systole and diastole (Type IIIa) such as rheumatic mitral valve stenosis (MS) or during systole only (Type IIIb). Type IIIb is seen in MR secondary to ischemic or nonischemic cardiomyopathy (restricted mitral motion).

In MVP, as a prototype of Carpentier's functional classification Type II, the placement of the coaptation point happens above the mitral annular plane in the PLA view in the systolic phase, especially at end-systole. According to this definition, the placement of the coaptation point above the mitral annular plane in the A4C view is not considered MVP because of the saddle shape of the MV.

Automatic morphological classification of MV that has clinical implications can be applied in the education program of novice trainees or in settings where those educational cases are restricted or there are busy experts or in the automatization of processes.

Therefore, in this research, we assessed the feasibility of the fully automatic recognition of MV diseases in echocardiographic images through deep neural networks. For MVP recognition, we applied the PLA view; and for the other three classes (viz, normal, rheumatic MS, and MV leaflets with restricted motions), we analyzed the A4C view.

The main contributions of this work are as follows:

- Echocardiography video data collected and annotated in two views (PLA and A4C views) with different heart rates.
- Analyzing eight common deep convolutional neural network (DCNN) models for the automate classification of MV diseases.
- Visually verifying the performance of the DCNNs based on the Gradient-weighted Class Activation Mapping (Grad-CAM) method.
- Finally, the experimental results demonstrate that our proposed method, combining two PLA and A4C view's information in a rule-based manner, achieves the best performance in the classification of MV diseases in echocardiographic images.

Related works

With the emergence of artificial intelligence methods such as deep learning techniques, medical imaging seems to have experienced a significant revolution. Deep learning has recently become one of the most widely used and efficient methods for diagnosing and classifying various diseases in medical imaging [4].

Headway, albeit slow by comparison with some other medical modalities, has been made in the application of deep learning to echocardiography to facilitate the interpretation of echocardiographic images. Currently, deep learning methods in 2D echocardiography are mainly applied to the classification of views [5], recognition of prosthetic MV [6], the assessment of quality [7], the detection of end-diastolic and end-systolic frames [8], segmentation [9], the estimation of the LV ejection fraction [10], and the classification of cardiovascular diseases [11].

Several studies have been conducted on MV morphology analysis which was dichotomized based on two categories: non-deep learning and deep learning techniques. Sotaquira et al. [12] presented a non-deep learning semiautomatic algorithm for mitral annulus (MA) and mitral leaflet segmentation, and Pedrosa et al. [13] proposed a framework for quantitative assessment of MV morphology from the 3D

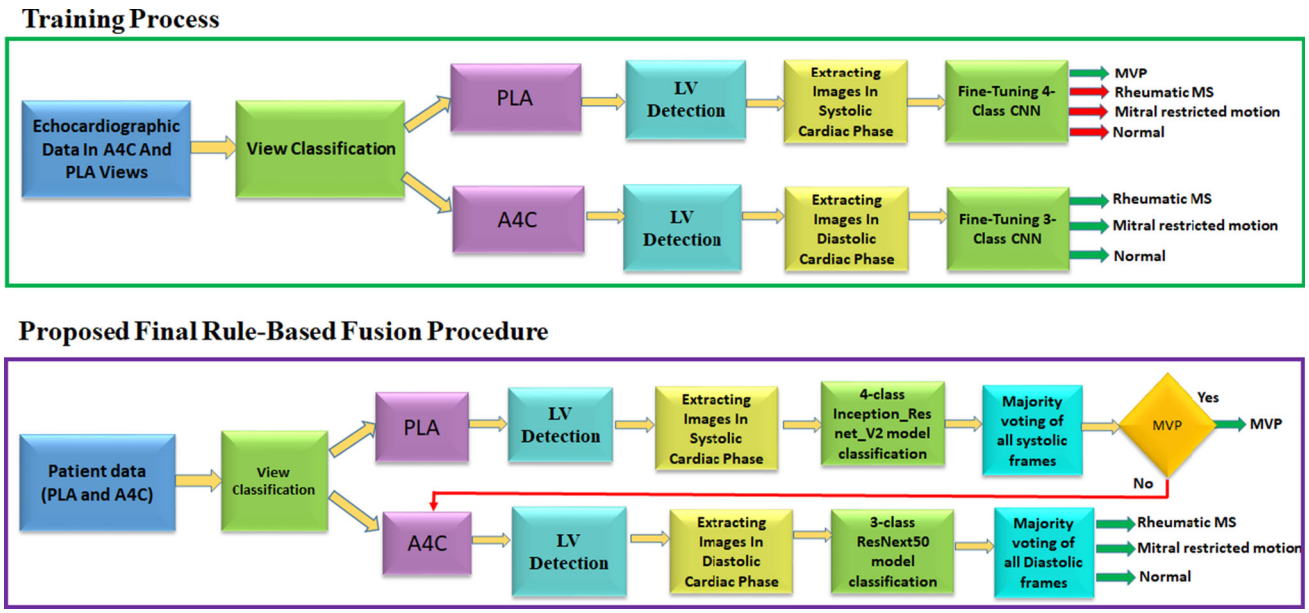

Fig. 2 The image depicts the flowchart of the training process and the proposed final rule-based fusion procedure

Transesophageal echocardiography (TEE) data. Andreassen et al. [14] proposed deep learning models for MA segmentation in 3D TEE, and the accuracy of prosthetic MV recognition by the deep learning algorithm was similar to that made by cardiologists [6].

Methods

In this part, we explain our suggested framework for MV diseases diagnosis in echocardiographic images. Because Type II (MVP) can be distinguished in the systolic phase of the PLA view, in the present study, this view was applied for MVP classification. For the other three types (viz, normal, rheumatic MS, and MV leaflets with restricted motions), the focus was placed on the A4C view. Figure 2 illustrates the training process and the proposed final rule-based fusion procedure.

In the training process, A4C and PLA views were automatically classified through the fine-tuning of a DCNN architecture. The systolic and diastolic phases of the cardiac cycle were extracted by calculating the LV area in PLA and A4C views, and all the frames of these phases were utilized in the next steps. Six popular 4-class DCNN architectures were fine-tuned to investigate which model could classify MVP from non-MVP (i.e., normal, rheumatic MS, and MV leaflets with restricted motions) by images within the systolic phase of the PLA view. In a parallel manner, 6 popular 3-class DCNN architectures were fine-tuned to train the classification of normal, rheumatic MS, and MV leaflets with restricted motion models by images in the diastolic phase of the A4C view. Afterward, between these six DCNN architectures, the best 4-class and 3-class models were selected through a comparison of model accuracy in the test data based on the majority voting ensemble of all the systolic and diastolic frames of PLA and A4C views, respectively. These best models were drawn upon for the final rule-based fusion procedure.

In the final rule-based fusion procedure, after view classification and cardiac phase recognition, for the PLA view, the Inception-ResNet-v2 model classified each case based on the majority voting ensemble of all the systolic frames. If the voting output was non-MVP, the decision was considered a 3-class A4C view model (ResNeXt50 [15]). In the following subsections the details of each step are described.

Data set

The study research protocol was approved by the institutional review board of Tehran Heart Center Hospital which is a well-known tertiary cardiovascular center in Iran. The transthoracic echocardiography for various patients with various pathologies is done in this high populated laboratory. The echocardiography laboratory provided us with echocardiography movies and echocardiography reports that were done by one of the authors between November 2018 and September 2020. Echocardiography was mostly done by (Philips, Affinity 70C, Andover, MA, USA) with an S5-1 probe with 41 ± 8 frames per second.

