ORIGINAL RESEARCH

Revealing Unforeseen Diagnostic Image Features With Deep Learning by Detecting Cardiovascular Diseases From Apical 4-Chamber Ultrasounds

Li-Hsin Cheng, MSc*; Pablo B. J. Bosch, MSc*; Rutger F. H. Hofman, PhD; Timo B. Brakenhoff, PhD; Eline F. Bruggemans, MSc; Rob J. van der Geest, PhD; Eduard R. Holman, MD, PhD

BACKGROUND: With the increase of highly portable, wireless, and low-cost ultrasound devices and automatic ultrasound acquisition techniques, an automated interpretation method requiring only a limited set of views as input could make preliminary cardiovascular disease diagnoses more accessible. In this study, we developed a deep learning method for automated detection of impaired left ventricular (LV) function and aortic valve (AV) regurgitation from apical 4-chamber ultrasound cineloops and investigated which anatomical structures or temporal frames provided the most relevant information for the deep learning model to enable disease classification.

METHODS AND RESULTS: Apical 4-chamber ultrasounds were extracted from 3554 echocardiograms of patients with impaired LV function (n=928), AV regurgitation (n=738), or no significant abnormalities (n=1888). Two convolutional neural networks were trained separately to classify the respective disease cases against normal cases. The overall classification accuracy of the impaired LV function detection model was 86%, and that of the AV regurgitation detection model was 83%. Feature importance analyses demonstrated that the LV myocardium and mitral valve were important for detecting impaired LV function, whereas the tip of the mitral valve anterior leaflet, during opening, was considered important for detecting AV regurgitation.

CONCLUSIONS: The proposed method demonstrated the feasibility of a 3-dimensional convolutional neural network approach in detection of impaired LV function and AV regurgitation using apical 4-chamber ultrasound cineloops. The current study shows that deep learning methods can exploit large training data to detect diseases in a different way than conventionally agreed on methods, and potentially reveal unforeseen diagnostic image features.

Key Words: 3-dimensional convolutional neural network ■ aortic valve regurgitation ■ apical 4-chamber cineloop ■ echocardiography ■ explainable artificial intelligence ■ impaired left ventricular function

Echocardiography is the main diagnostic imaging modality for the assessment of cardiovascular disease (CVD). However, although it is applicable in most settings, interpretation of echocardiograms is time-consuming and subject to intraobserver and interobserver variability. In addition, the image interpretation requires experienced experts, who are not always accessible. With the prevalence of CVD increasing, a scarcity of expert cardiologists to perform high-quality assessments is expected.[1] With the increase of highly portable, wireless, and low-cost ultrasound devices and automatic ultrasound acquisition techniques,[2] the availability of an automated interpretation method requiring only a limited set of views as input could make echocardiography-based CVD diagnosis more accessible. Such a system could become beneficial in

Correspondence to: Eduard R. Holman, MD, PhD, Department of Cardiology, Leiden University Medical Center, Albinusdreef 2, PO Box 9600, 2300 RC Leiden, the Netherlands. Email: e.r.holman@lumc.nl

*L.-H. Cheng and P. B. J. Bosch contributed equally.

CLINICAL PERSPECTIVE

What Is New?

- In this study, we developed deep learning models for automated detection of impaired left ventricular function and aortic valve regurgitation using apical 4-chamber ultrasound cineloops.
- With the feature importance analysis method DeepLIFT, we further identified the image features used by the models to detect the abnormalities.

What Are the Clinical Implications?

- With portable ultrasound devices becoming widely available, the study paves the way for an automated cardiovascular disease diagnosis tool requiring only a limited set of views as input, which could make preliminary cardiovascular disease diagnoses more accessible in the future.
- The study demonstrates that deep learning methods can be used to detect diseases in a way different from the predefined conventional way, and potentially help discovering unforeseen diagnostic image features.

Nonstandard Abbreviations and Acronyms

3D	3 dimensional
A4C	apical 4 chamber
AV	aortic valve
CNN	convolutional neural network
tSNE	T-distributed stochastic neighbor embedding

geographic regions with limited access to expert cardiologists and sonographers. It could also support general practitioners in the management of patients with suspected CVD, facilitating timely diagnosis and treatment of patients.

Recent developments in artificial intelligence technology provide an opportunity to achieve this goal. In particular, deep learning can automatically learn a hierarchy of features from a huge amount of image data,[3] thus having the potential to uncover diagnostic features in the data not previously recognized. Successful deep learning algorithms have already been developed to facilitate various steps in the workflow of echocardiography interpretation.[4–6] Among the models developed, 3-dimensional (3D) convolutional neural networks (CNNs) do allow the analysis of both spatial and temporal information of the input.[7–9] Therefore, in this study, we developed models to detect cardiovascular diseases with a 3D CNN-based approach, taking cineloops as input.

As a pilot study for a general-purpose automated CVD diagnosis model using simple inputs, we adopted the apical 4-chamber (A4C) view ultrasound cineloop as the input data, as we consider A4C a general view that contains comprehensive information in a single shot. We chose 2 abnormalities for the models to learn (namely, detection of impaired left ventricular [LV] function against normal and detection of aortic valve [AV] regurgitation against normal). Impaired LV function can be seen on the A4C view, but is recommended to be determined on multiple viewpoints.[10] This task would allow us to verify the feasibility of the 3D CNN approach, distinguishing the abnormality with limited but highly relevant information. On the other hand, AV regurgitation is typically diagnosed on the basis of color Doppler images using ≥1 viewpoints.[11] This detection task would allow us to further test the limit of a 3D CNN in distinguishing an abnormality that is not obvious on the A4C view. At the same time, it allows investigating whether the model identifies unforeseen image features while detecting the abnormality with an approach different from the clinical convention. Therefore, after training the models, we performed feature importance analysis to try to inspect what are the identified image features associated with each diagnostic task (Figure 1).

