



## Article

# Atrial Fibrillation Type Classification by a Convolutional Neural Network Using Contrast-Enhanced Computed Tomography Images

Hina Kotani <sup>1</sup>, Atsushi Teramoto <sup>2,\*</sup>, Tomoyuki Ohno <sup>3</sup>, Yoshihiro Sobue <sup>4</sup>, Eiichi Watanabe <sup>4</sup> and Hiroshi Fujita <sup>5</sup>

<sup>1</sup> Graduate School of Health Sciences, Fujita Health University, Toyoake 470-1192, Japan; hina.0306rv@gmail.com

<sup>2</sup> Faculty of Information Engineering, Meijo University, Nagoya 468-8502, Japan

<sup>3</sup> Department of Radiation, Fujita Health University Bantane Hospital, Nagoya 454-8509, Japan

<sup>4</sup> Department of Internal Medicine, Fujita Health University Bantane Hospital, Nagoya 454-8509, Japan

<sup>5</sup> Faculty of Engineering, Gifu University, Gifu 501-1193, Japan

\* Correspondence: teramoto@meijo-u.ac.jp

**Abstract:** Catheter ablation therapy, which is a treatment for atrial fibrillation (AF), has a higher recurrence rate as AF duration increases. Compared to paroxysmal AF (PAF), sustained AF is known to cause progressive anatomic remodeling of the left atrium, resulting in enlargement and shape changes. In this study, we used contrast-enhanced computed tomography (CT) to classify atrial fibrillation (AF) into paroxysmal atrial fibrillation (PAF) and long-term persistent atrial fibrillation (LSAF), which have particularly different recurrence rates after catheter ablation. Contrast-enhanced CT images of 30 patients with PAF and 30 patients with LSAF were input into six pretrained convolutional neural networks (CNNs) for the binary classification of PAF and LSAF. In this study, we propose a method that can recognize information regarding the body axis direction of the left atrium by inputting five slices near the left atrium. The classification was visualized by obtaining a saliency map based on score-class activation mapping (CAM). Furthermore, we surveyed cardiologists regarding the classification of AF types, and the results of the CNN classification were compared with the results of physicians' clinical judgment. The proposed method achieved the highest correct classification rate (81.7%). In particular, models with shallow layers, such as VGGNet and ResNet, are able to capture the overall characteristics of the image and therefore are likely to be suitable for focusing on the left atrium. In many cases, patients with an enlarged left atrium tended to have long-lasting AF, confirming the validity of the proposed method. The results of the saliency map and survey of physicians' basis for judgment showed that many patients tended to focus on the shape of the left atrium in both classifications, suggesting that this method can classify atrial fibrillation more accurately than physicians, similar to the judgment criteria of physicians.

**Keywords:** atrial fibrillation; catheter ablation; classification; convolutional neural network; contrast-enhanced computed tomography; deep learning



**Citation:** Kotani, H.; Teramoto, A.; Ohno, T.; Sobue, Y.; Watanabe, E.; Fujita, H. Atrial Fibrillation Type Classification by a Convolutional Neural Network Using Contrast-Enhanced Computed Tomography Images. *Computers* **2024**, *13*, 309. <https://doi.org/10.3390/computers13120309>

Academic Editors: Selene Tomassini and M. Ali Akber Dewan

Received: 23 October 2024

Revised: 19 November 2024

Accepted: 21 November 2024

Published: 24 November 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The number of patients with atrial fibrillation (AF) is increasing annually, and this trend is naturally related to the aging of the population [1]. In recent years, the aging of patients with AF has brought to light clinical problems that were previously invisible. The European Society of Cardiology (ESC) notes that six main problems are closely associated with AF: mortality, stroke, hospitalization, reduced quality of life, left ventricular dysfunction/heart failure, and cognitive decline/vascular dementia [2]. Therefore, the early detection and treatment of AF are important to prevent complications. AF is a disease that gradually shortens the interval between attacks over time, eventually becoming

persistent, long-lasting, and permanent. Thus, atrial fibrillation can be viewed as a disease that progresses through various stages [3]. Catheter ablation therapy, which is a treatment for AF, has been shown to be effective for paroxysmal atrial fibrillation (PAF). However, its efficacy is not well established in non-pharmacological guidelines for persistent atrial fibrillation and long-standing persistent atrial fibrillation (LSAF), for which the recommended level is Class IIa or Class IIb [4]. In other words, it is very important to determine which patients with persistent atrial fibrillation will benefit from catheter ablation therapy based on the results and possible complications of catheter ablation therapy for persistent AF, as described above [5]. However, it is difficult to predict postoperative recurrence, and the indications for catheter ablation therapy are currently determined based on the surgeon's empirical judgment and the patient's self-reported AF duration.

AF recurrence after catheter ablation therapy and its predictors have been the subject of many studies [6–9]. Njoku et al. showed that left atrial diameter predicts AF recurrence after radiofrequency catheter ablation treatment in a meta-analysis of the difference in left atrial volume between patients with and without recurrent AF after radiofrequency catheter ablation [6]. Other factors, such as the duration of AF, structural changes in the left atrium and pulmonary veins, and age, may also affect the outcome of catheter ablation therapy.

In recent years, many methods have been reported to classify AF types [10,11], and Nuria Ortigosa et al. proposed a method to classify AF subtypes with feature extraction from a general Fourier time-frequency transform using ECG waveforms and classification using a support vector machine [8]. The classification accuracy of the test data was approximately 77%. However, classification using ECG waveforms is often limited by the possibility of significant changes in the waveform characteristics when other diseases coexist.

Therefore, we attempted to classify AF types by extracting image features, such as left atrial diameter and structural changes in pulmonary veins due to persistent AF, from contrast-enhanced computed tomography (CT) images using convolutional neural networks (CNNs), which have been applied in medical practice in recent years [12–18]. Although previous studies using electrocardiogram waveforms have been reported in the classification of AF type, no method using contrast-enhanced CT images has been proposed. Furthermore, although there are research papers on the relationship between left atrial volume and AF type, there are no reported cases of applying that method to the classification of the type of disease. In this study, we propose a clinically novel method of classifying paroxysmal AF and long-term persistent AF on contrast-enhanced CT images using conventional CNN models, focusing on structural remodeling changes in the left atrium. The purpose of this study is to enable a standardized assessment using a deep learning approach that considers the information physicians need to evaluate the structural remodeling of the left atrium, including left atrial enlargement, poor contrast, structural changes in the pulmonary veins, the presence of thrombi in the left atrium, and coronary artery calcification. Based on this objective, contrast-enhanced CT imaging has an advantage over other dynamic modalities in that it can accurately capture the shape and focus on the structures around the left atrium. Furthermore, we hypothesize that the method using contrast-enhanced CT images will enable standardized evaluation with reduced subjective bias, even in cases in which the ECG waveform cannot detect sudden attacks, such as paroxysmal AF, or when there are concomitant diseases that may affect the ECG waveform. With the application of these systems to clinical workflows, it will be possible to evaluate the load on the atrial muscle when AF is first detected and, if signs of long-term persistence are confirmed, to begin treatment early.

In this study, we also compared the results of the CNN classification with those of physicians' clinical judgment by surveying cardiologists regarding AF type classification. Physicians estimate the type of atrial fibrillation based on factors such as the size of the left atrium, enlargement of the pulmonary veins, thrombus formation in the left atrial

appendage, and fibrosis of the atrial septum. Focusing on these features, we looked at images similar to those entered into the CNN to predict the corresponding disease type.

## 2. Materials and Methods

### 2.1. Outline

In this study, target slices were selected from contrast-enhanced CT images. The number of images was increased using data augmentation and then input into a CNN model. The output images were classified into two classes, PAF and LSAF, and the saliency map, which emphasized the pixels that contributed to the classification result using score-CAM according to their importance, was used to compare what each model focused on in the image to make its judgment. Persistent atrial fibrillation was excluded because its duration varies widely from 7 days to less than 1 year, making it difficult to accurately identify through the evaluation of the left atrial shape. This study was conducted with the approval of the ethics committee of the first author's institution (approval number HM22-095).

