






## Article

# Novel Domain Knowledge-Encoding Algorithm Enables Label-Efficient Deep Learning for Cardiac CT Segmentation to Guide Atrial Fibrillation Treatment in a Pilot Dataset

Prasanth Ganesan <sup>1,†</sup>, Ruibin Feng <sup>1,†</sup> , Brototo Deb <sup>1</sup> , Fleur V. Y. Tjong <sup>1,2</sup>, Albert J. Rogers <sup>1</sup>, Samuel Ruipérez-Campillo <sup>1,3</sup> , Sulaiman Somani <sup>1</sup>, Paul Clopton <sup>1</sup>, Tina Baykaner <sup>1</sup>, Miguel Rodrigo <sup>1,4</sup> , James Zou <sup>5</sup>, Francois Haddad <sup>1</sup>, Matei Zaharia <sup>6</sup> and Sanjiv M. Narayan <sup>1,\*</sup> 

<sup>1</sup> Department of Medicine and Stanford Cardiovascular Institute (CVI), Stanford University, Stanford, CA 94305, USA; prash030@stanford.edu (P.G.); ruibin@stanford.edu (R.F.)

<sup>2</sup> Heart Center, Department of Clinical and Experimental Cardiology, Amsterdam UMC, University of Amsterdam, 1105 AZ Amsterdam, The Netherlands

<sup>3</sup> Department of Computer Science, ETH Zurich, 8092 Zurich, Switzerland

<sup>4</sup> CoMMLab, Universitat de València, 46100 Valencia, Spain

<sup>5</sup> Department of Biomedical Data Science, Stanford University, Stanford, CA 94305, USA

<sup>6</sup> Department of Computer Science, University of California Berkeley, Berkeley, CA 94720, USA

\* Correspondence: sanjiv1@stanford.edu

† These authors contributed equally to this work.



**Citation:** Ganesan, P.; Feng, R.; Deb, B.; Tjong, F.V.Y.; Rogers, A.J.; Ruipérez-Campillo, S.; Somani, S.; Clopton, P.; Baykaner, T.; Rodrigo, M.; et al. Novel Domain Knowledge-Encoding Algorithm Enables Label-Efficient Deep Learning for Cardiac CT Segmentation to Guide Atrial Fibrillation Treatment in a Pilot Dataset. *Diagnostics* **2024**, *14*, 1538. <https://doi.org/10.3390/diagnostics14141538>

Academic Editor: Ali Gholamrezanezhad

Received: 15 May 2024

Revised: 7 July 2024

Accepted: 10 July 2024

Published: 17 July 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/>).

**Abstract: Background:** Segmenting computed tomography (CT) is crucial in various clinical applications, such as tailoring personalized cardiac ablation for managing cardiac arrhythmias. Automating segmentation through machine learning (ML) is hindered by the necessity for large, labeled training data, which can be challenging to obtain. This article proposes a novel approach for automated, robust labeling using domain knowledge to achieve high-performance segmentation by ML from a small training set. The approach, the domain knowledge-encoding (DOKEN) algorithm, reduces the reliance on large training datasets by encoding cardiac geometry while automatically labeling the training set. The method was validated in a hold-out dataset of CT results from an atrial fibrillation (AF) ablation study. **Methods:** The DOKEN algorithm parses left atrial (LA) structures, extracts “anatomical knowledge” by leveraging digital LA models (available publicly), and then applies this knowledge to achieve high ML segmentation performance with a small number of training samples. The DOKEN-labeled training set was used to train a nnU-Net deep neural network (DNN) model for segmenting cardiac CT in  $N = 20$  patients. Subsequently, the method was tested in a hold-out set with  $N = 100$  patients (five times larger than training set) who underwent AF ablation. **Results:** The DOKEN algorithm integrated with the nn-Unet model achieved high segmentation performance with few training samples, with a training to test ratio of 1:5. The Dice score of the DOKEN-enhanced model was 96.7% (IQR: 95.3% to 97.7%), with a median error in surface distance of boundaries of 1.51 mm (IQR: 0.72 to 3.12) and a mean centroid–boundary distance of 1.16 mm (95% CI:  $-4.57$  to  $6.89$ ), similar to expert results ( $r = 0.99$ ;  $p < 0.001$ ). In digital hearts, the novel DOKEN approach segmented the LA structures with a mean difference for the centroid–boundary distances of  $-0.27$  mm (95% CI:  $-3.87$  to  $3.33$ ;  $r = 0.99$ ;  $p < 0.0001$ ). **Conclusions:** The proposed novel domain knowledge-encoding algorithm was able to perform the segmentation of six substructures of the LA, reducing the need for large training data sets. The combination of domain knowledge encoding and a machine learning approach could reduce the dependence of ML on large training datasets and could potentially be applied to AF ablation procedures and extended in the future to other imaging, 3D printing, and data science applications.