The selection was done in keeping with Carpentier's functional classification Type II, Type IIIa, Type IIIb, and normal. In other words, the data selected were on MVP, rheumatic MS, MV leaflets with restricted motions due to regional wall motion abnormalities or LV dilation, and MV leaflets

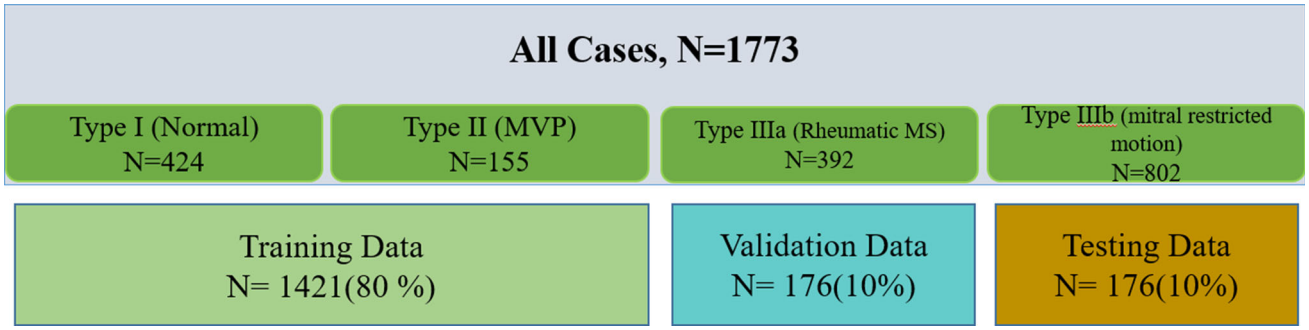

Fig. 3 The image presents the PLA and A4C views data set

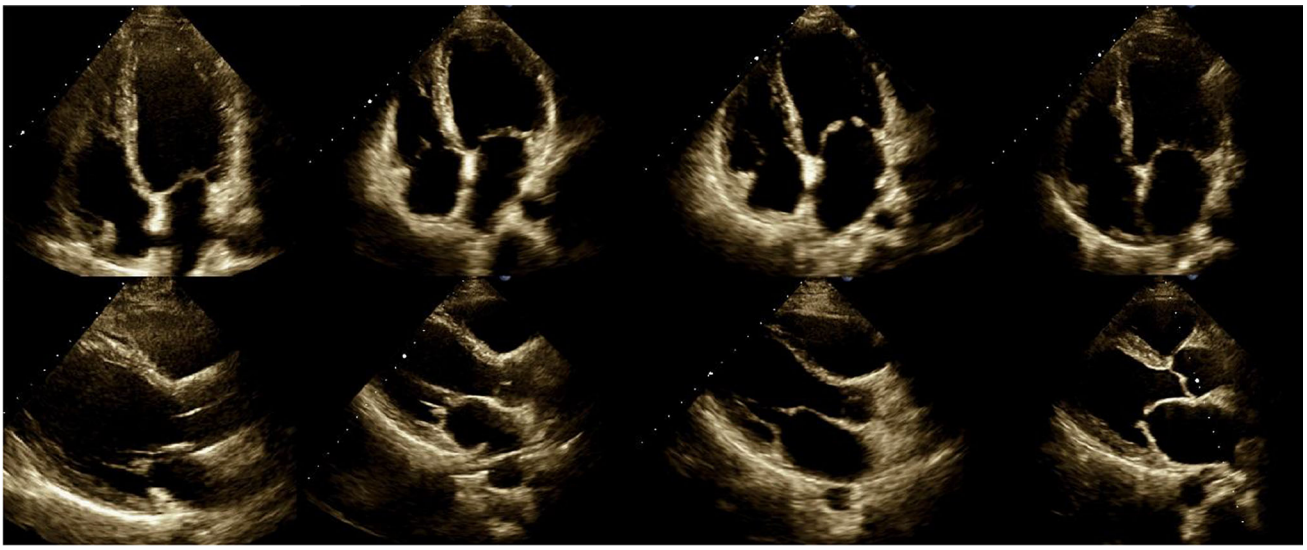

Fig. 4 The image presents a sample echocardiographic image of the different types of MV diseases in PLA and A4C views. The top row shows the A4C view, and the bottom row the PLA view for normal, MVP, Rheumatic MS, and mitral restricted motion from left to right

with normal motions. In Type IIIa pathological morphologies such as rheumatic MS, patients with prosthetic MVs were excluded. Images of patients with more than one pathology and the resultant change in the MV morphology were also excluded. Our network was not trained regarding mixed pathologies (morphologies) of the MV.

The A4C and PLA movies of each patient were extracted in Digital Imaging and Communications in Medicine (DICOM) format. Each movie was labeled by a highly experienced echocardiologist as normal, MVP, Rheumatic MS, and mitral restricted motion valve due to supporting MV apparatus pathologies such as left ventricular regional wall motion abnormalities or ischemic and non-ischemic dilated cardiomyopathies (type IIIb). In total, 1773 cases with A4C and PLA views were gathered, including 424 cases with normal MVs, 155 Type II cases (MVP), 392 Type IIIa cases (rheumatic MS), and 802 Type IIIb (mitral restricted motion) cases.

For all experiments, we divided our database into training (80%), validation (10%), and testing data sets (10%). The proportion of each class was kept in each set and the same division was kept for all experiments. Since we have two views (PLA and A4C) for each patient, we randomly do this division based on patient ID only once. We use this division to train, validation and test each of the A4C and PLA views. Therefore, for example, the patient ID whose A4C view is in the test subset, his/her PLA view is also in the test subset.

The number of samples in both views are illustrated in Fig. 3. A sample echocardiographic image of the different types of MV diseases in PLA and A4C views is displayed in Fig. 4.

A4C and PLA view classification

The essential first step in our automatic interpretation of echocardiograms was the classification of A4C and PLA views. In this stage, via transfer learning, the ImageNet pre-trained EfficientNetB0 [16] convolutional neural network model detected specific features in each view for the view classification of echocardiograms. In our procedure, we

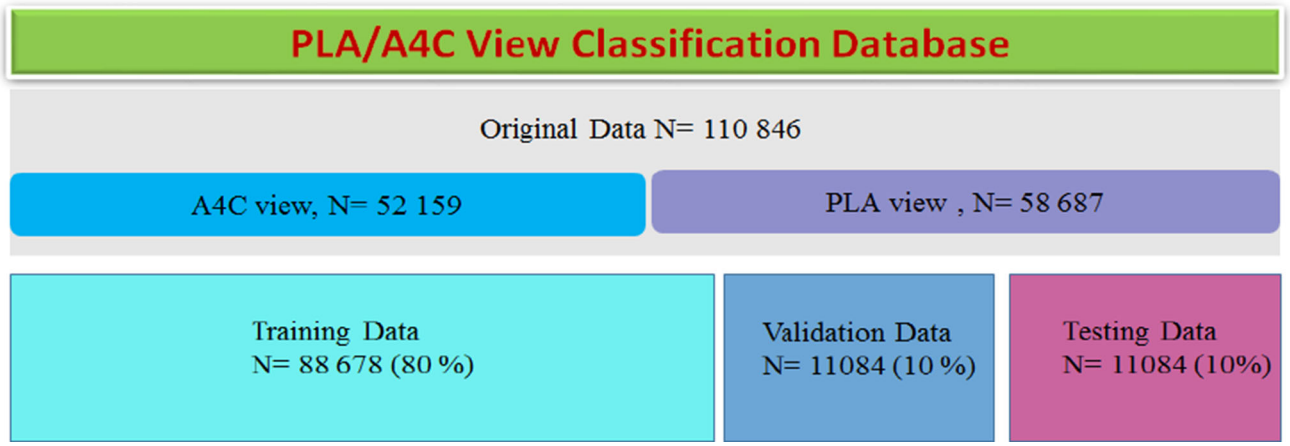

Fig. 5 The image presents the data set for the classification of PLA and A4C views

applied all 1773 cases which were described in the "Data set" section with 110,846 frames that included 52,159 frames for the A4C view and 58,687 frames for the PLA view (Fig. 5). The division of training/validation/test has been done according to the approach of the previous section.