### 2.2. Image Dataset

This study included 60 patients with AF who underwent CE-CT at Fujita Health University Bantane Hospital between May 2021 and July 2022. A total of 162 contrast-enhanced CT scans were performed during the period, including 116 patients with paroxysmal atrial fibrillation and 46 patients with long-standing persistent atrial fibrillation. From these, 30 patients of each disease type were randomly selected, and only those patients who did not undergo CT examinations due to contrast medium allergy or impaired renal function were excluded. The patients' disease types were diagnosed as defined in the guidelines [4]. Specifically, PAF was defined as AF that returns to sinus rhythm within 7 days of onset, and LSAF was defined as AF that persists beyond 1 year. The percentages of PAF and LSAF were each half of all patients. Basic patient information is shown in Table 1. An Aquilion ONE CT system (Canon Medical Systems, Inc., Tochigi, Japan) was used to obtain the images. The details of the imaging protocol are shown in Table 2. We used transaxial images with a matrix size of  $512 \times 512$  pixels and a pixel size of 0.625 mm. The images were stored in DICOM format, and all images were converted to 8-bit PNG images with a window level of 30HU and a window width of 1000 HU.

**Table 1.** Basic patient information.

Variables	PAF(N = 30)	LSAF(N = 30)	p-Value
Age (years) (mean $\pm$ SD)	65.3 $\pm$ 12.4	69.5 $\pm$ 8.6	0.093
Gender (male, %)	19(63.3%)	25(83.3%)	0.082
Height (cm) (mean $\pm$ SD)	164.23 $\pm$ 10.2	168.2 $\pm$ 8.75	0.131
Body weight (kg) (mean $\pm$ SD)	63.7 $\pm$ 12.4	68.7 $\pm$ 11.1	0.104
BMI (mean $\pm$ SD)	23.5 $\pm$ 3.35	24.3 $\pm$ 3.42	0.309
Hypertension (cases, %)	13(43.3%)	14(46.7%)	0.799
Diabetes mellitus (cases, %)	5(16.7%)	5(16.7%)	1.000
Heart failure (cases, %)	3(10.0%)	12(40.0%)	0.007
Cerebral infarction (cases, %)	4(13.3%)	5(16.7%)	0.723

**Table 2.** Imaging protocols.

Parameter	Value	
Imaging protocols	kV	120 kV
	mAs	CT-AEC
	Slice thickness	0.5 mm
	Scan time	0.35 s
	Scan method	ECG gated volume scan

Table 2. Cont.

	Parameter	Value
Reconstruction condition	Reconstruction method	AIDR-3D
	FOV	200 mm
	Slice thickness	0.5 mm
	Slice spacing	0.25 mm
	Reconstruction function	FC14
	Reconstruction cardiac phase	Systolic
Angiographic method	Iodine concentration	375 mgI/kg
	Injection time	15 s
	Imaging timing	Bolus tracking

### 2.3. Atrial Fibrillation Type Classification Using Contrast-Enhanced CT Images

The flow of this study is shown in Figure 1, and the details of each process are described below.

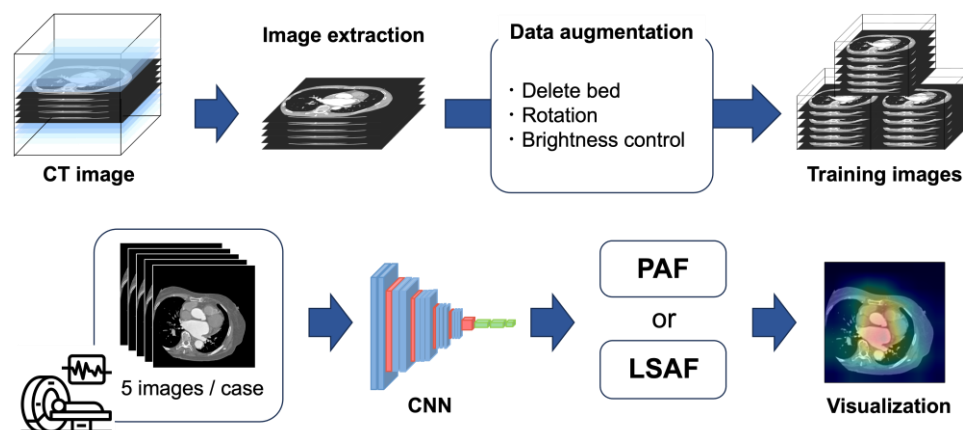


Figure 1. Process of this study.

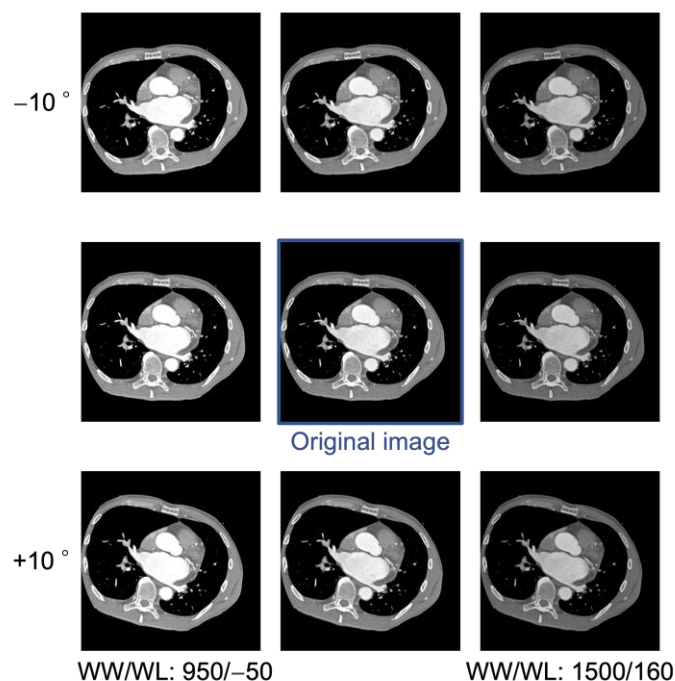
#### 2.3.1. Image Pre-Processing

Images centered on the slice, with the largest left atrium in the contrast-enhanced CT image and located 5 and 10 mm above and below, were selected, and five images per patient were used for analysis. If a bed was depicted in the image, it was removed by manually setting the CT value of the bed area to  $-1000$  HU.

#### 2.3.2. Data Augmentation

Data augmentation is a method of increasing data by “transforming” image data for training. For example, by rotating, flipping, shifting horizontally, scaling, distorting, adjusting brightness and contrast, and adding noise to an image, various variations can be created. In this study, the number of images increased nine times through data augmentation [19]. CT examinations are usually performed in the supine position; however, in some facilities, the patient is positioned so that the heart, which is located on the left side of the body, is centered in the FOV. In such cases, the curvature of the bed may cause the body to rotate about  $10^\circ$ . To simulate this, the heart was rotated by  $-10^\circ$  and  $+10^\circ$  for each image, aligning the heart’s tilt to match that observed in the actual CT image. In contrast-enhanced CT examinations, since the density of the contrast agent varies depending on the case, we augmented the pixel values to be robust to changes in pixel values. The CT values of the left atrium were observed across the entire dataset, and the window level (WL) and window width (WW) were adjusted so that the CT values after augmentation fell within the range of real CT images. As a result, in addition to the initial condition of WL = 30, WW = 1000, two variations, including WL =  $-50$ , WW = 950 and WL = 160, WW = 1500,

were added to increase the number of images threefold. An example of the created image is shown in Figure 2.