The current study proposes the use of deep learning to automatically derive CVD diagnoses from echocardiography cineloops with 2 specific focuses. First, we aimed to investigate the feasibility of using 3D CNNs to detect diseases based solely on the A4C view. Second, through feature importance analysis, we aimed to investigate whether the built models can identify anatomical and motion-related image features associated with the diseases, typically not being considered in conventional image interpretation.

METHODS

The code and the trained model weights of this study are publicly available on GitHub (https://github.com/LishinC/Disease-Detection-and-Diagnostic-Image-Feature). Additional supporting data are available from the corresponding author on reasonable request.

Data Extraction

Echocardiographic data appropriate for this retrospective study were anonymously extracted from the echocardiography database of the Heart Lung Center at the Leiden University Medical Center, Leiden, the Netherlands. The study was approved with waiver of informed consent by the Ethics Committee of the institution.
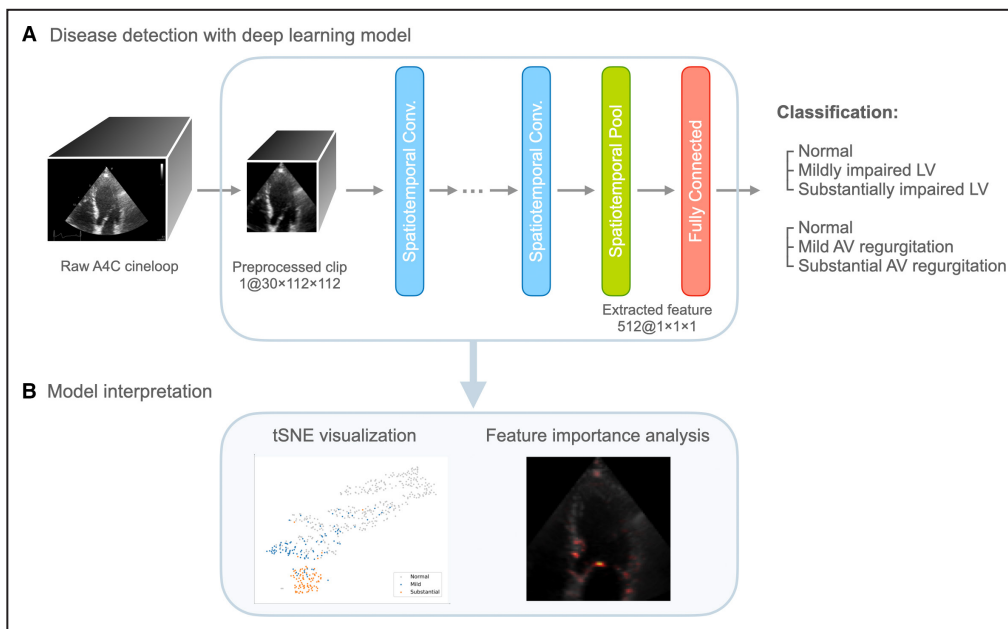
Figure 1. Study overview.
A, Two R(2+1)D models were trained to detect impaired left ventricular (LV) function and aortic valve (AV) regurgitation. **B**, Subsequently, T-distributed stochastic neighbor embedding (tSNE) was used to visualize the embedding of the extracted feature vectors, and DeepLIFT was used to identify important image features associated with the diagnostic tasks. A4C indicates apical 4 chamber; and Conv., convolutional layer.

All patients underwent echocardiography in the left lateral decubitus position, using a commercially available system (Vivid 7, E9, or E95; GE Vingmed Ultrasound AS, Horten, Norway) and 3.5-MHz transducers. Standard M-mode and 2-dimensional, color, pulsed, and continuous-wave Doppler images were acquired, according to the recommendations of the European Association of Echocardiography.[12] Offline analysis was performed using EchoPAC (version 203.59.0; GE Medical Systems). Only echocardiographic data of patients who were diagnosed as *normal* or with *impaired LV function* or *AV regurgitation* were included in this study. For patients diagnosed as *normal*, the images showed no significant abnormalities in anatomy or motion. Assessment of *impaired LV function* and *AV regurgitation* was done according to the recommendations of the American Society of Echocardiography and the European Association of Cardiovascular Imaging, based on complete acquisitions.[10,11] The LV volume was estimated by using the modified Simpson rule. The loss of LV function was qualified using the standard cutoff values for calculated LV ejection fraction[10] as mildly, moderately, or severely impaired. The severity of AV regurgitation was determined using the standard parameters (ie, vena contracta, proximal isovelocity surface area method, or pressure half time)[11] as mild, moderate, or severe regurgitation. All assessments were done by

experienced cardiologists. For our classification tasks, we combined the moderate and severe categories to form the "substantial" class. Therefore, both impaired LV function detection and AV regurgitation detection were formulated into a 3-class (normal, mild, and substantial) classification problem.