LV Region detection and cardiac phase analysis

The cardiac cycle consists of two phases: diastole and systole. In the diastolic phase, while the heart muscle relaxes, the heart fills with blood. In the systolic phase, a strong contraction and blood pumping occur. In our study, these two phases were extracted for a better classification of MV diseases. The end-systolic and end-diastolic frames were identified in a cardiac cycle. The end-diastolic frame was specified as the largest LV area, whereas the end-systolic frame represented the smallest LV area [17].

In the proposed approach, the LV region was detected via deep learning methods prior to the estimation of the LV area. Subsequently, according to the area of the LV during a cardiac cycle, the corresponding end-systolic and end-diastolic frames were detected. Afterward, on the basis of each view, the desired phase was considered for further analysis.

In the present study, for the detection of the LV and the MV, the RetinaNet architecture was fine-tuned with 2000 images (1000 for A4C and 1000 for PLA view) which were taken randomly from of the 1773 cases that were described in the "Data set" section. For the training of each object (viz, the LV and the MV), we divided our database into training (80%), validation (10%), and testing data sets (10%).

Model selection for fine-tuning the CNN

After preparation of proper data in systolic and diastolic cardiac phases based on PLA and A4C views, the fine-tuning procedure was accomplished based on ImageNet pre-trained models. When data are not sufficient to train a DCNN, the transfer learning concept can be drawn upon. Transfer learning benefits from a pre-trained model (the DCNN model) on an extensive database and helps the model learn a target task (e.g., classification of MV diseases) with limited training data [18, 19].

In our study, eight state-of-the-art DCNNs were applied: EfficientNet (B0,B2,B4) [16], NasNetLarge [20], NasNetMobile [20], ResNeXt50 [15], Xception [21], and Inception-ResNet-v2 [22].

In all networks, categorical cross-entropy was utilized as a loss function and all the models were fine-tuned for 50 epochs using stochastic gradient descent (SGD) optimization with a Nesterov's momentum rate of 0.9, a learning rate of 0.0001, and a batch size of 32, and in the final classification layer, which utilized the Softmax nonlinearity function. During the training process, data augmentation was applied with up to 10° rotation randomly, as well as 10% height and width shifts. After the passing of 50 epochs of training, the diagnostic accuracy was computed with the use of the test set. These parameters were applied for both view classification and mitral valve diseases classification procedures.

Results

View classification

Our model was able to classify A4C and PLA echocardiographic views with a 99.4% rate of overall test accuracy, 99.59% precision, 99.34% recall, 99.17% F1-score, and 99.98% AUC without overfitting.

Cardiac phase analysis

For object detection based on RetinaNet method, we applied three different backbones: ResNet50, ResNet101 and ResNet152. Comparison of different backbones for Reti-

Fig. 6 The image depicts the automatic detection of the left ventricle and the MV. Left: PLA view, right: A4C view

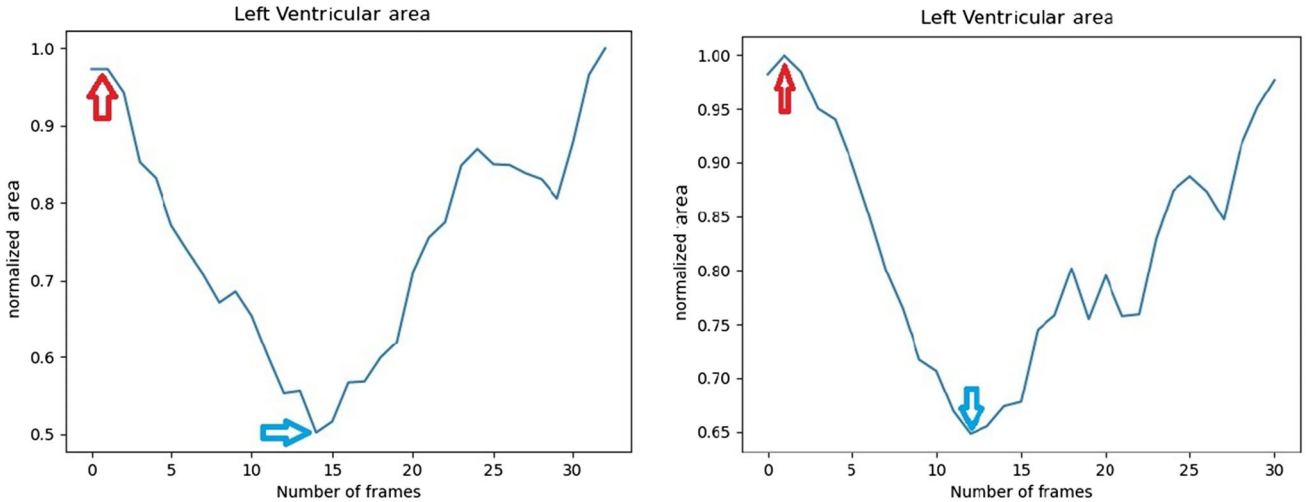

Fig. 7 The image shows the extraction of the end-diastolic frame (red arrow) and the end-systolic frame (blue arrow) in a cardiac cycle based on the left ventricular area in the A4C view

naNet conferred the best accuracy of the detection of the LV and the MV with the ResNet-152 architecture. Our model was able to detect the LV with 99.7% accuracy in the A4C view and 99.9% accuracy in the PLA view. Additionally, the MV detection accuracy rate was 99.2% in the A4C view and 99.4% in the PLA view (Fig. 6).

For the extraction of diastolic and systolic phases, the LV area during the cardiac cycle was detected. The area detected was then binarized by Otsu's adaptive thresholding. Next, the number of 0 pixels of the LV part was calculated, and its area was estimated. The maximum and minimum LV areas can be considered end-diastolic and end-systolic frames, respectively. Figures 7 and 8 illustrate samples of variations in the LV area in a cardiac cycle in A4C and PLA views, correspondingly.

For convenience, the detected end-systolic frame was focused upon. The previous frames of the end-systolic frame were regarded as the systolic phase, and the frames after the end-systolic frame to the end frame of the cardiac cycle were considered the diastolic phase. Afterward, the proper frames of cardiac diastolic and systolic phases for each case were extracted in four classes to feed the different state-of-the-art CNN models for the classification of MV diseases.

The 4-class model in the PLA view

The proposed framework fine-tuned six different state-of-the-art 4-class DCNN architectures on echocardiographic images for the differentiation between MVP and non-MVP in the PLA view in the systolic phase of the cardiac cycle. Figure 9 shows accuracy in the training sets of the six different 4-class and 3-class DCNN models in PLA and A4C views.