**Figure 2.** Examples of an original image and images created using data augmentation.

### 2.3.3. Classification by CNN

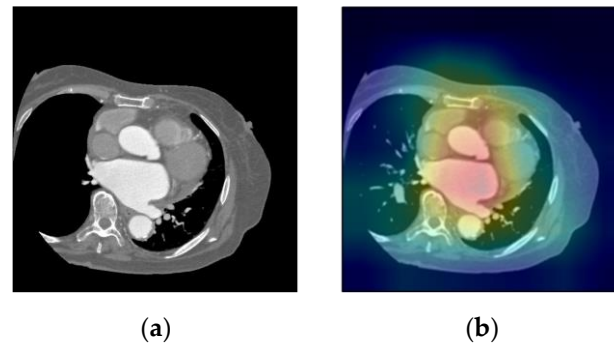
In this study, we used six network models (VGG16, VGG19, Resnet50, DenseNet121, DenseNet169, and DenseNet201). These networks were trained on 1.2 million images across 1000 categories in the ImageNet database [20–22]. To adapt these networks to PAF and LSAF classification, we removed the fully connected layers in each of the pretrained network models and replaced them with three new fully connected layers (the final layer being the output layer). The number of units in each layer was set to 1024, 256, and 2. In this study, finetuning was employed. Finetuning is a method to perform transfer learning using a different dataset for a different target task than the one used during pre-training that involves using a network model that has been pretrained from a large dataset as the initial parameters. Finetuning facilitates the learning of highly accurate models for each task from small datasets by simply recalibrating pretrained CNNs. In this case, the weights of the convolutional layer were initialized with the pretrained weights, and both the convolutional and fully connected layers were retrained (finetuning) using real images. The average of five continuous values obtained from the outputs of five slices output from the CNN was used as the patient's evaluation. In this evaluation, the cutoff value was fixed at 0.5.

For the CNN training conditions, we used a learning coefficient of 0.000001, early stopping (maximum number of epochs: 100) as the training frequency, a batch size of eight, and Adam as the optimization algorithm. The categorical cross entropy was employed for the loss function in the training of CNN. The training environment used was Windows 10 Pro OS, an AMD Ryzen 7 2700X CPU, and an NVIDIA TITAN RTX GPU.

### 2.4. Saliency Map

In this study, we used score-class activation mapping (CAM) to visualize the points of interest by highlighting the pixels that contributed to the classification results according to their importance. Score-CAM eliminates the dependence on gradients by obtaining the weight of each activation map through its forward passing score on the target class; the final result is obtained using a linear combination of weights and activation maps [23]. It

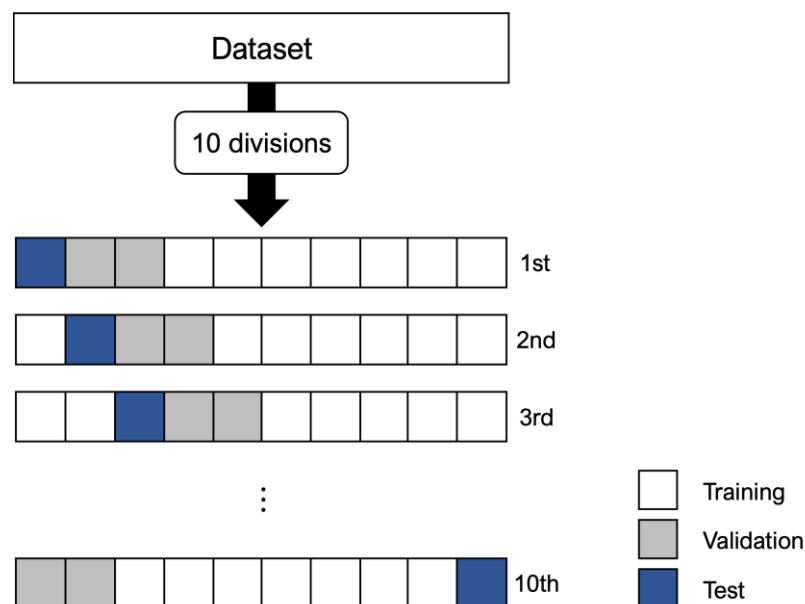
visualizes the importance based on the results obtained by providing the generated images to the CNN using the feature map obtained when the trained CNN infers a specific image. The resulting feature map was enlarged to the size of the input, normalized to a value between 0 and 1, and multiplied by the input image to generate a heatmap. The output of CAM is shown as a heatmap overlaid on the image. This heatmap is called a saliency map in CAM. The input and saliency map images are shown in Figure 3.



**Figure 3.** An example of visualization of decision basis in CNN (score-CAM). (a) Input image; (b) saliency map image.

### 2.5. Validation and Evaluation Metrics

In this study, cross-validation was used to assess the generalizability of the model. We also increased the number of folds and chose 10-fold cross-validation to improve generalization performance and reduce bias. The 10-fold cross-validation method divides the dataset into 10 subsets, 70% of which are training data and 20% of which are validation data, 10% of which are test data. Figure 4 shows a schematic of the 10-part cross-validation method.



**Figure 4.** Data assignment in the 10-part cross-validation method.

Using this method, the prediction results were compared based on patient-specific accuracy, sensitivity, specificity, and precision. The final classification performance evaluation was performed by determining the overall accuracy rate using the CNN classification results. The overall accuracy rate was calculated using the following Equation (1). TP,

TN, FP, and FN are the numbers of true positives, true negatives, false positives, and false negatives, respectively.

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100\% \quad (1)$$

The ROC curve represents the relationship between the true positive fraction (TP/TP + FN) and the false positive fraction (FP/FP + TN). It was created by plotting the false positive rate on the horizontal axis and the true positive rate on the vertical axis and continuously varying the cutoff value to separate positive and negative results. To smooth the ROC curve, the false positive fraction (FPF) and true positive fraction (TPF) were plotted on both normal probability papers to obtain an approximately straight line, and the curve depicted the relationship between the two.

The CNN was trained and evaluated thrice for each model, with the median value and standard deviation used as the final classification result. In this study, the slice with the largest left atrium and the two slices above and below it were used for training and evaluation to enable continuous evaluation of the left atrium in the direction of the body axis. In addition, the number of images used for training increased with data augmentation. To demonstrate the effectiveness of these methods, we performed an additional validation using only one central slice for training and evaluation (Additional Study 1) and a validation using an evaluation without data augmentation (Additional Study 2).

## 2.6. Classification by Physicians

In this study, we administered the same questionnaire to physicians regarding the classification of atrial fibrillation types based on only five images entered into the CNN classification, and the results were compared with the correct response rate and focus of the CNN classification.

### 2.6.1. Participants

A questionnaire survey was conducted among physicians in the Department of Cardiovascular Medicine at Fujita Health University Bantane Hospital, and responses were obtained from 18 physicians. In this survey, we asked patients to evaluate the type of AF in terms of structural changes around the left atrium. The purpose of this questionnaire was to compare the results of this study's classification with those of the physicians' clinical judgments.

### 2.6.2. Questionnaire Items

Questions included: (1) years of experience as a physician, (2) specialty, (3) number of catheter ablation procedures performed per year, (4) whether preoperative CT imaging could predict the efficacy of catheter ablation, and (5) type classification of atrial fibrillation (20 cases) and the basis for decision.

(3) The number of catheter ablation procedures performed in a year and (4) whether preoperative CT images could predict the efficacy of catheter ablation procedures were optional answers for physicians performing catheter ablation procedures. For AF classification (5), 10 cases of paroxysmal PAF and 10 cases of LSAF were randomly selected from the cases used in the CNN classification, and the results were tabulated on a 6-point scale. In addition, the basis for judgment was asked, e.g., "Please tell us the reason why you answered that way", for the answer of the disease type classification, and the answer was left open-ended. This question aimed to compare the points of interest of the CNN with those of physicians.