Echocardiographic examinations were anonymously extracted in DICOM format. As such, meta data, like sex, age, and weight of patients, were unknown. From the available database, a data set was created by manually selecting all A4C cineloops. The resulting 3554 ultrasounds were randomly split into separate data sets for training (70%), validation (10%), and testing (20%). Unfortunately, because of anonymization of

Table 1. Characteristics of the Data Set

Diagnosis	Total	Training	Validation	Test
Normal	1888	1322	189	377
Mildly impaired LV function	509	356	51	102
Substantially impaired LV function	419	293	42	84
Mild AV regurgitation	285	200	28	57
Substantial AV regurgitation	453	317	45	91

Per diagnosis, the total number of ultrasounds extracted (total) and ultrasounds per training, validation, and testing data set are presented. AV indicates aortic valve; and LV, left ventricular.

the data, we were not able to include the ultrasounds of individual patients in a single data set and enforce sample independence for patients. The Table 1 shows per class the total number of extracted ultrasounds and ultrasounds per data set for training, validation, and testing.

Data Preprocessing and Augmentation

In the raw ultrasound data, the frame rate for each acquisition could be different as well as the heart rate for each subject. As a result, in each raw ultrasound, a cardiac cycle could span a different number of frames. To make the temporal dimension and the contained information of the model input consistent, we wanted the input clip to be always 30 frames and corresponding to 1 cardiac cycle. To achieve this, we resampled each raw ultrasound (namely, adjusting the frame rate such that the duration of 1 cardiac cycle corresponded with 30 frames). After such resampling, taking a 30-frame clip from the frame rate-adjusted ultrasound would always lead to a clip covering exactly 1 cardiac cycle. During training time, we extracted a 30-frame clip with a random starting point as the model input. It is a part of the on-the-fly data augmentation to increase data diversity, analogous to the random shift in the spatial dimension. During validation and test time, we always extracted the clip from the starting point predetermined by the software of the ultrasound scanner, analogous to not performing any spatial augmentation during evaluation.

The raw ultrasound is embedded with electrocardiographic and text annotations. We filtered out all the embedded information so that our models would learn the 2 diagnoses based solely on the actual image information. The intensity of the filtered ultrasounds was subsequently aligned to the global intensity distribution of the whole data set through histogram matching.[13] This procedure helped ensure that the brightness and contrast of each video were roughly the same.

From the raw ultrasound of 708×1016 pixels, we cropped the center 549×549 pixels containing the fan-shaped field of view, then down-sampled the image to 112×112 pixels. During training, random translation (±5%) and rotation (±15°) were applied on the fly as augmentations to increase data diversity and prevent overfitting. The transformations were included to mimic variations that would happen in the real world because of different angles and positions of the transducer.

Model Development

We built 2 3D CNNs to separately classify impaired LV function and AV regurgitation cases against the normal cases. We decided to adopt the R(2+1)D model architecture[14] for the tasks. This particular architecture decomposes a 3D convolution into a spatial convolution followed by a temporal convolution, which was recently used to successfully predict the ejection fraction based also on A4C ultrasound cineloops.[7]

We used cross entropy as the loss function and the Adam optimizer to update network weights. The learning rate was set at 0.001, and the batch size was 16. Early stop with a patience of 50 epochs was applied. The models were implemented with the deep learning library Pytorch 1.7, and the training was performed on a NVIDIA Quadro RTX 6000 GPU. The code and the trained model weights are made available on GitHub (https://github.com/LishinC/Disease-Detection-and-Diagnostic-Image-Feature).

Feature Exploration Using T-Distributed Stochastic Neighbor Embedding Visualization

T-distributed stochastic neighbor embedding (tSNE)[15] is a dimensionality reduction method that is often used to embed high-dimensional data into a 2- (or 3-) dimensional embedding for the purpose of visual exploration. When performing embedding, the tSNE method tries to preserve the local relative distance between samples, such that closer data points in a tSNE plot imply more similar samples.

tSNE can be applied to visualize a variety of high-dimensional data. In this study, the method was used to visualize both the input video and the features extracted by the trained 3D CNN (512-dimensional vector). tSNE visualization of the input video shows the similarity of samples based directly on the pixel intensities, whereas visualization of the extracted features shows the similarity of samples based on the feature values. Comparison of the 2 can reveal whether the 3D CNN successfully extracted features relevant to the diagnosis and filtered out the noise, such that samples belonging to the same diagnosis were clustered closer together in the extracted-feature plot compared with the input-video plot. In addition, the extracted-feature plot could also indicate the relationship (the distance/similarity to each other) between several overlapping clusters.

Feature Importance Analysis

We used the feature importance analysis method DeepLIFT[16,17] to try to decrypt the reasoning behind the models' predictions, which could potentially help reveal diagnostic image features not considered before. DeepLIFT attributes a model's classification output to certain input features (pixels), which allows us to understand which region or frame in an ultrasound is the key that makes the model classify it as a certain diagnosis.

DeepLIFT decomposes the output activation difference between the input and the baseline as a sum of layer-wise relevance values, thus obtaining the

contribution of each input feature (pixel) to the output prediction. On a given input of interest, DeepLIFT would return an analysis result with the same size as the input for each class. The values in the analysis result reflect the importance of that pixel to the class, and the sign of the values indicates either a positive or a negative contribution to the class.

RESULTS

Model Performance

Figure 2 summarizes the predictive performance of the 2 models. Figure 2A and 2C shows the normalized confusion matrices for the impaired LV function and AV regurgitation detection models, respectively. Figure 2B and 2D lists the detailed recall (sensitivity), precision, and F1-score values for each diagnosis for both models.