Table 1 illustrates classification metrics include recall, precision, the F1 score, and accuracy on the test data set, which attests to the difficulty in distinguishing between the

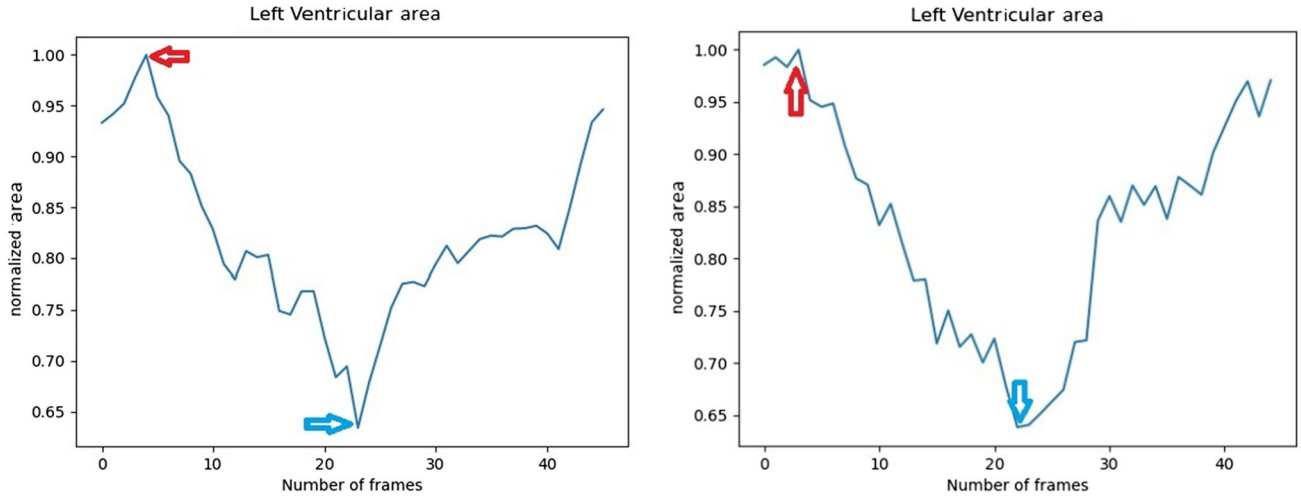

Fig. 8 The image demonstrates the extraction of the end-diastolic frame (red arrow) and the end-systolic frame (blue arrow) in a cardiac cycle based on the left ventricular area in the PLA view

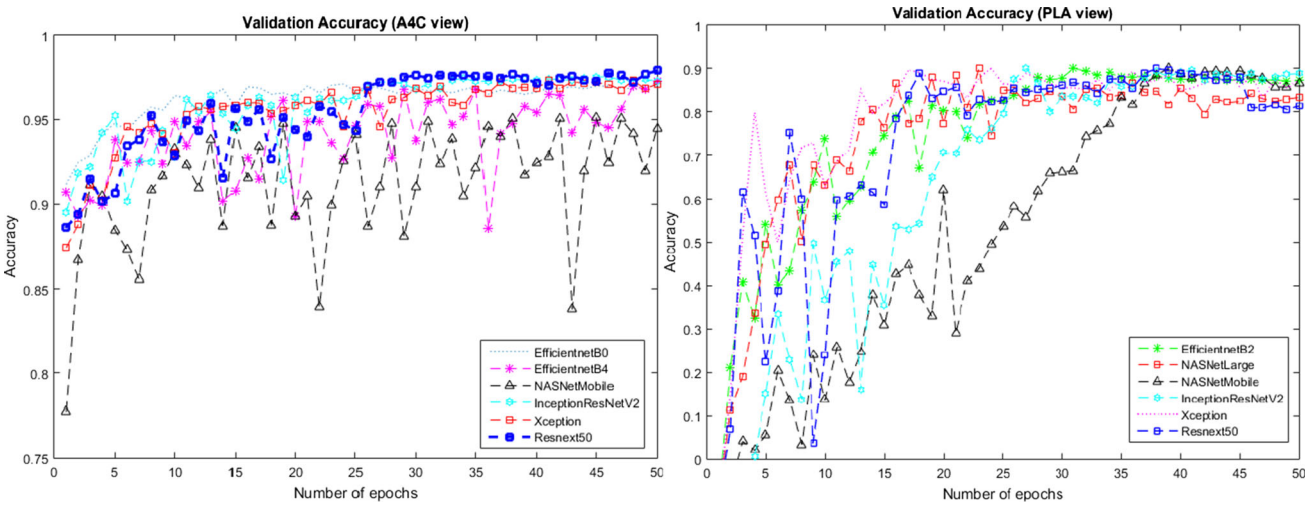

Fig. 9 The image demonstrates accuracy in the training sets for the training of the different DCNN models. Left: 3-class for the A4C view, Right: 4-class for the PLA view

four classes and takes into account the PLA/A4C classification and ED/ES detection. In the PLA view, of all the architectures, the Inception-ResNet-v2 framework yielded the highest accuracy in the classification of MVP as opposed to the other classes in echocardiographic images.

For a detailed comparison of the identification of frequent confusion in distinguishing between MVP and the other classes, the confusion matrices of the PLA view are illustrated in Fig. 10.

Although the false acceptance rate of NasNetLarge was smaller than that of Inception-ResNet-v2, the total accuracy of the proposed method was better by considering Inception-ResNet-v2 as the best model in the PLA view. Accordingly, in the present study, the Inception-ResNet-v2 4-class model was utilized for the detection of MVP in the PLA view and another classifier was employed for the other 3 classes, which is described in the next section.

The 3-class model in the A4C view

After decision-making in regard to MVP in the PLA view, the focus was shifted onto non-MVP. The A4C view of the non-MVP cases was analyzed in a 3-class task to classify the images into normal, rheumatic MS, and mitral restricted motion in the systolic phase of the cardiac cycle. Table 2 illustrates the results of common classification metrics for the assessment of the 3-class model in the A4C view in the test data set and takes into account the PLA/A4C classification and ED/ES detection.

As is shown in Table 2, in the A4C view, ResNeXt50, EfficientNetB4, and Inception-ResNet-v2 achieved the first three

Table 1 Results of common classification metrics for the evaluation of the 4-class (normal, MVP, Rheumatic MS, and mitral restricted motion) model in the PLA view in the test data

Model	Precision	Recall	F1_score	Kappa_score	Accuracy
EfficientNetB2	0.6372	0.5655	0.5820	0.4117	0.59
Inception_ResNet_v2	0.6704	0.6873	0.6697	0.5441	**0.69**
NasNetLarge	0.6002	0.6032	0.5871	0.4123	0.58
NasNetMobile	0.6821	0.5146	0.5508	0.3755	0.59
ResNeXt50	0.5270	0.5086	0.47047	0.2905	0.48
Xception	0.5644	0.5918	0.5623	0.4539	**0.61**