**Keywords:** cardiac CT segmentation; machine learning; domain knowledge encoding; atrial fibrillation; ablation

## 1. Introduction

The segmentation of cardiac computed tomography (CT) images has historically been performed by semi-automated algorithms such as graph-cuts [1], region growing [2] with manual seed inputs, and other traditional image-processing methods. Deep neural networks (DNN) showed superior performance to traditional image processing even for complex tasks such as segmenting a person of interest in an image of a crowded street [3] or classifying complex diseases from radiology scans [4,5]. However, DNN models require a large amount of training data, which, in the context of cardiac CT segmentation, is challenging to obtain. Several publicly available medical datasets include <100 cases [6–8] due to technical, privacy, and regulatory concerns. Since deep learning typically reserves the majority of cases for training, models are thus often tested on <40 cases [8], which may limit generalizability [9,10].

This begs the question as to whether high DNN performance can be obtained when the number of training samples is smaller than that of test-set samples. The focus of this work is to explore the idea of achieving high DNN segmentation performance in cardiac CT images from a small number of training samples. A DNN was applied to raw CT images to segment the left atrium (LA) body and other LA substructures: four pulmonary veins (PVs) and one LA appendage (LAA), which are central to treating patients with atrial fibrillation (AF). Although this paper focuses on a label-efficient segmentation approach of the six LA substructures for AF application, the approach can theoretically be extended to segment other chambers of the heart as well.

This article proposes a novel approach called the domain knowledge-encoding (DOKEN) algorithm, which extracts “anatomical knowledge” by leveraging digital LA models (available publicly) and then applies this knowledge to achieve high DNN segmentation performance with small number of training samples. The DOKEN algorithm essentially pre-processes the training samples before inputting them for DNN training. The pre-processing involves automatic labeling to obtain robust ground-truth labels of LA substructures. The performance of the DOKEN-labeled DNN model was tested in a hold-out dataset >5 times larger than the training set.

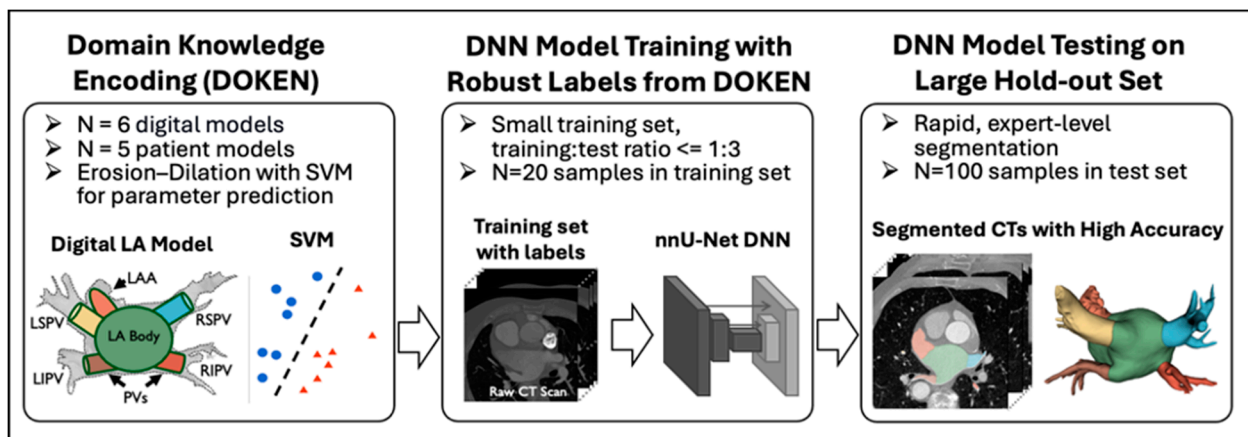
The purpose of this study is to test the hypothesis that ML models could be trained using very small datasets if combined with some domain knowledge of the task at hand. This method of training using conceptual domain knowledge principles rather than massive training data [11,12] is analogous to how humans can learn from small data [12]. Lake et al. used this approach to generate handwritten characters with human-level performance from one exemplar by parsing characters into simple primitives that were composited to create new characters [13]. However, for medical image analysis, such domain knowledge has rarely been used to reduce training sizes for DNN [14,15]. In the following sections, we describe the methods, results, discussion and conclusions.