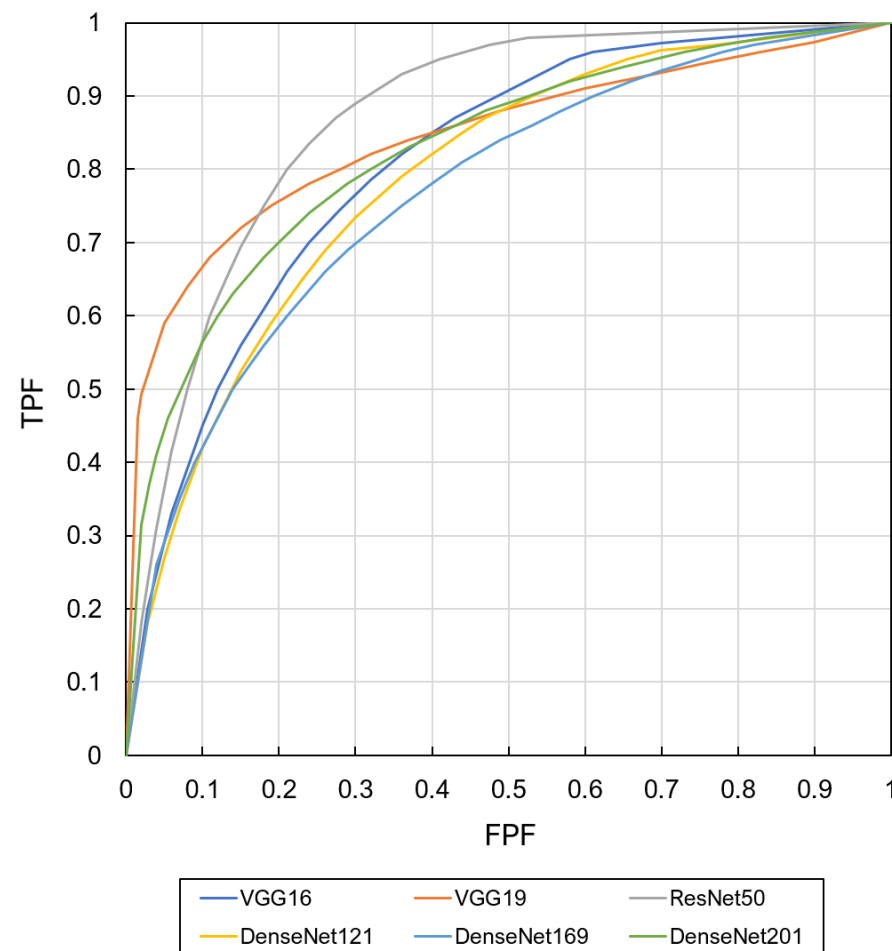
### 3. Results

#### 3.1. Classification Results by CNN

First, we describe the results of the AF type classification using a CNN. The classification results and AUC for the six CNN models are listed in Table 3, and the ROC curves are shown in Figure 5. ResNet50 exhibited the highest accuracy for all classification results.

**Table 3.** Classification results for each CNN model (proposed method).

Model	Sensitivity	Specificity	Precision	Accuracy	AUC
VGG16	$80.0 \pm 1.56$	$63.3 \pm 4.71$	$68.6 \pm 3.28$	$71.7 \pm 2.84$	$0.80 \pm 0.03$
VGG19	$80.0 \pm 1.56$	$76.7 \pm 4.15$	$77.4 \pm 2.54$	$78.3 \pm 1.56$	$0.79 \pm 0.00$
ResNet50	$83.3 \pm 5.65$	$80.0 \pm 4.15$	$80.6 \pm 3.27$	$81.7 \pm 3.60$	$0.88 \pm 0.07$
DenseNet121	$76.7 \pm 4.71$	$66.7 \pm 3.16$	$69.7 \pm 2.68$	$71.7 \pm 3.45$	$0.80 \pm 0.02$
DenseNet169	$80.0 \pm 2.74$	$63.3 \pm 4.15$	$68.6 \pm 2.59$	$71.7 \pm 2.08$	$0.76 \pm 0.03$
DenseNet201	$83.3 \pm 3.11$	$63.3 \pm 4.16$	$69.4 \pm 2.63$	$73.3 \pm 2.36$	$0.82 \pm 0.01$



**Figure 5.** ROC curves of CNN models.

The results of the additional validation are presented in Tables 4 and 5. In addition, Figure 6 shows a comparison of the classification correctness rate between the proposed method and Additional Studies 1 and 2. When learning and evaluation were performed on the central slice only (Additional Study 1), the classification correctness increased for VGG16, VGG19, and ResNet50 but decreased for the other three DenseNet models. Without augmentation of the training data (Additional Study 2), the accuracy remained the same or decreased for models other than DenseNet169.

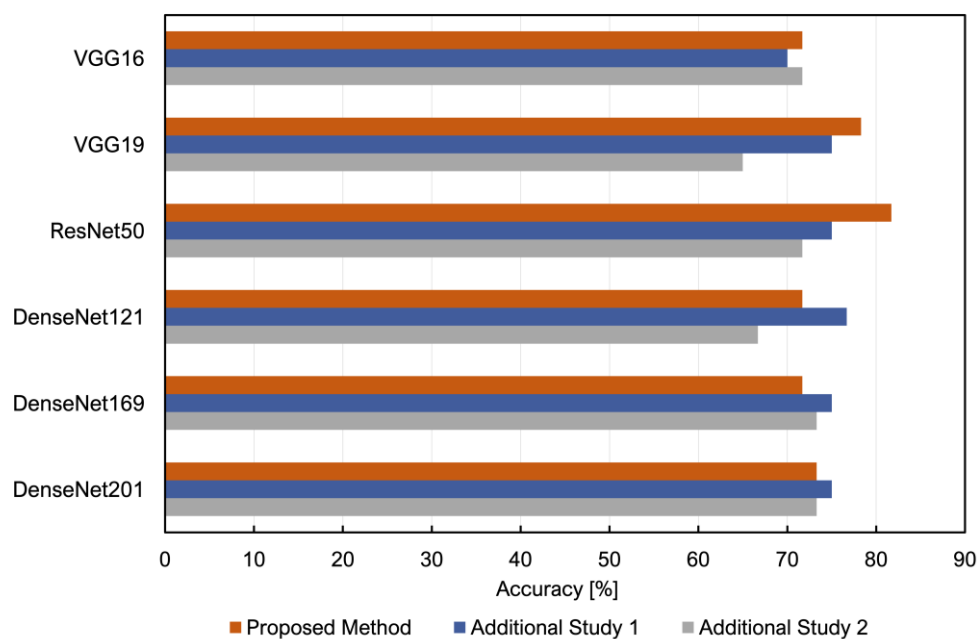


**Table 4.** Classification results for evaluation of central slices only (Additional Study 1).

Model	Sensitivity	Specificity	Precision	Accuracy	AUC
VGG16	76.7 ± 1.52	63.3 ± 4.71	67.6 ± 3.34	70.0 ± 2.84	0.75 ± 0.02
VGG19	70.0 ± 4.16	80.0 ± 1.56	77.8 ± 2.05	75.0 ± 2.36	0.78 ± 0.01
ResNet50	73.3 ± 3.16	76.7 ± 2.74	75.9 ± 1.39	75.0 ± 0.80	0.83 ± 0.01
DenseNet121	73.3 ± 1.56	80.0 ± 1.56	78.6 ± 1.27	76.7 ± 0.80	0.82 ± 0.01
DenseNet169	73.3 ± 2.74	76.7 ± 0.00	75.9 ± 0.69	75.0 ± 1.39	0.77 ± 0.01
DenseNet201	66.7 ± 3.11	83.3 ± 0.00	80.0 ± 0.71	75.0 ± 1.56	0.82 ± 0.02

**Table 5.** Classification results without data augmentation (Additional Study 2).

Model	Sensitivity	Specificity	Precision	Accuracy	AUC
VGG16	66.7 ± 3.11	76.7 ± 4.15	74.1 ± 3.98	71.7 ± 2.81	0.75 ± 0.03
VGG19	63.3 ± 4.16	66.7 ± 6.27	65.5 ± 5.26	65.0 ± 4.08	0.70 ± 0.03
ResNet50	60.0 ± 5.43	83.3 ± 4.16	78.3 ± 2.11	71.7 ± 0.75	0.81 ± 0.01
DenseNet121	63.3 ± 1.56	70.0 ± 6.86	67.9 ± 6.19	66.7 ± 3.60	0.72 ± 0.04
DenseNet169	70.0 ± 2.69	76.7 ± 0.00	75.0 ± 0.73	73.3 ± 1.35	0.77 ± 0.02
DenseNet201	63.3 ± 4.16	83.3 ± 1.60	79.2 ± 2.50	73.3 ± 3.37	0.84 ± 0.02

**Figure 6.** Comparison of proposed method and additional study.

The images that were correctly classified by ResNet50 are shown in Figure 7, and those that were incorrectly classified are shown in Figure 8.