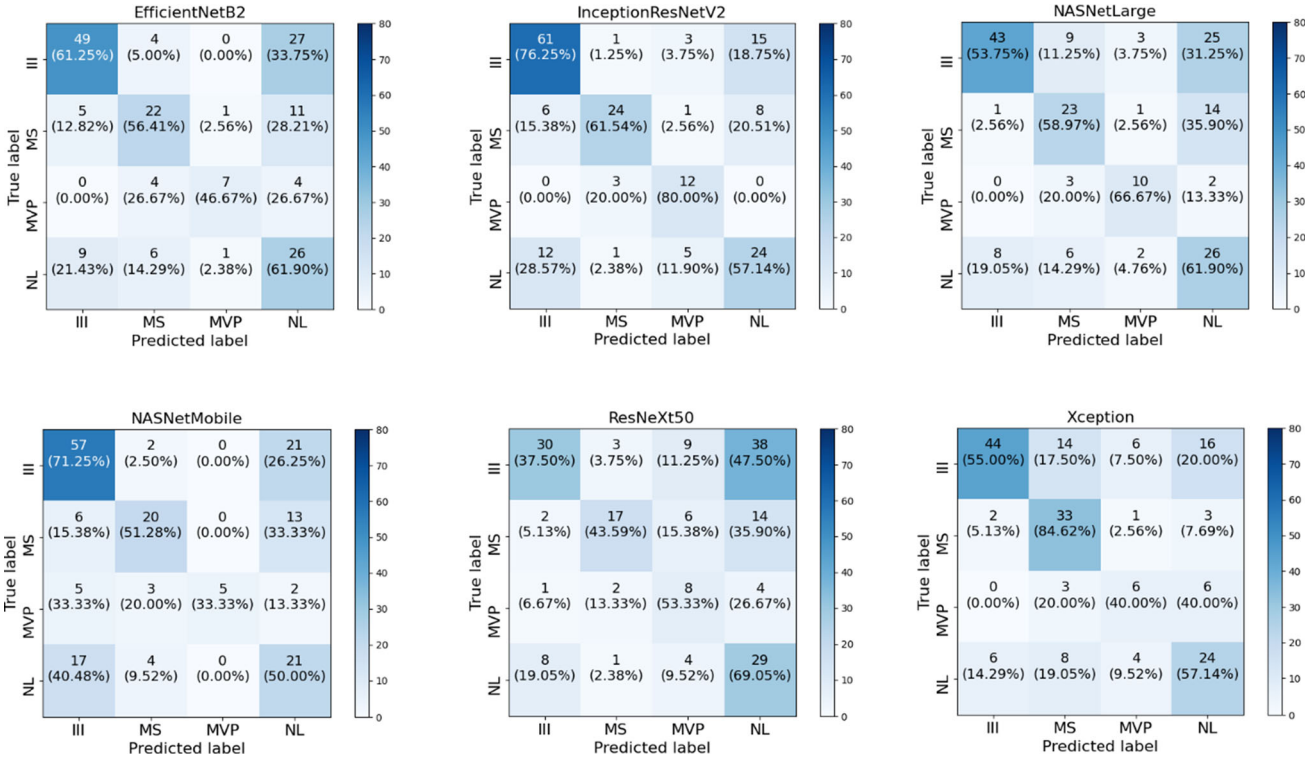

Fig. 10 The image depicts the results of the confusion matrix for the 4-class (normal, MVP, MS, and mitral restricted motion) model in the PLA view in the test data

Table 2 Results of common classification metrics for the evaluation of the 3-class (normal, MS, and mitral restricted motion) model in the A4C view in the test data

Model	Precision	Recall	F1_score	Kappa_score	Accuracy
EfficientNetB0	0.7164	0.6563	0.6762	0.5021	0.70
EfficientNetB4	0.7105	0.6800	0.6855	0.5351	0.72
Inception_ResNet_v2	0.7519	0.7035	0.7210	0.5615	**0.73**
NasNetMobile	0.6869	0.6879	0.6839	0.4948	0.68
ResNeXt50	0.8034	0.7794	0.7892	0.6765	**0.80**
Xception	0.7395	0.7397	0.7277	0.5706	0.72

ranks in accuracy metrics for distinguishing the three classes. These confusion matrices (Fig. 11) denoted the reliability of ResNeXt50, Inception-ResNet-v2 and EfficientNetB4 for the A4C view. The next section presents the additional analyses conducted with the Grad-CAM algorithm in the current investigation to achieve higher accuracy.

Grad-CAM visualization of trained CNNs

While deep learning techniques have delivered unprecedented accuracy and speed in the application of computer image analysis, such as classification, object detection, and segmentation, one of their major drawbacks is model interpretability, which is a primary constituent in model

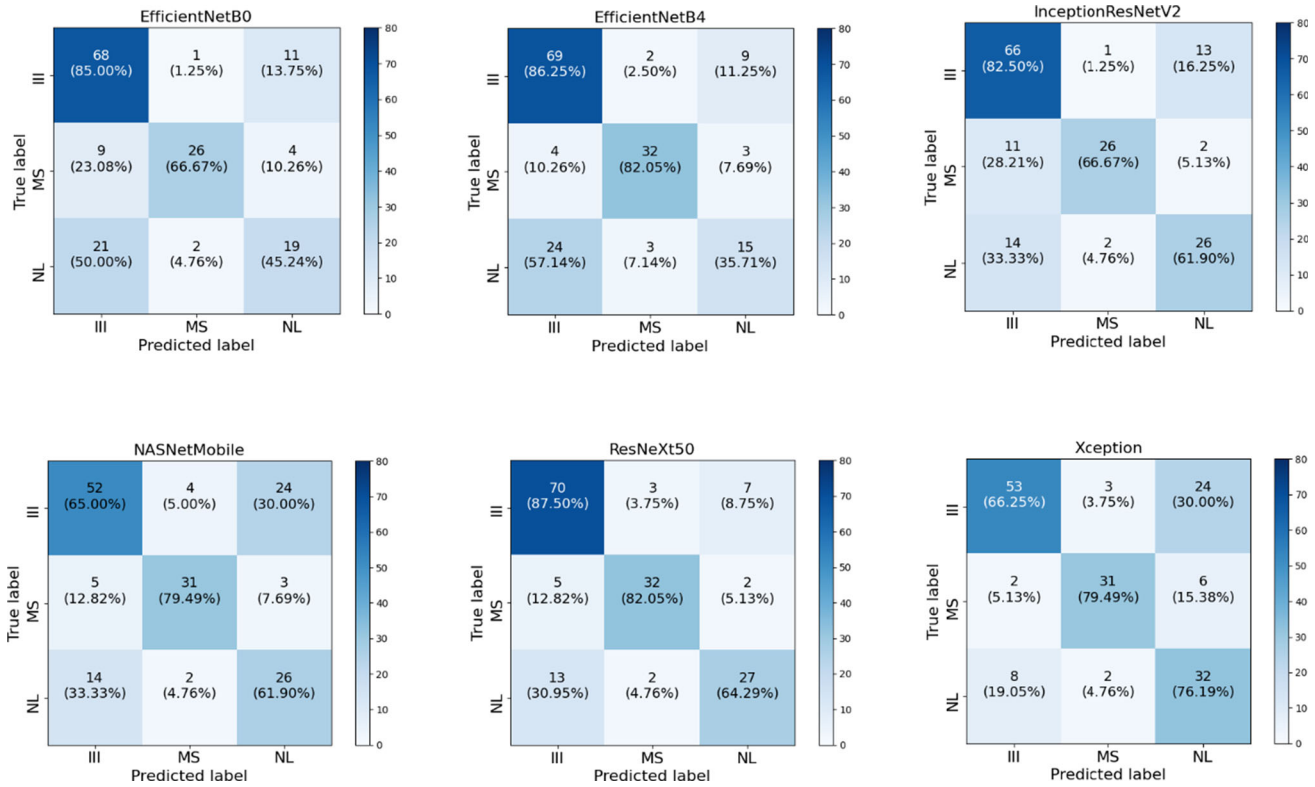

Fig. 11 The image presents the results of the confusion matrix for the 3-class (normal, Rheumatic MS, and mitral restricted motion) model in the A4C view

understanding and model debugging that currently limits the utilization of these techniques in critical contexts (e.g., medicine). CNNs are treated as "black box" methods; as a result, it may not always be prudent to confirm that the network is considering and activating the correct locations in the input image or to confirm which series of neurons are activated in the forward pass during inference/prediction.