## 2. Methods

Figure 1 outlines the method. (1) The proposed DOKEN algorithm encoded domain knowledge of the LA body and other anatomies; (2) the algorithm was used to train a nnU-Net DNN to segment cardiac CT images using only a small training set; and (3) the trained DNN was tested in a large hold-out set.

### 2.1. Dataset for Training and Testing

The CT dataset used in this study consists of  $N = 120$  patients who had undergone AF ablation between October 2014 and July 2019 and had cardiac CT scans. All patients signed informed consent at Stanford Health Care. We split this dataset randomly into  $N = 20$  for DNN model training (Training Set), with  $N = 100$  patients as a hold-out test set (Test Set). Note that the number of samples in the training set is 5 times smaller than the test set samples, which is one of the key contributions of this study. Separately, for developing the DOKEN algorithm,  $N = 6$  publicly available 3D digital heart models built using Gaussian process morphable models [16] was used.



**Figure 1.** Method overview: proposed domain knowledge-encoding algorithm used to label CT images for efficient DNN training on a training set significantly smaller than the test set. LA: left atrium, LSPV: left superior pulmonary vein, LIPV: left inferior pulmonary vein, RSPV: right superior pulmonary. Each color represents an LA sub-structure.

## 2.2. Domain Knowledge-Encoding (DOKEN) Algorithm

The goal of the DOKEN algorithm is to automatically generate robust ground-truth labels of LA substructures for DNN training. The algorithm consists of the following two steps:

- I. Segmentation of digital LA models:  $N = 6$  digital LA models (publicly available) were segmented based on an iterative erosion–dilation (ED) process (Figure 2).
- II. Tuning ED parameter using patient LA models: The iterative ED process requires optimal iteration number as a parameter, which decides the accurate segmentation of the LA body and other substructures. To determine this parameter, 5 manually segmented LA models were used to train support vector machines (SVMs) to predict the optimal iteration in the ED process.

The two steps are used to develop the DOKEN algorithm, and once its developed, it takes training images as input and generates ground-truth labels as output. The two development steps are detailed below.