The impaired LV function detection model achieved an overall accuracy of 86%. The model was able to detect 92% of the ultrasounds qualified by the cardiologist as substantially impaired. Of the mildly impaired class,

67% were correctly identified. On the other hand, the AV regurgitation detection model reached an overall accuracy of 83% and was able to detect 71% of the substantial class, but only 25% of the mild class.

Feature Exploration Using tSNE Visualization

Figure 3 shows the tSNE plots for impaired LV function (Figure 3A and 3B) and AV regurgitation (Figure 3C and 3D), with each sample colored by the corresponding diagnosis. By comparing the plots before being processed by the models (Figure 3A and 3C) and after (Figure 3B and 3D) (ie, the input-video plot with the extracted-feature plot), we can observe that samples of the same diagnosis are clustered closer to each other after being processed by the models. This implies that the models have successfully extracted diagnosis-relevant features and filtered out irrelevant noise throughout cascades of convolutional layers.

Especially, for the case of impaired LV function detection, it can be seen in Figure 3B that, after processing, the normal, mild, and substantial clusters are even

B

	Recall	Precision	F1-score
Normal	0.9	0.93	0.91
Mild	0.67	0.62	0.64
Substantial	0.92	0.9	0.91

D

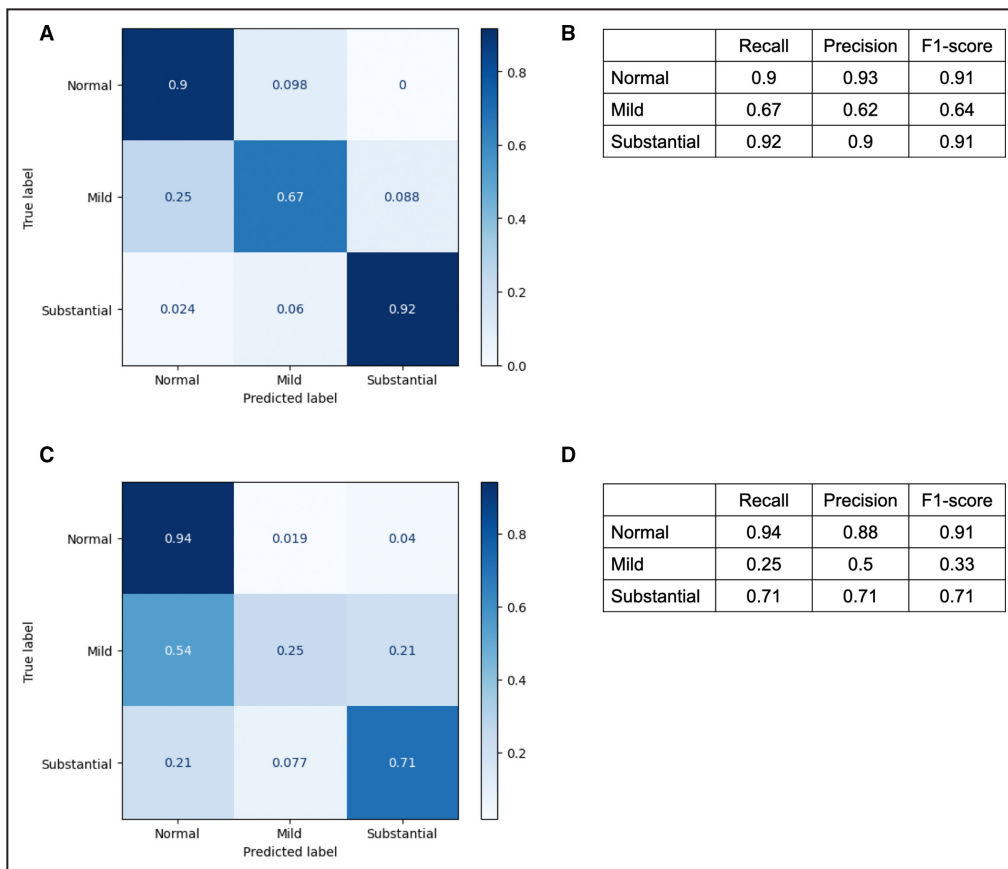|	Recall	Precision	F1-score
Normal	0.94	0.88	0.91
Mild	0.25	0.5	0.33
Substantial	0.71	0.71	0.71