Figure 9 shows the saliency map output when ResNet50 correctly classifies a case, and Figure 10 shows the heatmap output when ResNet50 incorrectly classifies a case. Note that the presented case is the same patient as the one presented in Figures 7 and 8.

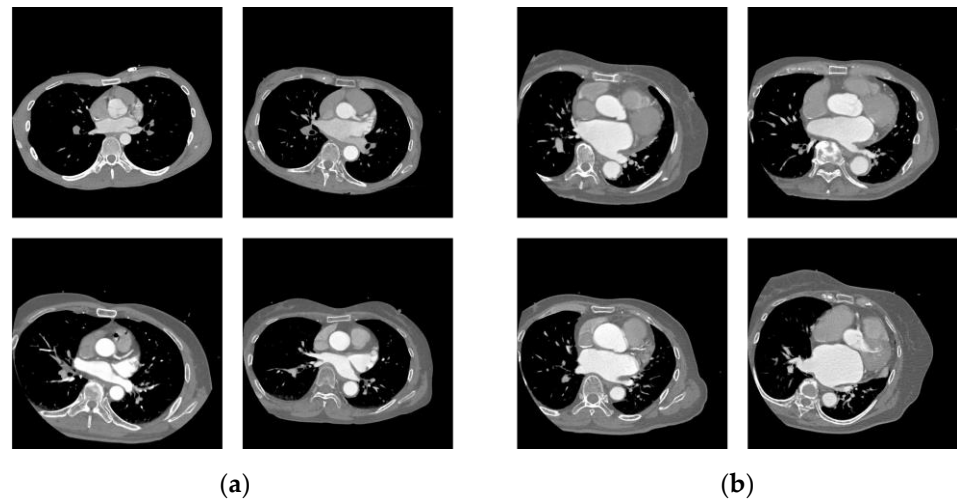


Figure 7. Correctly classified cases. (a) PAF; (b) LSAF.

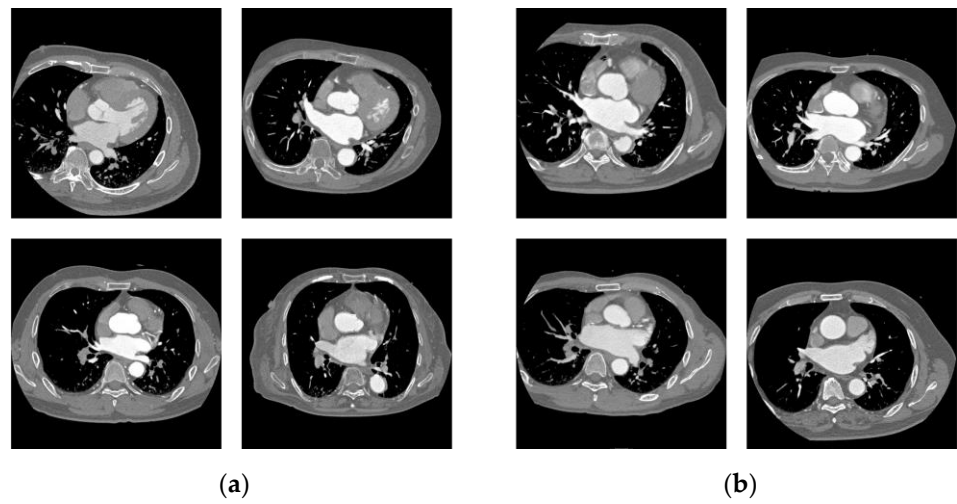


Figure 8. Incorrectly classified cases. (a) PAF; (b) LSAF.

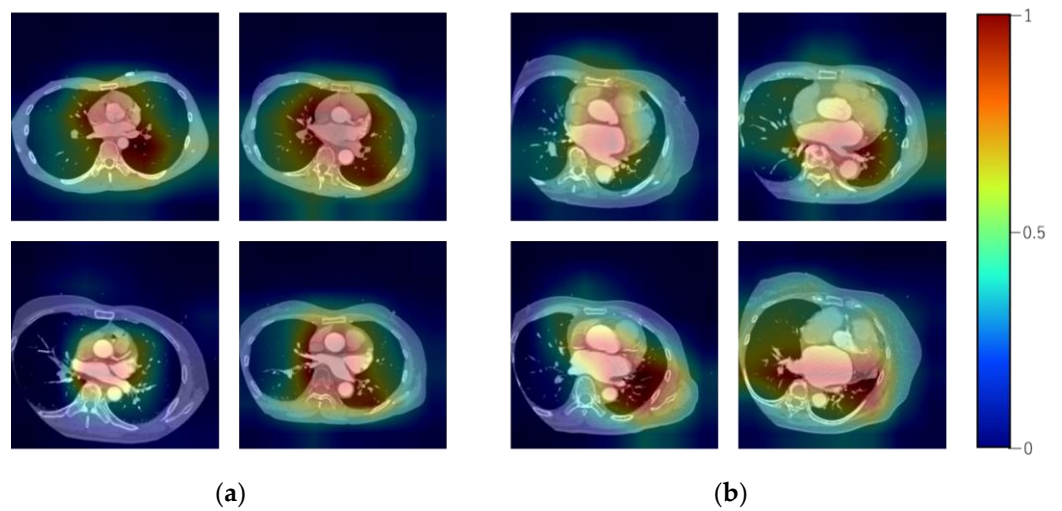
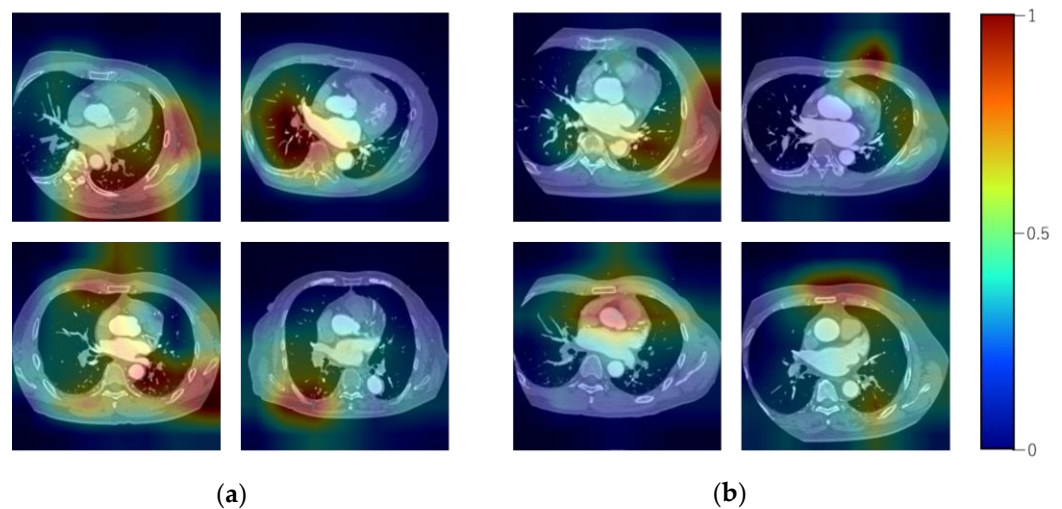


Figure 9. Saliency maps of correctly classified cases. (a) PAF; (b) LSAF.