### I. Segmentation of digital LA models

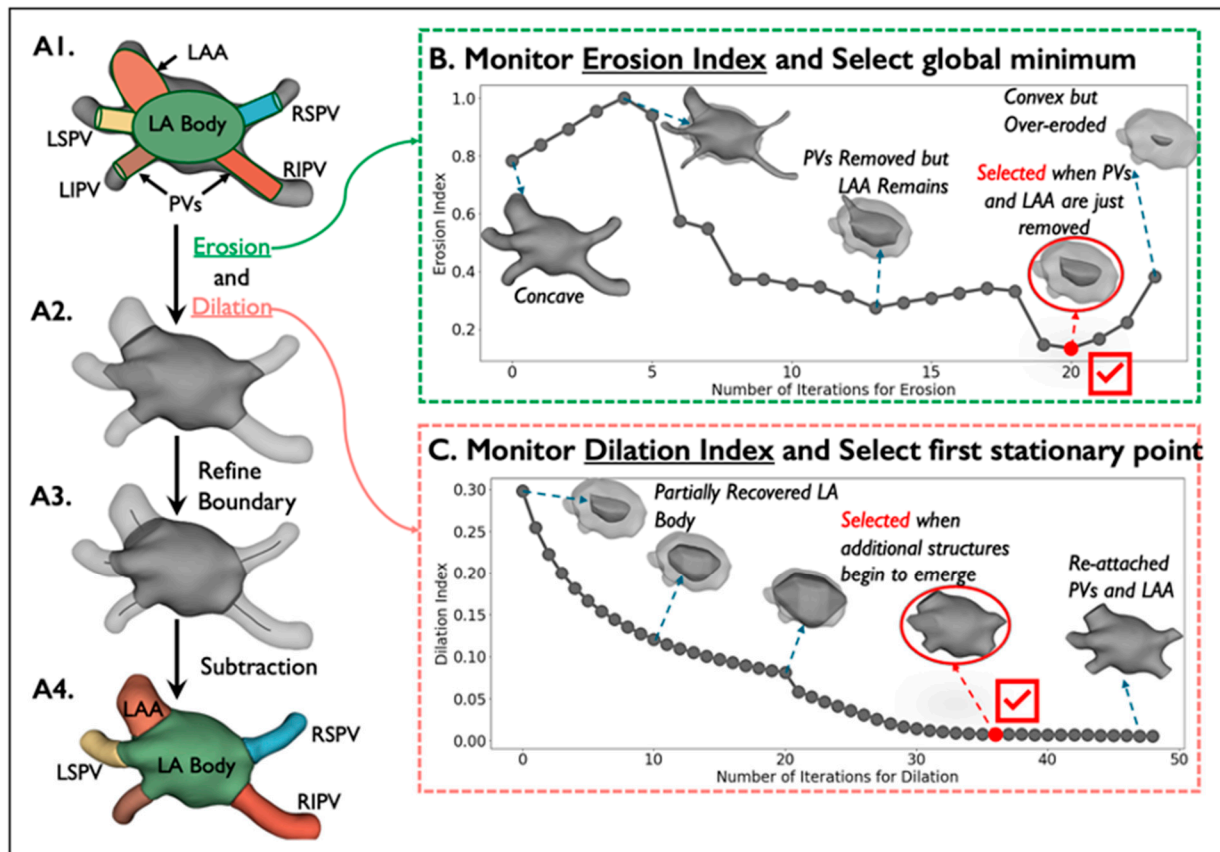
We reasoned that heart structures can be geometrically parsed by separating the convex LA body from the concave whole heart. Three-dimensional voxel erosion, dilation [17], and subtraction were used for this purpose.

To segment PV and LAA from the digital heart, a binary erosion operation was used, which can be defined as  $A \ominus B = \{x \in E^N \mid x + b \in A \text{ for every } b \in B\}$ . Then, in order to recover the original dimension of LA, binary dilation was applied, defined as  $A \oplus B = \{x \in E^N \mid c = a + b \text{ for some } a \in A \text{ and } b \in B\}$ , where  $A$  and  $B$  are sets in  $N$ -space ( $E^N$ ) with elements  $a$  and  $b$ . In our case,  $A$  is the heart model and  $B$  is a structuring element, which is a  $3 \times 3 \times 3$  cube where the center and its 6 neighbors are set to 1 and the remaining elements are 0s.

First, the digital shells were segmented by the application of erosion to concave junctions between PVs and LAA with the LA (Figure 2(A1)). The PVs and LAA are smaller and consist of more 1-connected voxels than the LA body and thus erode more rapidly. However, it is non-trivial to iteratively erode just the PVs and LAA to leave the residual convex LA. To do so, an *Erosion Index* was proposed to monitor the progression of erosion:

$$\text{Erosion Index} = \frac{V(\text{Convex}(E_i)) - V(E_i)}{V(E_i)},$$

where  $E_i$  is the 3D model after the  $i$ th erosion,  $Convex(\cdot)$  is the convex hull, and  $V(\cdot)$  returns the volume of a 3D shape. The erosion index approaches 0 as the shape becomes convex. The index data are preprocessed with a Savitzky–Golay filter and fitted with a polynomial function. The global minimum of the fitting function is calculated to determine the number of iterations for erosion (Figure 2B).



**Figure 2.** Segmentation of digital LA models by erosion–dilation: Detailed description of the first step of the DOKEN algorithm. (A1–A4) The pipeline of our DOKEN algorithm. (B,C) Iterative variations of the erosion and dilation indices along with variations in LA model corresponding to iterations. The ED parameters from this step are then learned by an SVM for labeling clinical data.