Figure 2. Predictive performance of the impaired left ventricular function detection model (A and B) and the aortic valve regurgitation detection model (C and D).
A and C, The normalized confusion matrices for each classification task. B and D, The detailed recall (sensitivity), precision, and F1-score of each class. The performance demonstrates the feasibility of detecting the diseases using apical 4-chamber cineloops.
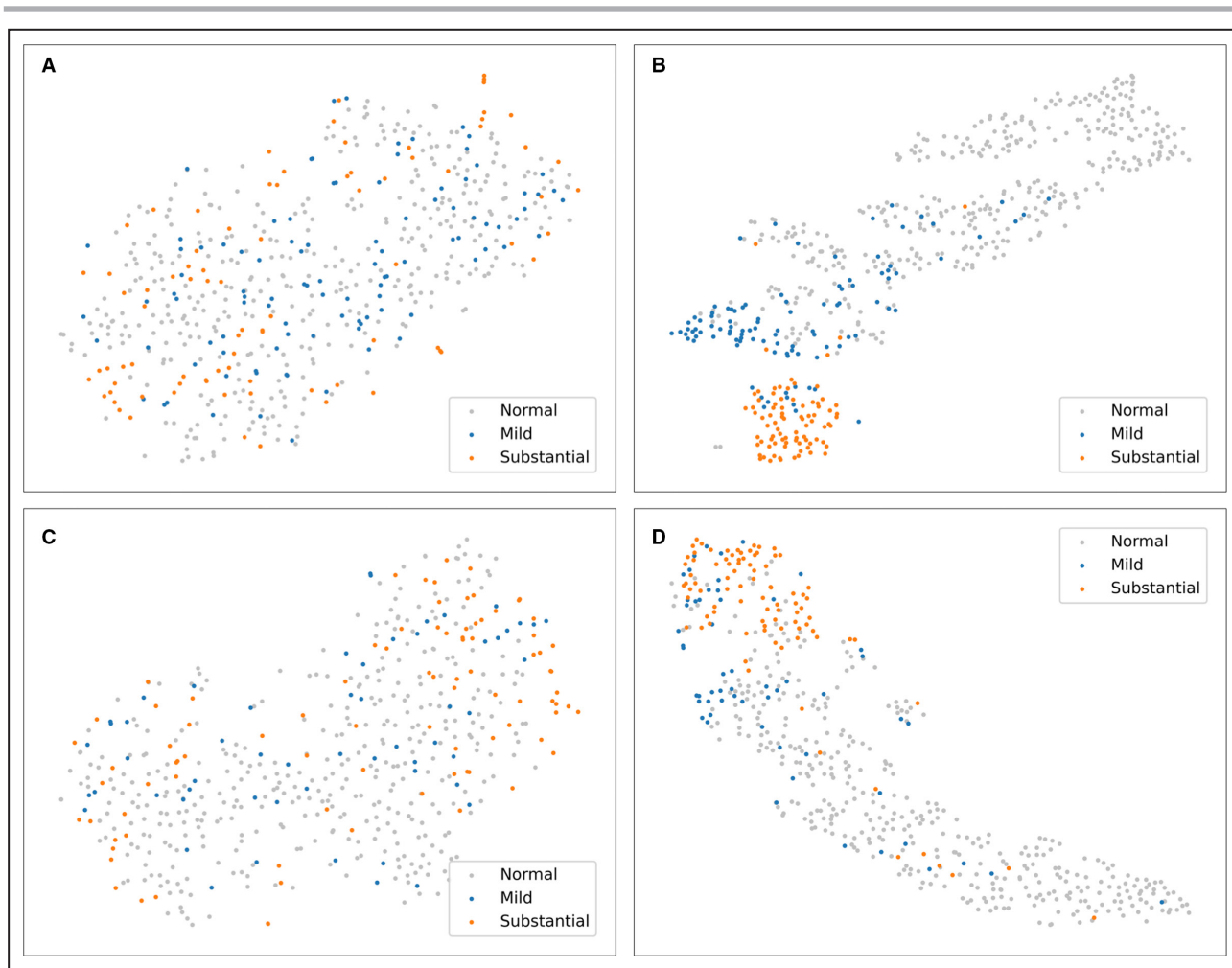
Figure 3.   T-distributed stochastic neighbor embedding (tSNE) visualization for impaired left ventricular function (A and B) and aortic valve regurgitation (C and D).

A and **C**, The visualizations of input ultrasound, in which the input video was directly reduced into a 2-dimensional embedding by tSNE and visualized. **B** and **D**, The visualizations of the model-extracted feature vector, in which the 512-dimensional feature extracted by the model was reduced into a 2-dimensional embedding by tSNE and visualized. Nearby dots in the plots imply similar data in the original high-dimensional space. It can be observed (**B** and **D**) that samples of the same diagnosis are clustered closer to each other, as compared with other data (**A** and **C**). This indicates that the models have filtered out irrelevant information and extracted diagnosis-relevant features. **B**, Especially, it can be seen that the model might have obtained information on the level of severity.

ordered in the level of severity. The information about the severity (ie, the correct ordering of the 3 classes) was actually not given to the model. The 1-hot categorical label fed to the model implies only that normal, mild, and substantial were 3 different classes, which does not contain hints about the relative similarity between each class. Therefore, besides being able to differentiate the 3 classes, the model had also learned the correct relative similarity relationship between the 3 classes.

Feature Importance Analysis

Figure 4 presents the feature importance analysis results produced by DeepLIFT, which attributes (per query) the model's prediction of an output class to certain input features (pixels). For both the impaired LV function and AV regurgitation detection models, we present the analysis calculated on the basis of the same representative normal case. The analysis results are presented as heat maps in Figure 4. The brighter pixels in the heat maps are the input features that positively contributed to the normal class (ie, image features that made the model classify the case as normal). The highlighted regions can thus be interpreted as the image information that makes the models distinguish the normal case from the disease cases. More analyses in video format can be found in our GitHub repository.

