BRIEF RESEARCH COMMUNICATION

Deep Learning for Assessment of Left Ventricular Ejection Fraction from Echocardiographic Images

For automated measurements of left ventricular ejection fraction (LVEF), obtaining accurate border detection is a difficult task due to the complicated temporal deformation of the left ventricle (LV). Recently, deep learning (DL) has been developed as a state-of-the-art method for the classification of cardiovascular diseases.[1,2] Our study aim was to evaluate whether a three-dimensional convolutional neural network (3DCNN) could estimate and differentiate preserved ejection fraction (EF) or reduced EF independently of volumes using echocardiographic images.

The 3DCNN model was trained on a selected data set of 340 heart failure (HF) patients with homogeneously distributed EF range (185 patients had LVEF < 50%, and 155 patients had LVEF ≥ 50%). We selected cases with good or adequate acoustic detail to test the DL algorithm on images obtained from two vendors' machines. To test for generalizability, we gathered a separate validation group of 189 consecutive patients who were referred to our laboratory using six vendors with various image qualities (68 patients had LVEF < 50%, and 121 patients had LVEF ≥ 50%). The Institutional Review Board approved the study protocol.

Echocardiographic measurements were obtained according to guidelines.[3] Apical two-chamber (AP2), apical four-chamber (AP4), apical three-chamber (AP3), parasternal long-axis (PLAX), and parasternal short-axis (PSAX) views were stored digitally. The import data process is shown in Supplemental Figure 1. The process for the DL model is shown in Figure 1. The preprocessing is described in the Supplemental Material. Estimation of LVEF was accomplished by a 3DCNN algorithm. Using the Python version 3.6 programming language, described in detail in the Supplemental Material, the 3DCNN model automatically captures features that are relevant for LVEF from image data and was applied in the regression of LVEF quantity. To serve as reference, LV end-diastolic volume, LV end-systolic volume, and LVEF were calculated by the biplane method of disks using the AP2 and AP4 views, and this measurement was confirmed by the other echocardiographic views (AP3, PLAX, and PSAX) by two experts (K.K. and H.Y.). We used reference LVEF values obtained by averaging measurements by the two experts.

The results of the comparison between LVEF by a 3DCNN model and reference LVEF are shown in Figure 2. We built two models based on AP2/AP4 or the five views (AP2, AP4, AP3, PLAX, and PSAX). A good correlation was found between estimated LVEF based on the AP2 and AP4 views and reference LVEF ($r = 0.88$, $P < .001$). An excellent correlation was found between estimated LVEF based on an average of all five views (AP2, AP4, AP3, PLAX, and PSAX) and reference LVEF ($r = 0.92$, $P < .001$).

Receiver operating characteristic analysis was used to assess the diagnostic ability for classification of HF subtypes (heart failure reduced ejection fraction: LVEF < 50%; and heart failure preserved ejection fraction: LVEF ≥ 50%). The area under the curve (AUC) for LVEF by the averaged five views (AUC, 0.99 ± 0.01 on five-fold cross validation) was larger than the AUC for LVEF based on the AP2/AP4 views (AUC, 0.95 ± 0.01 on five-fold cross validation, compared $P < .05$).

For the correlation between LVEF by a 3DCNN model and reference LVEF in the separate validation group, a good correlation was found between estimated LVEF based on an average of all five views ($r = 0.82 \pm 0.02$, $P < .001$). The AUC was 0.92 ± 0.01. The root-mean squared error in this independent cohort was almost comparable with that in test samples of the original cohort (root-mean squared error = 7.10 ± 0.02 and 7.12 ± 0.01, respectively). According to the result from the independent cohort, we believe that the algorithm would also perform well in subjects regardless of HF.

To help understand the 3DCNN assessment, we have selected the top 20 cases of over- or underestimation of LVEF by the 3DCNN algorithm (10 cases for overestimation and 10 cases for underestimation) in the test cohort of 68 patients. There was no statistically significant difference for patient characteristics between the correctly estimated group and the over- or underestimated group. However, heart rate trended higher and LV size trended larger in the over- or underestimated group than in the correctly estimated group. One reason is that the sample size with large LV size and high heart rate is relatively small. Thus, in the future it is important to include large LVs and cases with high heart rate in the development of DL models.