Because erosion may remove outer layers of the LA, a dilation operation was applied to recover its original dimension (Figure 2(A2)) by paving voxels on the contour and stopping just before the PVs and LAA are re-attached (Figure 2(A3)), which is monitored by the proposed *Dilation Index* by measuring the number of added voxels after each dilation:

$$Dilation\ Index = \frac{V(D_{i+1}) - V(D_i)}{V(D_i)},$$

where  $D_i$  is the 3D shell after the  $i$ th dilation and  $V(\cdot)$  returns the volume of a 3D shape. Similarly, we processed the index data using a Savitzky–Golay filter then fitted them with a polynomial function. The first stationary point of the fitting function determines the number of dilation iterations (Figure 2C).

After the left atrium body is isolated after erosion and dilation, the boundaries between the LA body and the PVs and LAA were refined by calculating centerlines from the LA centroid to the centroid of each segmented structure. This approach has been used to extract and segment the aorta and great vessels [6,18,19]. Below is a step-by-step algorithm of boundary refinement and centerline calculation:



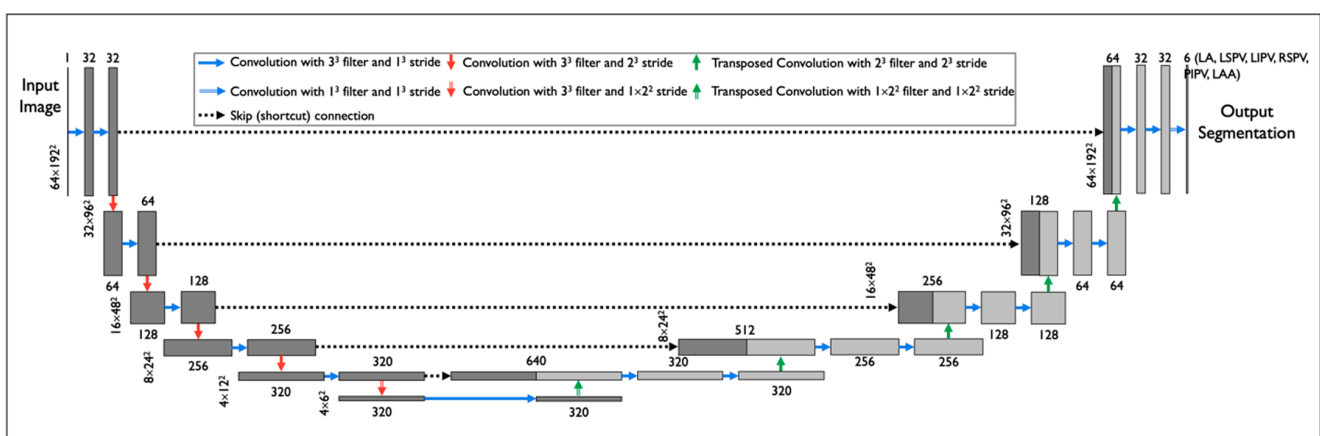
1. Extrapolate a Voronoi diagram [20] from the shell (Figure 2(A1)) to all internal points to create a maximal sphere centered at that point.
  2. Calculate the centroid of the LA body and the centroid of each virtually dissected substructure (4 PVs and LAA).
  3. For each substructure centroid, create a centerline automatically by minimizing the integral of the radius of maximal inscribed spheres along the path that connects the substructure centroid to the LA body centroid.
  4. Replace the boundary between the left atrium and each substructure by a plane orthogonal to the corresponding centerline and close to the original boundary generated by the ED process (Figure 2(A4)).
- II. Tuning the ED parameter using patient LA models

The parameters for the ED process, i.e., the optimal number of iterations, that are suitable for digital models may not apply to clinical data due to heterogeneities such as anatomy variability and imaging artifacts present in the clinical data. The parameters were made suitable for clinical data using a support vector machine (SVM) to predict the parameter value for input clinical CT. Two SVMs (one for each parameter) were trained with manually annotated seed samples ( $N = 5$ ) to predict the optimal number of erosion and dilation iterations. The ED process with parameters predicted by SVMs forms the DOKEN algorithm and will be used to generate robust labels for training the DNN model.