**Figure 10.** Saliency maps of incorrectly classified cases. (a) PAF; (b) LSAF.

### 3.2. Classification Results by Physicians

Table 6 shows the number (%) of responses to each of the following questions: (1) years of experience as a physician, (2) specialty, and (3) number of catheter ablation therapies performed per year. In Case (2), all 18 physicians specialized in cardiovascular medicine.

**Table 6.** Survey results on years of experience as a physician and areas of specialization.

Experience (years)	Responses (%)	Specialty	Responses (%)
5–10	3 (16.7)	cardiovascular	18 (100)
11–15	4 (22.2)		
16–20	7 (38.9)		
21–25	2 (11.1)		
26–30	1 (5.6)		
31–35	1 (5.6)		
Total	18 (100)	Total	18 (100)

Six physicians responded to question (3), the number of catheter ablation therapy performed in a year. The results are summarized in Table 7.

**Table 7.** Survey results of the number of catheter ablation therapies performed in a year.

Number of Treatments (Cases)	Responses (%)
1–50	2 (33.3)
51–100	2 (33.3)
101–150	1 (16.7)
151–200	1 (16.7)
Total	6 (100)

Nine physicians responded to the question about (4) whether preoperative CT images could predict the efficacy of catheter ablation therapy. Of these, eight physicians answered that preoperative contrast-enhanced CT could predict the efficacy of catheter ablation therapy.

Figure 11 shows the percentage of correct answers for the 20 cases used in the questionnaire classified by ResNet50, the percentage of correct answers for 18 physicians, and the average percentage for all physicians. In addition, Figure 12 shows the ROC of the physicians' classification results, and Table 8 shows details of the physicians' classification accuracy and AUC. The mean accuracy was 73.6% and the median was 75%. The mean

AUC was 0.802. The 20 cases used to evaluate ResNet50 were the same as those used in the survey of physicians, and the overall correct response rate for physicians was widely distributed, ranging from 55% to 90%; however, the average correct response rate was 73.6%, which was slightly lower than that of ResNet50.

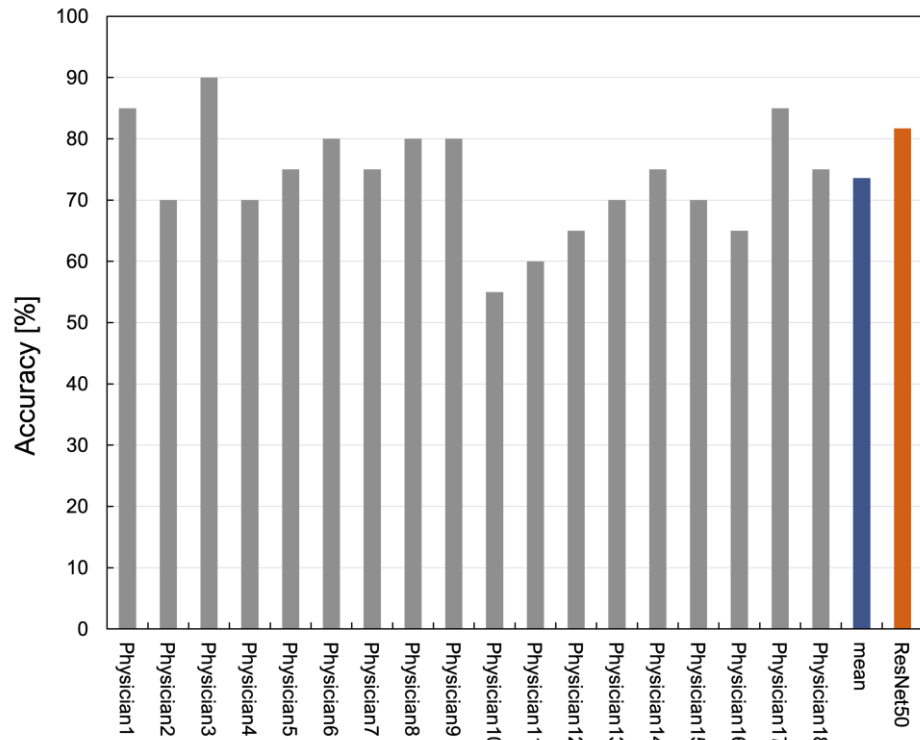


Figure 11. Physicians’ classification results and comparison between CNN models.

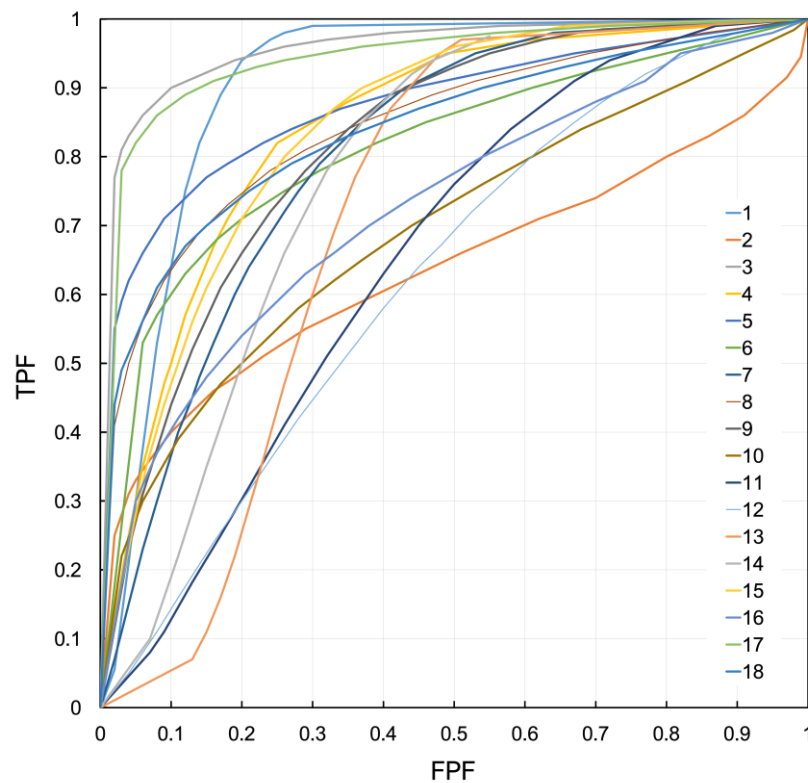


Figure 12. ROC curves of physicians.

**Table 8.** Physicians' classification results.

Physician	Accuracy(%)	AUC	Physician	Accuracy(%)	AUC
1	85.0	0.895	10	55.0	0.690
2	70.0	0.640	11	60.0	0.670
3	90.0	0.955	12	65.0	0.675
4	70.0	0.820	13	70.0	0.770
5	75.0	0.865	14	75.0	0.780
6	80.0	0.825	15	70.0	0.800
7	75.0	0.825	16	65.0	0.740
8	80.0	0.830	17	85.0	0.945
9	80.0	0.885	18	55.0	0.825

The respondents had diverse opinions based on their judgments. Generally, LSAF is characterized by left atrial enlargement, roundness of the left atrium, coronary artery calcification, left auricular enlargement, poor contrast, auricular thrombus closure, uneven contrast density, retraction of the comb muscle, atrial wall thickening, and fibrosis of the atrial septum. The most common finding of persistent atrial fibrillation is enlargement of the left atrium.