Theoretically, estimated LVEF based on AP2/AP4 is an appropriate method because we used the reference LVEF based on the biplane method of disks. A recent paper showed that the LVEF based on AP2/AP4 views by machine learning algorithms had a good correlation with reference LVEF.[4] Our analysis adds to this by demonstrating an even better performance of a DL algorithm when five views are utilized. It is well known that the prediction accuracy of DL can be improved by averaging the several models. According to our results, it may be more accurate to make a prediction model for LVEF from multilevel images in the clinical setting.

In conclusion, the 3DCNN algorithm using multiviews for volume-independent LVEF estimation has the potential to estimate and classify LVEF in the clinical setting.

Kenya Kusunose, MD, PhD
Department of Cardiovascular Medicine
Tokushima University Hospital
Tokushima, Japan

Akihiro Haga, PhD
Department of Medical Image Informatics
Graduate School of Biomedical Sciences
Tokushima University, Tokushima, Japan

Natsumi Yamaguchi, MS
Department of Cardiovascular Medicine
Tokushima University Hospital
Tokushima, Japan

Takashi Abe, MD, PhD
Department of Radiology and Radiation Oncology
Graduate School of Biomedical Sciences
Tokushima University, Tokushima, Japan

Daiju Fukuda, MD, PhD
Department of Cardiovascular Medicine
Tokushima University Hospital
Tokushima, Japan

Hirotsugu Yamada, MD, PhD
Department of Community Medicine for Cardiology
Graduate School of Biomedical Sciences

This work was partially supported by Japan Society for the Promotion of Science (JSPS) Kakenhi grants (no. 17K09506 to K.K. and 19H03654 to M.S.) and the Takeda Science Foundation (to K.K.).

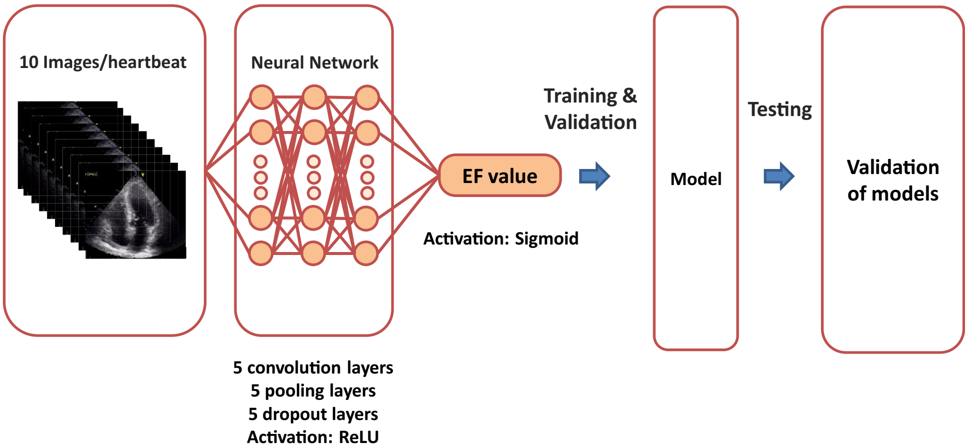


Figure 1 Neural networks for the estimation of LVEF. The fully connected layers transform the image features into the final scores by adjusting weights for neuron activations during training. *ReLU*, Rectified linear unit.