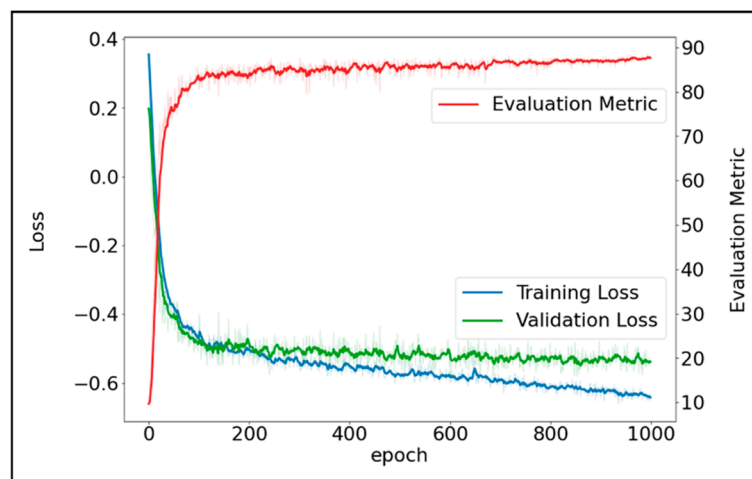
### 2.3. Training the DNN for CT Segmentation from a Small Training Set

DOKEN was applied to  $N = 20$  training data to label the different LA structures in each sample. This was used as ground truth for training the DNN.

We implemented nnU-Net (Figure 3)—a DNN model which has been widely used in 23 public datasets [21]. To train the nnU-Net model, first, each input CT scan was z-score normalized by subtracting its mean, followed by division by its standard deviation. Then the images were re-sampled using third-order spline interpolation. The target voxel spacing was set as the median spacing of the training samples. To improve the generalizability, a set of data augmentation techniques were randomly applied on the fly during training, including rotations, flipping, scaling, Gaussian noise and blur, and random changes in brightness, contrast, and gamma. During the training process, the batch size was set to 2 due to the GPU memory limitation, and the DL model was trained for 1000 epochs. Stochastic gradient descent [22] was used to optimize the model. The initial learning rate and Nesterov momentum were set to 0.01 and 0.99, respectively. The sum of cross-entropy and Dice loss were used as training loss. Figure 4 shows the convergence of training loss, validation loss, and validation accuracy (measured by Dice) during training.



**Figure 3.** DNN model architecture: nn-Unet was applied to segment raw cardiac CT images. The model was trained using a DOKEN-labeled training set (small size) and was tested on a large hold-out test set.



**Figure 4.** Convergence of training loss, validation loss, and the Dice validation accuracy.

#### 2.4. Experimental Setting for Performance Evaluation

The DOKEN algorithm's ability was empirically evaluated to parse cardiac geometry and the DNN model's ability to segment cardiac structures from CT images. The large test set ( $N = 100$ ) was used to manually annotate the ground-truth labels for the 6 substructures by a panel of clinical experts. The manual annotation was performed using a commercially available software tool (EnSite Verismo Segmentation Tool v.2.0.1; Abbott/St Jude Medical, Inc., St. Paul, MN, USA) to manually segment a shell containing the LA body with 4 PVs and the LAA. This whole shell was further parsed ("refined") into its 6 substructures using 3D Slicer [23], manually. The parsing performance of the DOKEN algorithm was measured by centroid–boundary distances against manual annotations. The CT segmentation performance of the DNN model was measured by Dice scores, average surface distance, and centroid–boundary distances, also against manual annotations.

#### 2.5. Performance Evaluation

A newly designed metric, the centroid–boundary distance, was used along with two standard metrics for segmentation tasks [6–8,24–27]—Dice similarity coefficient and average surface distance—to evaluate the model's accuracy in capturing the 2D LA-PV/LAA boundaries, the global 3D structures, and the local 3D shapes and contours, respectively. Mathematically, the centroid–boundary distance is calculated as the average of all the distances from the centroid of the heart to points on the LA-PV/LAA boundary. The Dice similarity score measures spatial overlap between the model prediction and the ground truth, while 0 indicates no overlap and 1 indicates complete overlap, which can be mathematically expressed as

$$\text{Dice Similarity Score} = \frac{2 \times \text{True Positive}}{2 \times \text{True Positive} + \text{False Positive} + \text{False Negative}}.$$

The average surface distance is calculated as the average of all the distances from points on the boundary from model prediction to the ground-truth boundary. The success rate of the DOKEN algorithm was also calculated, where success was defined as an intersect over union (IoU) between the algorithm prediction and expert manual annotation larger than 0.5. This metric has been widely used for detection tasks [28].