The Grad-CAM [23] algorithm can visualize the class activation maps of trained CNNs. This concept increases the trust in the decisions of the trained model. Figures 12, 13 and 14 show the Grad-CAM randomly selected image samples of the three classes. Corresponding to the accuracy results in Table 2, as is expected in some trained models such as EfficientNetB0 and NasNetMobile, the network exhibited low activation around the MV in the image, suggesting that the network had failed to learn the underlying patterns properly. This can be the reason for the low accuracy of the classification of MV diseases.

In contrast, for ResNeXt50 and EfficientNetB4, the network focused on the LV and MV regions. According to the results of the previous section, presented in Table 2, Inception-ResNet-v2 was one of the top three models. Still, as is shown in Figs. 12, 13 and 14, the focal area of this model was scattered in the image, and it focused on more expansive areas rather than on the MV. As is demonstrated in Figs. 12 and 13, the focal area of EfficientNetB4 concerning the visualization of Type I (normal) and Type IIIa (Rheumatic MS) was on the MV; nevertheless, Fig. 14 shows that the focal area of the model vis-à-vis Type IIIb was the right atrium and the right ventricle. The best model to focus on the MV area was ResNeXt50. Figure 12 demonstrates that with respect to Type I (normal), while ResNeXt50 focused on the right ventricle slightly, its central area of interest was the MV. Therefore, of all the models, ResNeXt50 exhibited the highest focus on the MV region, with the results practically confirmed and aligned with the knowledge of the aforementioned cardiologist.

With regard to the models with almost similar results for the test samples by the Grad-CAM algorithm (top 3 models from Table 2), the explainability of the models was investigated. ResNeXt50 was the best explainable model with attention to a meaningful region of interest; it was, thus, a more appropriate choice than the mode of the blind use of other deep network models. The ability of this model to classify each of the 3 classes was comparable with the medical knowledge of the cardiologist.

Rule-based fusion

As was mentioned in the previous section, in the present study, PLA and A4C view information was aggregated in a

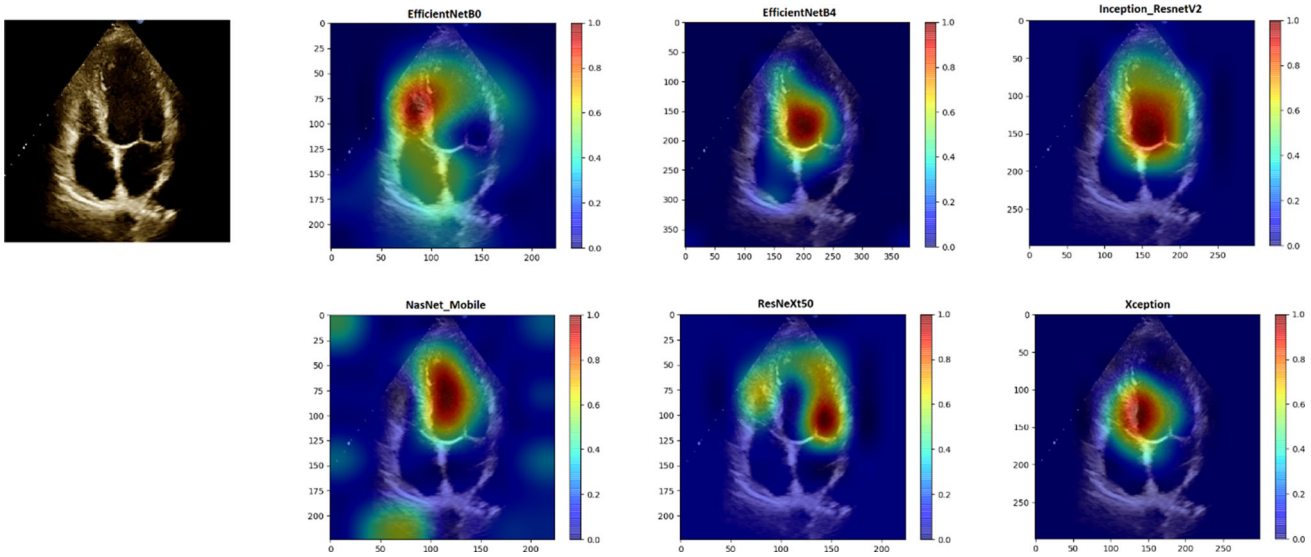

Fig. 12 The image presents the visualization of activation maps applied to an A4C image for Type IIIb using Grad-CAM for all the trained DCNN architectures. This heat map verifies the region of interest of the DCNN in the image

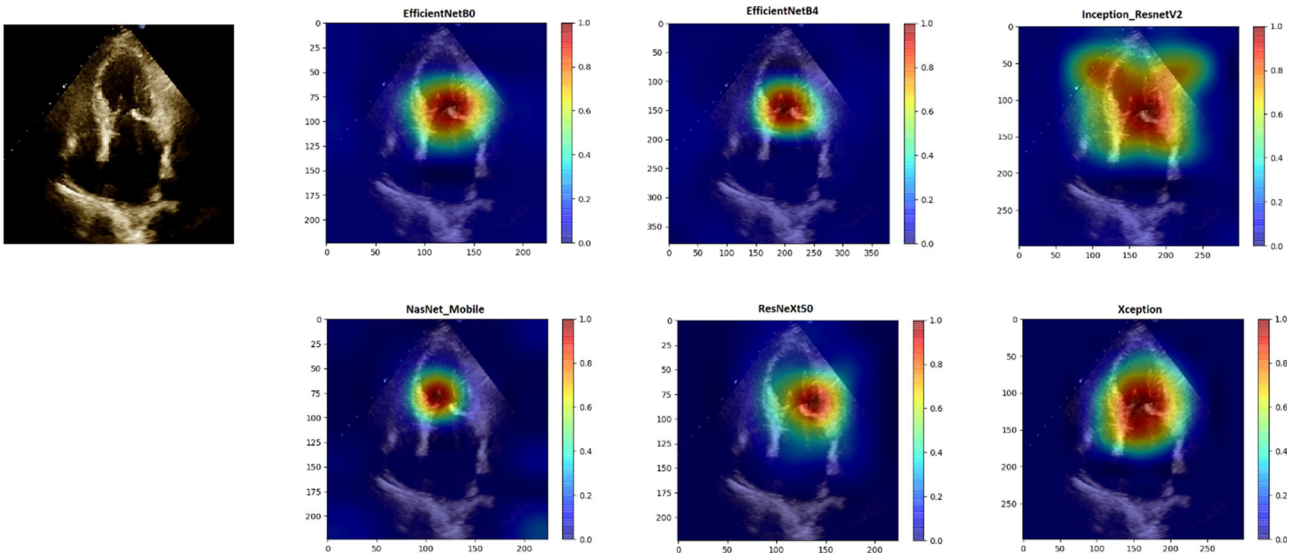

Fig. 13 The image shows the visualization of activation maps applied to an A4C image for Type IIIa using Grad-CAM for all the trained DCNN architectures. This heat map verifies the region of interest of the DCNN in the image

rule-based manner. In the PLA view, the final label of each case was determined based on the majority voting ensemble of all the systolic frames. If the result was non-MVP, the decision was considered a 3-class A4C view model.

The results of common classification metrics were displayed by combining the result from each of the previous two sections and taking into account the PLA/A4C classification and ED/ES detection (Table 3). For the detection of MVP in the PLA view, Inception-ResNet-v2 and ResNeXt50 were applied to identify and classify the three classes. The total accuracy rate for the four classes was 80%.

Discussion

In our suggested approach, eight state-of-the-art pre-trained networks were fine-tuned on the target task given the paucity of training instances to differentiate between MV diseases in echocardiographic images in A4C and PLA views. To improve classification accuracy, we developed a rule-based procedure based on morphological information regarding MVP (Type II) in the systolic phase of the PLA view and three other types in the diastolic phase of the A4C view.