#### 4. Discussion

##### 4.1. Comparison of CNN Models

In this study, six CNN models were evaluated on their performance in classifying the AF types. ResNet50 performed the best in terms of overall accuracy, followed by VGG19. The reason these CNN outperformed DenseNet121, 169, and 201 could be that the number of layers in the network was shallow, which made it possible to extract features in a localized region. The long-term persistence of AF results in structural remodeling, such as left atrial shape changes and auricular enlargement, also affected the results. Therefore, ResNet50 and VGG19 should focus on these localized areas for classification purposes. The best overall correct response rate for ResNet50 was achieved because ResNet50 is optimized using a residual function and performs batch normalization for each residual block. We hypothesize that this resulted in stable learning without the gradient vanishing problem.

In addition, Figure 6 shows a comparison of the classification correctness rate between the proposed method and Additional Studies 1 and 2. In most cases, the proposed method performs better than Additional Studies 1 and 2. The reason for the better accuracy rate than that of Additional Study 1 is that the proposed method uses a total of five slices (located 5 mm above and below) centered on the slice with the most enlarged left atrium for training; therefore, it is possible to analyze information in the body axis direction, in addition to the slice direction, and classification is more accurate than when only one slice is used for evaluation. The reason for the higher rate of correct answers compared to Additional Study 2 is thought to be that the data augmentation increased the number of pseudo-variations because of the various body inclinations and CT values due to the contrast agent and was able to respond to the effects on the image caused during imaging. Furthermore, data augmentation increased the number of images used for training by a factor of nine; therefore, it was assumed that efficient training was possible.

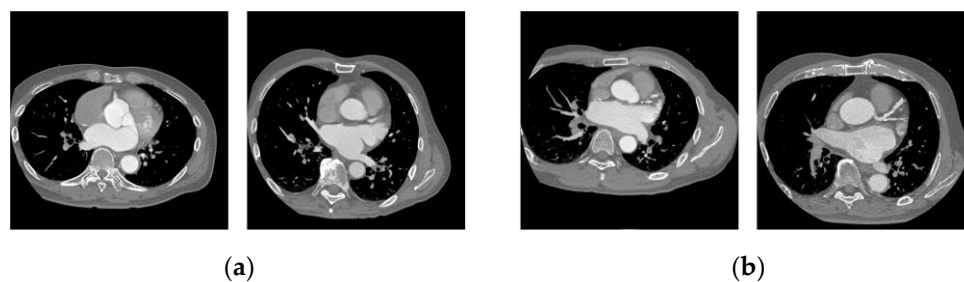
##### 4.2. Insights from Saliency Map in CNN Classifications

Score-CAM was used to output a color map showing the pixels contributing to the CNN classification results. In the heatmap output for the correct classification in Figure 8, the left atrium and pulmonary veins tended to attract more CNN attention. In addition, when attention was focused on structures other than the heart, which was often seen in the heatmap output when the patient was incorrectly classified, as shown in Figure 10, there was a tendency toward incorrect classification. Focusing on the left atrium, cases of PAF were misclassified with findings of major LSAF, including an enlarged left atrium, the loss of comb-like muscular structures, and large rounded anterior and posterior structural

left atria. In the cases of LSAF, there was also a tendency to misclassify cases in which the left atrium was not enlarged, especially when the anteroposterior diameter of the left atrium was short. Based on these findings, CNN classification focuses on the shape and surrounding structures of the left atrium and is considered a valid classification for the findings of LSAF.

#### 4.3. Comparison with Physician's Results

In response to the physician's description of the basis for judgment, enlargement of the left atrium is a feature of LSAF in many cases. In the correctly classified cases shown in Figure 7, (a) the PAF has a small, flat left atrial structure, whereas (b) the LSAF has a large, rounded left atrial structure in the front and back. The CNN model is expected to classify patients using the same criteria as physicians because the heatmap also shows that the left atrium area attracts more attention. The cases in which the CNN model and averaged results of the physicians' responses differed are shown in Figure 13. Case (a) involved LSAF, but the left atrium was relatively small (left), and there was no loss of the pectinate muscle structure (right). The CNN model can classify these cases. However, it was misclassified, even when the typical findings of LSAF in the size of the left atrium were observed, as shown in (b). The possible reason is that by using the entire CT image as the input image, information other than the left atrial region may have led to misclassification. This problem could be improved by increasing the variation with more training data and narrowing the field of view to the left atrial region alone.



**Figure 13.** LSAF cases with different results between physicians and the proposed method. (a) Correctly classified only by CNN model; (b) correctly classified only by physician.

#### 4.4. Comparison with Previous Studies

The results of this study showed a higher accuracy than those of the study by Ortigosa et al. using ECG (classification accuracy rate 77.1%) [10]. Furthermore, the method of this study has the advantage of being able to classify the pathology of AF using the assessment of structural remodeling of the left atrium, even when other diseases that affect the ECG waveform are present at the same time. AF is usually detected using an ECG, but we think that the limitation of using an ECG is that the time of detection of an attack is considered to be the moment of the first appearance of the attack. The advantage of this study using contrast CT images is that it allows for an objective evaluation of the state of the atrium regardless of the type of disease. We think that by evaluating the strain on the atrial muscle when atrial fibrillation is first discovered and confirming long-term findings, it will be possible to get closer to starting treatment at an earlier stage.

#### 4.5. Practical Applications in Clinical Settings

We hypothesize that by using deep learning to classify AF types from CT images, this study will facilitate a standardized assessment of structural remodeling of the left atrium, which was originally determined subjectively by physicians, thereby reducing subjective bias. By integrating these systems into clinical workflows, it will become possible to evaluate the strain on the atrial muscle at the initial detection of AF. Additionally, if signs of long-term persistence are confirmed, early treatment can be initiated. This approach could potentially reduce unnecessary catheter ablation procedures, allow for more tailored

treatment recommendations, and decrease healthcare costs. Furthermore, computational resources and processing time need to be discussed for practical application. Although model training requires substantial hardware resources and prolonged processing time (2–5 h), we believe that once the model is trained, the prediction process can be completed in under one minute, making it sufficiently feasible for clinical use because of the reduced hardware requirements for inference.

#### 4.6. Limitation of This Study

There are two limitations of this study. The first is that it is a small and single-facility dataset. Furthermore, potential confounding factors, such as patient comorbidities, are not discussed. When the number of data is increased and external validation is performed in the future, comorbidities should be included in the analysis and evaluated. In addition, contrast-enhanced CT provides a clearer image of the left atrium than simple CT, but patients who cannot use contrast media and variations in contrast media and image quality among facilities remain a challenge. We hypothesize that this challenge can be resolved by using simple CT images or by preparing a dataset that includes images taken at other facilities and performing data augmentation, as in this study. The second limitation is that the classification does not include persistent AF, which we think does not allow for continuous evaluation. The definition of the duration of persistent AF ranges from 7 days to less than 1 year, making it difficult to accurately identify it through the assessment of left atrial geometry. Therefore, persistent atrial fibrillation was excluded from classification in this study and classified as paroxysmal and long-standing persistent; these cases have predominantly different results in ablation therapy and can be evaluated for structural remodeling based on imaging features. In the future, it is necessary to develop a method to evaluate AF types continuously by adding cases of persistent AF. The use of left atrial volume, dynamic modality information, additional machine learning models, and natural language processing models is also possible and will be explored.

## 5. Conclusions

Catheter ablation therapy is a treatment for AF; however, its efficacy is not well established due to the high recurrence rate in patients with PAF. In this study, we attempted to classify AF types using a convolutional neural network based on features obtained from contrast-enhanced CT images. As a result of the classification, ResNet50, which is a CNN model, showed the best performance in terms of the overall correct response rate and AUC value. The output of the heatmap and the survey of physicians' judgment criteria indicated that many patients tend to focus on the shape of the left atrium in both classifications, suggesting that this method can classify AF types more accurately than physicians in a manner similar to the physicians' judgment criteria. In the future, we plan to address the challenges of this study, such as using plain CT images, preparing a dataset that includes images from other facilities, and conducting continuous evaluations that include persistent AF. Furthermore, once these issues are resolved, this study can potentially be applied in predicting the efficacy of catheter ablation therapy. A future direction is to predict the efficacy of catheter ablation therapy in patients with atrial fibrillation based on contrast-enhanced CT images with the goal of providing quality information for patients to choose their treatment options.