For the impaired LV function detection model, DeepLIFT highlighted the basal part of the myocardium from the early systolic to the early diastolic phase. In addition,
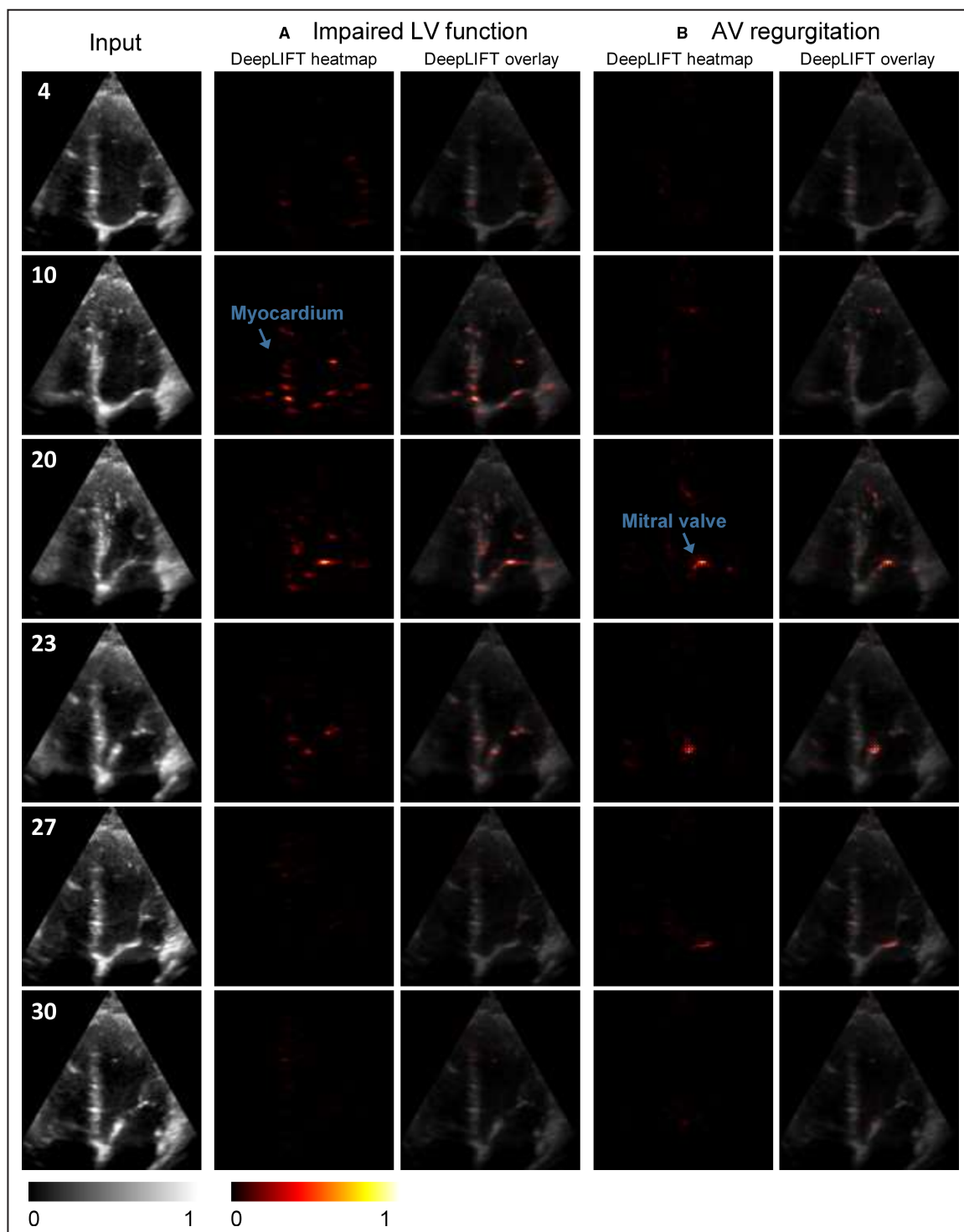
Figure 4. Feature importance analysis with DeepLIFT.
The highlighted regions are image features considered important by the respective models to distinguish normal cases from the disease cases. **A**, The analysis of the impaired left ventricular (LV) function detection model. The LV myocardium at the basal level was highlighted from the early systolic to the early diastolic phase. The mitral valve was highlighted as well during the early diastolic phase. **B**, The analysis of the aortic valve (AV) regurgitation detection model. The tip of the mitral valve anterior leaflet was highlighted particularly at a short time, centering around the moment of valve opening. This indicates that the model focuses not only on a certain anatomical structure but also on a certain temporal phase within the cardiac cycle. This representative case shown herein is the one with the highest average probability of being the normal class, as predicted by the 2 models (namely, a confident case). More examples in video format are available in our GitHub repository.

the mitral valve was highlighted at the early diastolic phase (Figure 4A).

For the AV regurgitation detection model, DeepLIFT highlighted the tip of the mitral valve anterior leaflet, particularly at a short time centering around the moment of valve opening (Figure 4B). This indicates that the model not only focused on a specific anatomical structure but also on a specific temporal phase within the cardiac cycle.

DISCUSSION

As a pilot study to make preliminary diagnoses of cardiovascular diseases automated and thus more accessible, we built 2 3D CNNs to detect impaired LV function and AV regurgitation using the A4C-view ultrasound. The impaired LV function model was able to detect 92% of the substantial class, and the AV regurgitation model was able to detect 71% of the substantial class. On the basis of the lower recall, we conclude that detecting AV regurgitation was the more difficult task among the 2. This is in line with the fact that AV regurgitation is usually diagnosed using Doppler imaging and not from an A4C view. However, our results also reveal that abnormalities derived from the A4C view, although not obvious to the human eye, were sufficient for the AV regurgitation model to reach an overall detection accuracy of 83%. The success of building the impaired LV function detection model demonstrates the feasibility of deep learning algorithms in identifying the abnormality with limited input information. Furthermore, the success of AV regurgitation detection verifies that the model can detect a disease with an approach different from the current practice (ie, based on Doppler information). We attribute this to the models' ability to learn from huge amounts of data and derive important features independently.