We ensured the success of our approach by implementing several different procedures for training networks in terms

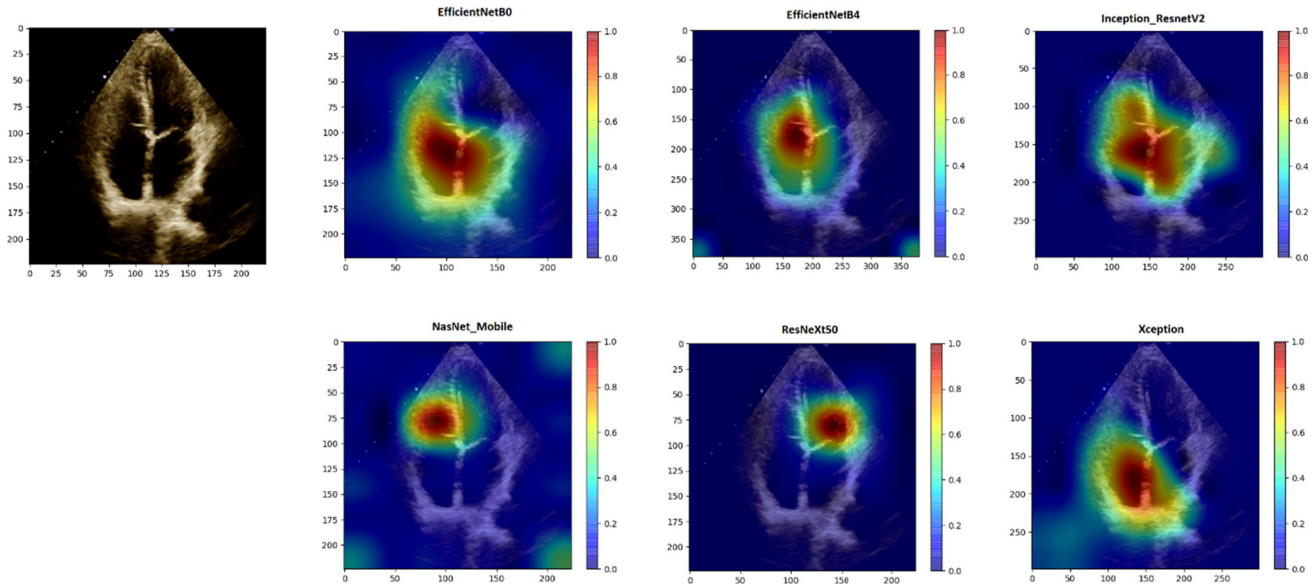

Fig. 14 The image demonstrates the visualization of activation maps applied to an A4C image for Type I using Grad-CAM for all the trained DCNN architectures. This heat map verifies the region of interest of the DCNN in the image

Table 3 Results of common classification metrics for the evaluation of the rule-based fusion

	Precision	Recall	F1_score	Kappa_score	Accuracy
Rule-based fusion	0.7731	0.7845	0.7708	0.7065	0.80

of the number of input images, the number of output classes, and the number of frames for majority voting. We considered several procedures on the data utilized for training and testing processes. In all these procedures, we considered one cardiac cycle for each case. Experimental results showed that fine-tuning of a DCNN based on ImageNet pre-trained models gave better results as compared to training from scratch.

In the first procedure, we achieved a light and fast model by training 4-class models through the use of the first single frame (this frame corresponded to the end-diastolic frame of the cardiac cycle) of each case in both PLA and A4C views separately. We also performed this test on the first frame. In this method, the best model accuracy rates on the test data in A4C and PLA views were 60% and 57%, respectively.

The second procedure was the averaging of all the frames of a cycle to create more informative single frames for the training, validating, and testing of the data set. In this procedure, we trained 4-class models using the average single frames in both PLA and A4C views separately and achieved the best model accuracy rates of 64% and 60% for A4C and PLA views, respectively.

In the third procedure, the model input comprised all frames in 1 cardiac cycle from each case, and the model output was all 4 Carpentier's functional classes. In the testing process, we reported the final label based on the majority voting of all the frames. In this procedure, the best model accuracy rates on the test data in A4C and PLA views were 66% and 63%, correspondingly. Nonetheless, the disadvantage of this solution was its low accuracy for the detection of the MVP class in the test samples.

In the fourth procedure, we automatically extracted diastolic and systolic phases, based on the estimation of the detected LV area during the cardiac cycle. We trained 4-class DCNN architectures with diastolic phase frames for the A4C view and systolic phase frames for the PLA view, and we achieved model accuracy rates of 71% and 69%, respectively.

Another approach to this classification problem was performing subclass and hierarchical procedures. This approach enabled us to identify the normal type from abnormalities before we could classify the other three classes of abnormalities. With this method, normal cases can be identified, and unhealthy cases can be examined more closely by artificial intelligence approaches. To that end, we trained a 2-class model: the normal class and the abnormal class (viz, MVP, rheumatic MS, and MV leaflets with restricted motions), and the network input was the total frames of each case. For each of the 2 PLA and A4C views, we trained the model separately. In this procedure, the best model accuracy rates on the test data in the A4C and PLA views were 68% and 65%, respectively.

Concerning the subclass and hierarchical perspective on the better diagnosis of the MVP class, we considered another solution. The MVP class is better classified in the PLA view and the systolic phase by cardiologists. Accordingly, in the fifth procedure, we trained 2-class DCNN architectures for MVP cases against the three other classes (viz, normal, rheumatic MS, and MV leaflets with restricted motions) and subsequently trained 3-class DCNN architectures to distinguish between the 3 classes of normal, rheumatic MS, and MV leaflets with restricted motions. An examination of the confusion matrix in the first stage showed little improvement in the accuracy of the classification. In general, the accuracy was low compared with the 4-class DCNN architectures for A4C and PLA views.

For more confidence, we carried out another procedure whereby the input of the networks was the bounding box of the MV and the focus of the network was only on the features and characteristics of this area. Via this approach, we extracted the location of the MV according to the determination of end-systolic/end-diastolic frames with the use of ventricular area measurement. We trained 8 DCNN architectures for 2-class, 3-class, and 4-class models based on the previous procedures. Still, the model accuracy failed to better that of the other models.

Finally, by summarizing the results of each of the above procedures and the cardiologist's recommendation, we applied both views of patients. In the final rule-based fusion procedure, the first step was the automatic classification of A4C and PLA views. Afterward, we extracted systolic and diastolic cardiac cycle phases by calculating the LV area in PLA and A4C views. The subsequent application of the 4-class Inception-ResNet-v2 architecture to classify Type II in the PLA view and the 3-class ResNeXt50 architecture to classify the other 3 classes in the A4C view helped us achieve an 80% rate of model accuracy in the test data set.

We applied the Grad-CAM algorithm to visualize areas of echocardiographic images where the different models generated a prediction. Corresponding to the accuracy results in Table 2, as is expected in some trained models such as EfficientNetB0 and NasNetMobile, the network exhibited low activation around the MV in the image, suggesting that the network had failed to learn the underlying patterns properly. This can be the reason for the low accuracy of the classification of MV diseases.