**Author Contributions:** Conceptualization, H.K. and A.T.; formal analysis, H.K. and A.T.; methodology, H.K. and A.T.; data curation, T.O. and Y.S.; software, H.K. and A.T.; writing—original draft preparation, H.K. and A.T.; writing—review and editing, T.O., Y.S., E.W., and H.F. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** This study was approved by the Ethical Review Committee of Fujita Health University (HM22-095) and was carried out in accordance with the World Medical Association's Declaration of Helsinki.

**Informed Consent Statement:** Informed consent was obtained in the form of an opt-out at Fujita Health University Bantane Hospital, and all data were anonymized.

**Data Availability Statement:** The source code and additional information used to support the findings of this study are available from the corresponding author upon request.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Morillo, C.A.; Banerjee, A.; Perel, P.; Wood, D.; Jouven, X. Atrial fibrillation: The current epidemic. *J. Geriatr. Cardiol.* **2017**, *14*, 195–203. Available online: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5460066> (accessed on 9 December 2023). [PubMed]
2. Kirchhof, P.; Benussi, S.; Kotecha, D.; Ahlsson, A.; Atar, D.; Casadei, B.; Castella, M.; Diener, H.-C.; Heidbuchel, H.; Hendriks, J.; et al. 2016 ESC guidelines for the management of atrial fibrillation developed in collaboration with EACTS. *Eur. Heart J.* **2016**, *37*, 2893–2962. [CrossRef] [PubMed]
3. Developed with the special contribution of the European Heart Rhythm Association (EHRA); Endorsed by the European Association for Cardio-Thoracic Surgery (EACTS); Camm, A.J.; Kirchhof, P.; Lip, G.Y.; Schotten, U.; Savelieva, I.; Ernst, S.; Van Gelder, I.C.; Al-Attar, N.; et al. Guidelines for the management of atrial fibrillation: The Task Force for the Management of Atrial Fibrillation of the European Society of Cardiology (ESC). *Eur. Heart J.* **2010**, *31*, 2369–2429. [CrossRef]
4. Nogami, A.; Kurita, T.; Abe, H.; Ando, K.; Ishikawa, T.; Imai, K.; Usui, A.; Okishige, K.; Kusano, K.; Kumagai, K.; et al. 2018 Revised Guidelines for Non-Pharmacologic Treatment of Arrhythmia. *Circ. J.* **2021**, *85*, 1692–1700. [CrossRef] [PubMed]
5. Sultan, A.; Lüker, J.; Andresen, D.; Kuck, K.H.; Hoffmann, E.; Brachmann, J.; Hochadel, M.; Willems, S.; Eckardt, L.; Lewalter, T.; et al. Predictors of Atrial Fibrillation Recurrence after Catheter Ablation: Data from the German Ablation Registry. *Sci. Rep.* **2017**, *7*, 16678. [CrossRef] [PubMed]
6. Njoku, A.; Kannabhiran, M.; Arora, R.; Reddy, P.; Gopinathannair, R.; Lakkireddy, D.; Dominic, P. Left atrial volume predicts atrial fibrillation recurrence after radiofrequency ablation: A meta-analysis. *EP Eur.* **2018**, *20*, 33–42. [CrossRef] [PubMed]
7. Zhou, X.; Nakamura, K.; Sahara, N.; Takagi, T.; Toyoda, Y.; Enomoto, Y.; Hara, H.; Noro, M.; Sugi, K.; Moroi, M.; et al. Deep Learning-Based Recurrence Prediction of Atrial Fibrillation After Catheter Ablation. *Circ. J.* **2022**, *86*, 299–308. [CrossRef] [PubMed]
8. Kim, J.Y.; Kim, Y.; Oh, G.-H.; Choi, Y.; Hwang, Y.; Kim, T.-S.; Kim, S.-H.; Kim, J.-H.; Jang, S.-W.; Oh, Y.-S.; et al. A deep learning model to predict recurrence of atrial fibrillation after pulmonary vein isolation. *Int. J. Arrhythmia* **2020**, *21*, 19. [CrossRef]
9. McGann, C.; Akoum, N.; Patel, A.; Kholmovski, E.; Revelo, P.; Damal, K.; Wilson, B.; Cates, J.; Harrison, A.; Ranjan, R.; et al. Atrial Fibrillation Ablation Outcome Is Predicted by Left Atrial Remodeling on MRI. *Circ. Arrhythmia Electrophysiol.* **2014**, *7*, 23–30. [CrossRef] [PubMed]
10. Ortigosa, N.; Cano, Ó.; Ayala, G.; Galbis, A.; Fernández, C. Atrial fibrillation subtypes classification using the General Fourier-family Transform. *Med. Eng. Phys.* **2014**, *36*, 554–560. [CrossRef] [PubMed]
11. Alcaraz, R.; Sandberg, F.; Sörnmo, L.; Rieta, J.J. Classification of Paroxysmal and Persistent Atrial Fibrillation in Ambulatory ECG Recordings. *IEEE Trans. Biomed. Eng.* **2011**, *58*, 1441–1449. [CrossRef] [PubMed]
12. Fujita, H. AI-based computer-aided diagnosis (AI-CAD): The latest review to read first. *Radiol. Phys. Technol.* **2020**, *13*, 6–19. [CrossRef] [PubMed]
13. Suman, G.; Panda, A.; Korfiatis, P.; Goenka, A.H. Convolutional neural network for the detection of pancreatic cancer on CT scans. *Lancet Digit. Health* **2020**, *2*, 453. [CrossRef] [PubMed]
14. Xiang, L.; Wang, Q.; Nie, D.; Zhang, L.; Jin, X.; Qiao, Y.; Shen, D. Deep embedding convolutional neural network for synthesizing CT image from T1-Weighted MR image. *Med. Image Anal.* **2018**, *47*, 31–44. [CrossRef] [PubMed]
15. Teramoto, A.; Fujita, H.; Yamamuro, O.; Tamaki, T. Automated detection of pulmonary nodules in PET/CT images: Ensemble false-positive reduction using a convolutional neural network technique. *Med. Phys.* **2016**, *43*, 2821–2827. [CrossRef] [PubMed]
16. Wang, Q.; Shen, F.; Shen, L.; Huang, J.; Sheng, W. Lung Nodule Detection in CT Images Using a Raw Patch-Based Convolutional Neural Network. *J. Digit. Imaging* **2019**, *32*, 971–979. [CrossRef] [PubMed]
17. Liu, C.; Cao, Y.; Alcantara, M.; Liu, B.; Brunette, M.; Peinado, J.; Curioso, W. TX-CNN: Detecting tuberculosis in chest X-ray images using convolutional neural network. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; Volume 23, pp. 14–18.
18. Rohini, A.; Praveen, C.; Mathivanan, S.K.; Muthukumar, V.; Mallik, S.; Alqahtani, M.S.; Al-Rasheed, A.; Soufiene, B.O. Multimodal hybrid convolutional neural network based brain tumor grade classification. *BMC Bioinform.* **2023**, *24*, 382. [CrossRef] [PubMed]
19. Shorten, C.; Khoshgoftaar, T.M. A survey on image data augmentation for deep learning. *J. Big Data* **2019**, *6*, 60. [CrossRef]
20. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In Proceedings of the 3rd International Conference on Learning Representations (ICLR), San Diego, CA, USA, 7–9 May 2015; pp. 11–14.
21. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.



22. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269.
23. Wang, H.; Wang, Z.; Du, M.; Yang, F.; Zhang, Z.; Ding, S.; Mardziel, P.; Hu, X. Score-CAM: Score-Weighted Visual Explanations for Convolutional Neural Networks. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 111–119.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.