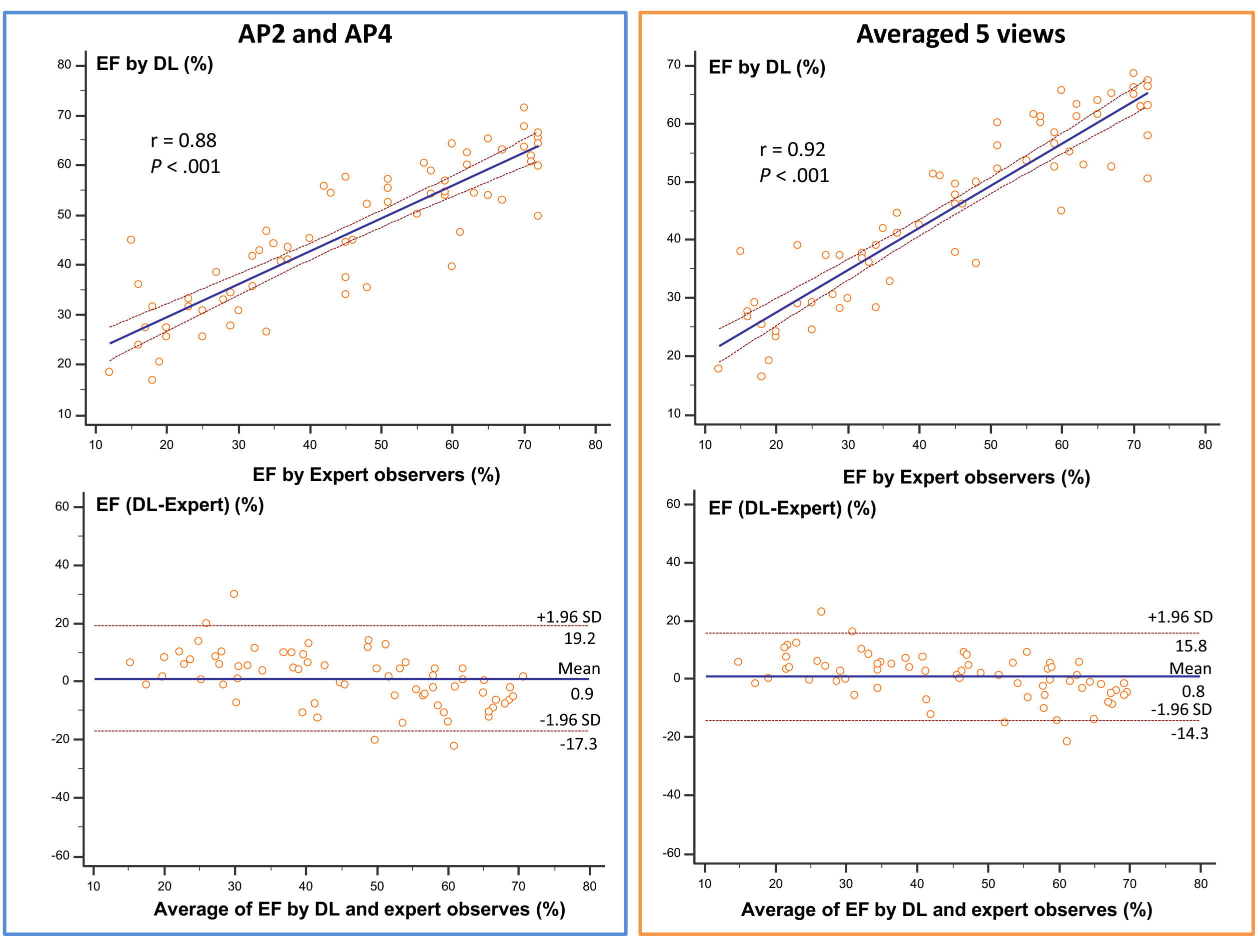


Figure 2 Correlation between estimated LVEF and biplane LVEF. A good correlation was found between estimated EF based on the AP2 and AP4 views and reference EF (r = 0.88, P < .001). An excellent correlation was found between estimated EF based on an average of all five views by DL and reference EF (r = 0.92, P < .001).

Tokushima University
Tokushima, Japan

Masafumi Harada, MD, PhD
Department of Radiology and Radiation Oncology
Graduate School of Biomedical Sciences
Tokushima University
Tokushima, Japan

Masataka Sata, MD, PhD
Department of Cardiovascular Medicine
Tokushima University Hospital
Tokushima, Japan