Using tSNE visualization, we verified that the models had transformed the input ultrasounds into a diagnosis-relevant feature representation. Especially, similar to the reconstruction of disease progression, as shown in the study by Eulenberg et al,[18] the impaired LV function detection model had mapped the normal, mild, and substantial classes in the correct order. This indicates that the model might have obtained information about the severity of the disease, although to which extent it can accurately rank the severity of each individual case requires further evaluation. Nevertheless, this indicates that the trained model could potentially serve as a tool to systematically evaluate the severity of the disease, which would otherwise be hard to accurately quantify by the human eye, merely from a single view and without additional annotation.

Finally, to see which signs in the input cineloop the models focus on to detect the abnormalities, we analyzed the models with the feature importance analysis method DeepLIFT. The analysis suggests that the mitral valve and the LV myocardium at the basal level are crucial for distinguishing the normal class from impaired LV function. This observation verifies that the model works in a reasonable way to detect the disease, because the movement of the myocardium is strongly related to LV function, as is the movement of the mitral valve.[19] On the other hand, the analysis suggests that the tip of the mitral valve anterior leaflet, during the opening of the valve, is the most important feature that the model focuses on to distinguish the normal class from AV regurgitation. It is possible that the movement of the mitral valve is affected by the abnormal regurgitation jet[20] and, hence, identified by the model as a key difference. It is also possible that morphological changes, such as mitral valve leaflet enlargement, were the key characteristic that the model used to distinguish cases, as supported by a recent study.[21] Although the exact mechanism remains unclear, the analysis shows that certain regions of the heart or phases in the cardiac cycle that people often neglected previously might also have a strong link to the disease. Potentially, a trained model could identify image features that are not yet known to be related to a disease, hence bringing insights to the disease diagnosis.

For the respective models, the DeepLIFT-highlighted regions in the shown normal case represent the general highlighting pattern that we observed in most of the normal cases. These consensus highlighted regions are the diagnostic image features concluded by the respective models after learning from the training data. However, if we input disease cases to the analysis, the highlighted regions in different queries would be different. (Examples of the DeepLIFT analysis on disease cases can be found in our GitHub repository.) We speculate that this might be attributable to a higher heterogeneity in the appearance of the disease cases. This links to a major bottleneck in the current feature importance analysis workflow, where the method can only show important input features per query, instead of directly deriving high-level information from the trained model weights. For instance, the analysis cannot tell us directly from the trained model weights that the wall motion abnormality at the end systolic phase is the most critical sign for a disease. Human interpretation is still required to obtain insights from the trained model by going through the highlighted regions in multiple input queries. Another shortcoming of the currently available feature importance analysis methods, such as DeepLIFT, is that the analysis result is often noisy.[22] This is especially true in our case, where the input ultrasound was already noisy. The noise would often hinder the further interpretation process.

Apart from these shortcomings of DeepLIFT, this study has several limitations. A first limitation refers to the lack of patient-relevant information. Because of

data protection regulations, all ultrasound data used in this study were anonymized and stripped of identifying meta data. Therefore, we were not able to maintain subject-level independence for the training-validation-testing splits. Also, further analyses on age, sex, and relevant clinical information and echocardiographic findings were not possible. If these data would become available in the future, the current analyses could be extended to investigate the possibility of integration of nonimaging data in the model. Second, there are currently no other public data sets with diagnostic labels available as in our data set. If, in the future, an independent validation data set becomes available, we would be able to further verify the generalization ability of our models. We would also like to train our models on an extended data set with higher sample variety (multiple medical centers and various ultrasound machines), such that the generalizability of the models on unseen independent validation data can be potentially improved. At the moment, we provide our code and trained model weights on GitHub for everyone to use to externally validate or fine-tune in a transfer learning manner on private data sets. Third, the misclassification rates of the models, particularly for the mild classes, were still high, especially in AV regurgitation detection. Failing to identify the mild cases might have important clinical implications, such as delayed diagnosis and treatment. Therefore, further refinement of the models to decrease the misclassification rates is needed before their deployment in clinical routine. Finally, this pilot study served mainly as a proof of concept of using a simple input for automated CVD diagnosis. Whether a best single view or best view combination exists for detecting each distinct CVD, and whether A4C is an acceptable input already for detecting a variety of CVDs, remain topics for further investigation.

CONCLUSIONS

In conclusion, this pilot study shows the feasibility of a 3D CNN approach in the detection of impaired LV function and AV regurgitation based on A4C-view ultrasound cineloops, which paves the way for an automated CVD diagnosis that can be made more accessible. Moreover, it demonstrated that deep learning methods can learn from large training data to detect diseases in a way different from the predefined conventional way, and potentially discover diagnostic image features not previously paid attention to by humans.

ARTICLE INFORMATION

Affiliations

Division of Image Processing, Department of Radiology, Leiden University Medical Center, Leiden, the Netherlands (L.C., R.J.v.d.G.); Department of Science, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands (P.B.B., R.F.H.); Ynformed, Utrecht, the Netherlands (P.B.B., T.B.B.); Department of Cardiothoracic Surgery (E.F.B.) and Department of Cardiology, Leiden University Medical Center, Leiden, the Netherlands (E.R.H.).

Acknowledgments

Sources of Funding

Disclosures

None.