In contrast, for ResNeXt50 and EfficientNetB4, the network focused on the LV and MV regions. According to the results of the previous section, presented in Table 2, Inception-ResNet-v2 was one of the top 3 models. Still, as is shown in Figs. 12, 13 and 14, the focal area of this model was scattered in the image, and it focused on more expansive areas rather than on the MV. As is demonstrated in Figs. 12 and 13, the focal area of EfficientNetB4 concerning the visualization of Type I (normal) and Type IIIa (Rheumatic MS) was on the MV; nevertheless, Fig. 14 shows that the focal area of the model vis-à-vis Type IIIb was the right atrium and the right ventricle. The best model to focus on the MV area was ResNeXt50. Figure 12 demonstrates that with respect to Type I (normal), while ResNeXt50 focused on the right ventricle slightly, its central area of interest was the MV. Therefore, of all the models, ResNeXt50 exhibited the highest focus on the MV region, with the results practically confirmed and aligned with the knowledge of the aforementioned cardiologist.

With regard to the models with almost similar results for the test samples by the Grad-CAM algorithm (top 3 models from Table 2), the explainability of the models was investigated. ResNeXt50 was the best explainable model with attention to a meaningful region of interest; it was, thus, a more appropriate choice than the mode of the blind use of other deep network models. The ability of this model to classify each of the 3 classes was comparable with the medical knowledge of the cardiologist.

In our proposed method, the diastolic and systolic periods are automatically detected without dependency on electrocardiographic signals that is beneficial for low-cost portable hand-held echocardiographic devices.

In our study, the differentiation between normal and mitral restricted motion was not possible. Also, the concomitant analysis of PLA and A4C views was not done in one model.

Future research can aim at the analysis of images in various echocardiography views, and full training a CNN network without transfer learning by defining and building new network architecture. Future research could also draw upon temporal and spatial video echocardiographic information and video deep learning methods. It seems that our research can provide a basis for new approaches for the diagnosis of cardiovascular disorders.

Conclusions

In this research, we suggested an explainable, fully automated deep learning procedure drawing upon TTE images to classify MV morphologies based on Carpentier's functional classification.

The results of this study imply that deep learning models can be applied for rapid, accurate, and consistent evaluation of echocardiography movies regarding MV pathologies that accompany distorted MV morphologies. According to our knowledge, our study is the first one that provides a public data set regarding the Carpentier classification of MV pathologies. Moreover, we show the explainability of our DCNN model by exploiting activation areas in echocardiographic images in the discrimination between Carpentier's functional classifications of MV diseases.

Funding Not applicable.

Data availability The datasets generated during the current study are available from the corresponding author on reasonable request at https://github.com/medical-dataset/Mitral-Valve-Echocardiography.

Code availability Code for data cleaning and analysis are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

Ethical approval Ethical approval for this study was granted by the Ethics Committee of Tehran Heart Center, Iran.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Bonow RO, O'Gara PT, Adams DH, Badhwar V, Bavaria JE, Elmariah S, Hung JW, Lindenfeld J, Morris AA, Satpathy R (2020) 2020 Focused update of the 2017 ACC expert consensus decision pathway on the management of mitraláregurgitation: a report of the american college of cardiology solution set oversight committee. J Am Coll Cardiol 75(17):2236–2270. https://doi.org/10.1016/j.jacc.2020.02.005
2. Carpentier A (1983) Cardiac valve surgery—the "French correction." J Thorac Cardiovasc Surg 86(3):323–337. https://doi.org/10.1016/S0022-5223(19)39144-5
3. Carpentier A, Adams DH, Filsoufi F (2011) Carpentier's Reconstructive Valve Surgery E-Book. Elsevier Health Sciences
4. Fourcade A, Khonsari R (2019) Deep learning in medical image analysis: a third eye for doctors. J Stomatol Oral Maxillofac Surg 120(4):279–288. https://doi.org/10.1016/j.jormas.2019.06.002
5. Østvik A, Smistad E, Aase SA, Haugen BO, Lovstakken L (2019) Real-time standard view classification in transthoracic echocardiography using convolutional neural networks. Ultrasound Med Biol 45(2):374–384. https://doi.org/10.1016/j.ultrasmedbio.2018.07.024
6. Vafaeezadeh M, Behnam H, Hosseinsabet A, Gifani P (2021) A deep learning approach for the automatic recognition of prosthetic mitral valve in echocardiographic images. Comput Biol Med 133:104388–1
7. Ge R, Yang G, Chen Y, Luo L, Feng C, Ma H, Ren J, Li S (2019) K-net: Integrate left ventricle segmentation and direct quantification of paired echo sequence. IEEE Trans Med Imaging 39(5):1690–1702
8. Jafari MH, Girgis H, Van Woudenberg N, Liao Z, Rohling R, Gin K, Abolmaesumi P, Tsang T (2019) Automatic biplane left ventricular ejection fraction estimation with mobile point-of-care ultrasound using multi-task learning and adversarial training. Int J Comput Assist Radiol Surg 14(6):1027–1037. https://doi.org/10.1007/s11548-019-01954-w
9. Leclerc S, Smistad E, Pedrosa J, Østvik A, Cervenansky F, Espinosa F, Espeland T, Berg EAR, Jodoin P-M, Grenier T (2019) Deep learning for segmentation using an open large-scale dataset in 2D echocardiography. IEEE Trans Med Imaging 38(9):2198–2210
10. Smistad E, Østvik A, Salte IM, Melichova D, Nguyen TM, Haugaa K, Brunvand H, Edvardsen T, Leclerc S, Bernard O (2020) Real-time automatic ejection fraction and foreshortening detection using deep learning. IEEE Trans Ultrason Ferroelectr Freq Control 67(12):2595–2604. https://doi.org/10.1109/tuffc.2020.2981037
11. Kusunose K, Abe T, Haga A, Fukuda D, Yamada H, Harada M, Sata M (2020) A deep learning approach for assessment of regional wall motion abnormality from echocardiographic images. Cardiovasc Imaging 13(2):374–381
12. Sotaquira M, Pepi M, Fusini L, Maffessanti F, Lang RM, Caiani EG (2015) Semi-automated segmentation and quantification of mitral annulus and leaflets from transesophageal 3-D echocardiographic images. Ultrasound Med Biol 41(1):251–267
13. Pedrosa J, Queirós S, Vilaça J, Badano L, D'hooge J. Fully automatic assessment of mitral valve morphology from 3D transthoracic echocardiography. In: 2018 IEEE International ultrasonics symposium (IUS). IEEE; 2018. p. 1–6.
14. Andreassen BS, Veronesi F, Gerard O, Solberg AHS, Samset E (2019) Mitral annulus segmentation using deep learning in 3-D transesophageal echocardiography. IEEE J Biomed Health Inform 24(4):994–1003
15. Xie S, Girshick R, Dollár P, Tu Z, He K. Aggregated residual transformations for deep neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 1492–1500.
16. Tan M, Le Q. Efficientnet: rethinking model scaling for convolutional neural networks. In: International conference on machine learning. PMLR; 2019. p. 6105–6114.
17. Darvishi S, Behnam H, Pouladian M, Samiei N (2013) Measuring left ventricular volumes in two-dimensional echocardiography image sequence using level-set method for automatic detection of end-diastole and end-systole frames. Res Cardiovasc Med 2(1):39
18. Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, Liang J (2016) Convolutional neural networks for medical image analysis: Full training or fine tuning? IEEE Trans Med Imaging 35(5):1299–1312
19. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. Imagenet: a large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. IEEE; 2009. p. 248–255.
20. Zoph B, Vasudevan V, Shlens J, Le QV. Learning transferable architectures for scalable image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2018. p. 8697–8710. https://doi.org/10.1109/CVPR.2018.00907.
21. Chollet, F. Xception: deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 1251–1258.
22. Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, inception-resnet and the impact of residual connections on learning. In: Proceedings of the AAAI conference on artificial intelligence, vol. 1; 2017.
23. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision; 2017. p. 618–626. https://doi.org/10.1109/ICCV.2017.74.