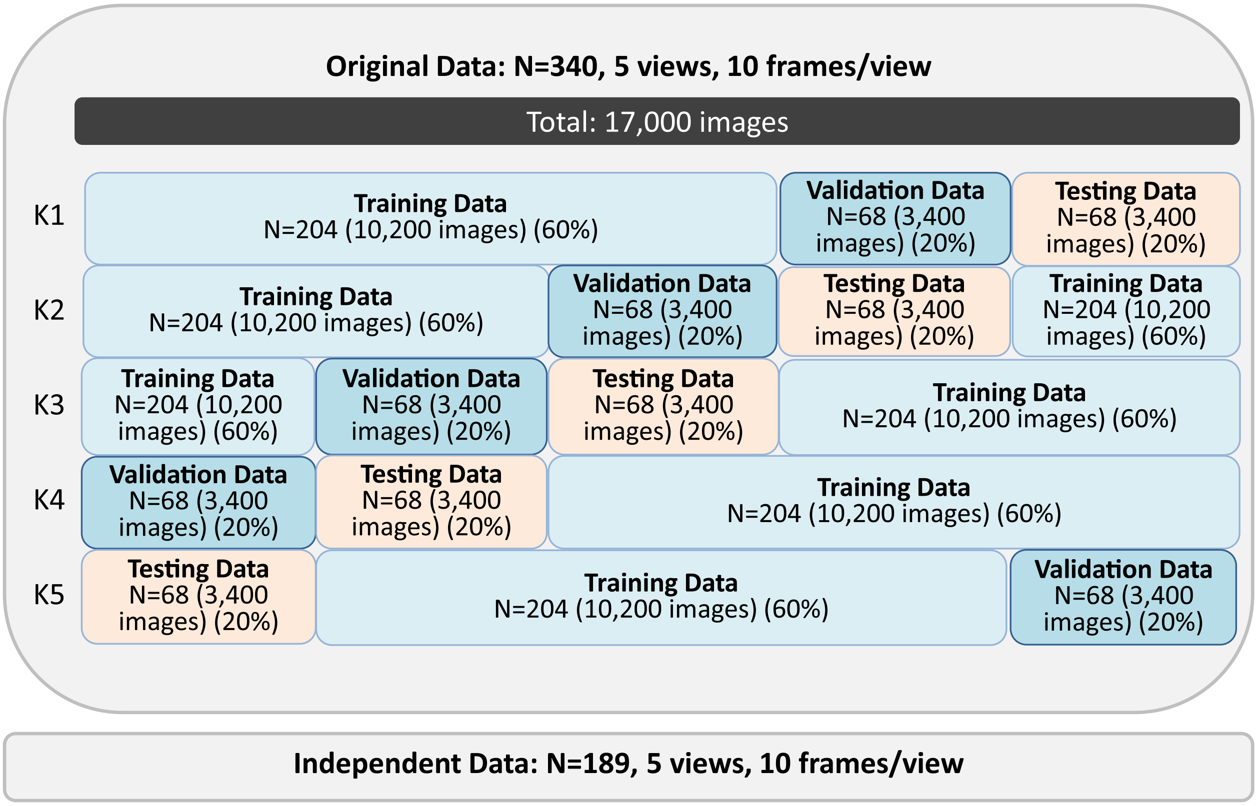
SUPPLEMENTARY DATA

Supplementary data related to this article can be found at https://doi.org/10.1016/j.echo.2020.01.009.

REFERENCES

1. Kusunose K, Haga A, Abe T, Sata M. Utilization of artificial intelligence in echocardiography. Circ J 2019;83:1623-9.
2. Kusunose K, Abe T, Haga A, Fukuda D, Yamada H, Harada M, et al. A deep learning approach for assessment of regional wall motion abnormality from echocardiographic images. JACC Cardiovasc Imaging 2019. https://doi.org/10.1016/j.jcmg.2019.02.024.
3. Mitchell C, Rahko PS, Blauwet LA, Canaday B, Finstuen JA, Foster MC, et al. Guidelines for performing a comprehensive transthoracic echocardiographic examination in adults: recommendations from the American Society of Echocardiography. J Am Soc Echocardiogr 2019;32:1-64.
4. Asch FM, Poilvert N, Abraham T, Jankowski M, Cleve J, Adams M, et al. Automated echocardiographic quantification of left ventricular ejection fraction without volume measurements using a machine learning algorithm mimicking a human expert. Circ Cardiovasc Imaging 2019;12:e009303.

https://doi.org/10.1016/j.echo.2020.01.009



Supplemental Figure 1 Import data. All data were divided into five groups, and then four of the groups were used as a training and validation to create a model; the rest were used in a test of the model. A total of 340 cases with 17,000 images were split with 204 cases (10,200 images) as the training set, 68 cases (3,400 images) as the validation set, and 68 cases (3,400 images) as the test set. Then five-fold cross validation (K1-K5) was employed to show a model performance.