REFERENCES

1. Dall TM, Gallo PD, Chakrabarti R, West T, Semilla AP, Storm MV. The care span: an aging population and growing disease burden will require a large and specialized health care workforce by 2025. *Health Aff*. 2013;32:2013–2020. doi: 10.1377/hlthaff.2013.0714

2. Wang S, Singh D, Johnson D, Althoefer K, Rhode K, Housden RJ. Robotic ultrasound: view planning, tracking, and automatic acquisition of transesophageal echocardiography. *IEEE Robot Autom Mag*. 2016;23:118–127. doi: 10.1109/MRA.2016.2580478

3. Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. *Eur Conf Comput Vis*. 2014;8689 LNCS:818–833.

4. Schuuring MJ, Išgum I, Cosyns B, Chamuleau SAJ, Bouma BJ. Routine echocardiography and artificial intelligence solutions. *Front Cardiovasc Med*. 2021;8:1–10. doi: 10.3389/fcvm.2021.648877

5. Kusunose K, Haga A, Abe T, Sata M. Utilization of artificial intelligence in echocardiography. *Circ J*. 2019;83:1623–1629. doi: 10.1253/circj. CJ-19-0420

6. Gandhi S, Mosleh W, Shen J, Chow CM. Automation, machine learning, and artificial intelligence in echocardiography: a brave new world. *Echocardiography*. 2018;35:1402–1418. doi: 10.1111/echo.14086

7. Ouyang D, He B, Ghorbani A, Yuan N, Ebinger J, Langlotz CP, Heidenreich PA, Harrington RA, Liang DH, Ashley EA, et al. Video-based AI for beat-to-beat assessment of cardiac function. *Nature*. 2020;580:252–256. doi: 10.1038/s41586-020-2145-8

8. Kusunose K, Haga A, Yamaguchi N, Abe T, Fukuda D, Yamada H, Harada M, Sata M. Deep learning for assessment of left ventricular ejection fraction from echocardiographic images. *J Am Soc Echocardiogr*. 2020;33:632–635.e1. doi: 10.1016/j.echo.2020.01.009

9. Silva JF, Silva JM, Guerra A, Matos S, Costa C. Ejection fraction classification in transthoracic echocardiography using a deep learning approach. In: Hollmén J, McGregor C, Soda P, Kane B, eds. *2018 IEEE 31st Symposium on Computer-Based Medical Systems*. IEEE; 2018:123–128.

10. Lang RM, Badano LP, Mor-Avi V, Afilalo J, Armstrong A, Ernande L, Flachskampf FA, Foster E, Goldstein SA, 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. doi: 10.1093/ehjci/jev014

11. Lancellotti P, Tribouilloy C, Hagendorff A, Popescu BA, Edvardsen T, Pierard LA, Badano L, Zamorano JL, Bruder O, Cosyns B, et al. Recommendations for the echocardiographic assessment of native valvular regurgitation: an executive summary from the European Association of Cardiovascular Imaging. *Eur Heart J Cardiovasc Imaging*. 2013;14:611–644. doi: 10.1093/ehjci/jet105

12. Evangelista A, Flachskampf F, Lancellotti P, Badano L, Aguilar R, Monaghan M, Zamorano J, Nihoyannopoulos P. European Association of Echocardiography recommendations for standardization of performance, digital storage and reporting of echocardiographic studies. *Eur J Echocardiogr*. 2008;9:438–448. doi: 10.1093/ejechocard/jen174

13. Gonzalez RC, Woods RE. *Digital Image Processing, Global Edition*. Pearson Education Limited; 2017:1-140–144.

14. Tran D, Wang H, Torresani L, Ray J, LeCun Y, Paluri M. A closer look at spatiotemporal convolutions for action recognition. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE; 2018:6450–6459.

15. Van Der Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res*. 2008;9:2579–2625.

16. Shrikumar A, Greenside P, Kundaje A. Learning important features through propagating activation differences. *Proceedings of the 34th International Conference on Machine Learning*. PMLR; 2017:3145–3153.

17. Kokhlikyan N, Miglani V, Martin M, Wang E, Alsallakh B, Reynolds J, Melnikov A, Kliushkina N, Araya C, Yan S, et al. Captum: a unified and generic model interpretability library for PyTorch. *arXiv*. 2020. Available at: http://arxiv.org/abs/2009.07896. Accessed January 25, 2021.

18. Eulenberg P, Köhler N, Blasi T, Filby A, Carpenter AE, Rees P, Theis FJ, Wolf FA. Reconstructing cell cycle and disease progression using deep learning. *Nat Commun*. 2017;8:1–6. doi: 10.1038/s41467-017-00623-3

19. Jugdutt BI, Lee SJK, McFarlane D. Noninvasive assessment of left ventricular function from the mitral valve echogram. Relation of final anterior mitral leaflet closing velocity to peak dp/dt and aortic velocity. *Circulation*. 1978;58:861–871. doi: 10.1161/01.CIR.58.5.861

20. Akinseye OA, Pathak A, Ibebuogu UN. Aortic valve regurgitation: a comprehensive review. *Curr Probl Cardiol*. 2018;43:315–334. doi: 10.1016/j.cpcardiol.2017.10.004

21. Hirasawa K, Izumo M, Sasaoka T, Ashikaga T, Suzuki K, Harada T, Isobe M, Akashi YJ. Effect of aortic regurgitant jet direction on mitral valve leaflet remodeling: a real-time three-dimensional transesophageal echocardiography study. *Sci Rep*. 2017;7:1–10. doi: 10.1038/s41598-017-09252-8

22. Singh A, Sengupta S, Lakshminarayanan V. Explainable deep learning models in medical image analysis. *J Imaging*. 2020;6:52. doi: 10.3390/jimaging6060052