#### 2.6. Statistical Analysis

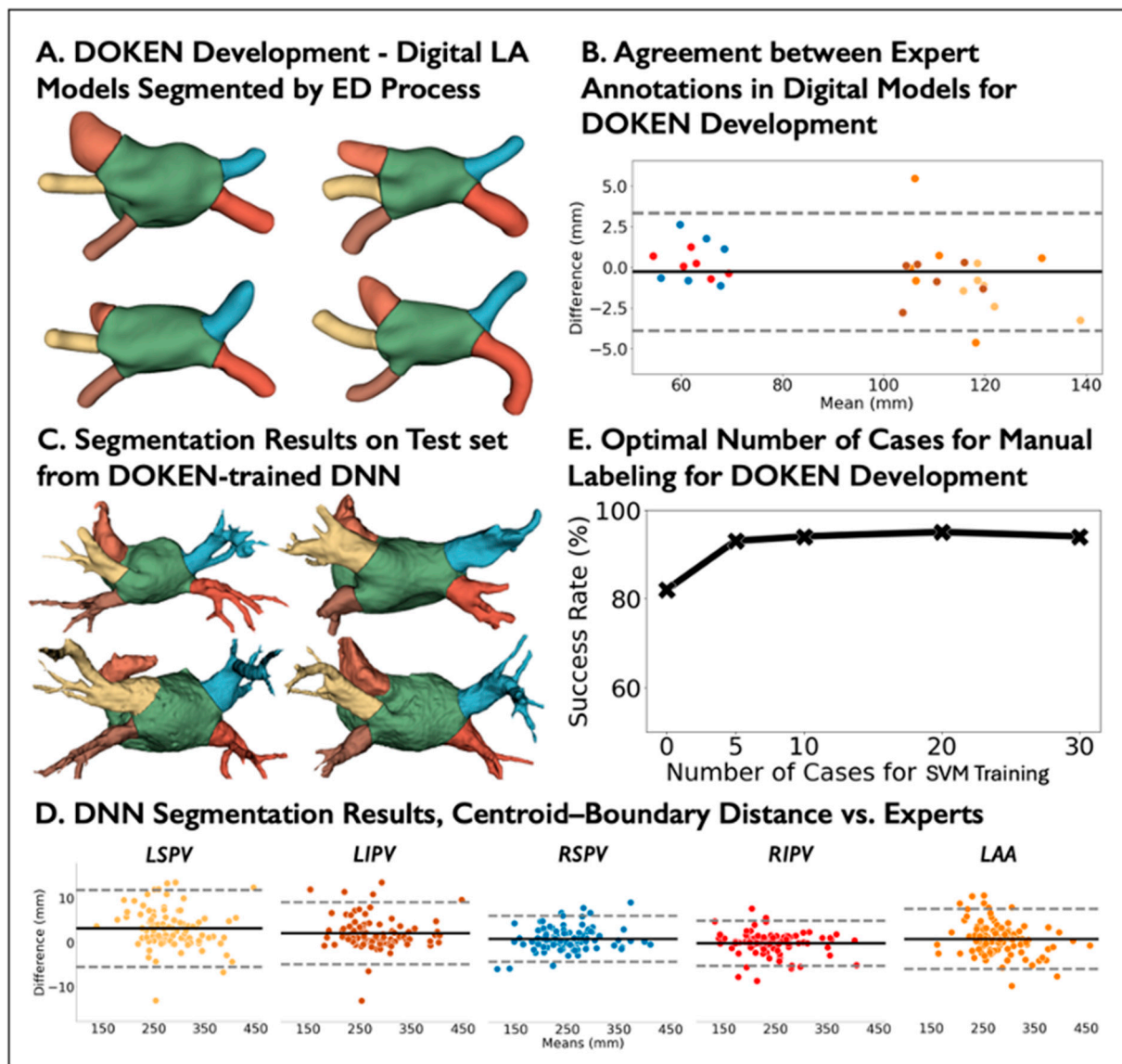
Continuous data are expressed by mean  $\pm$  SD and categorical data by percentages. The distance and Dice scores were summarized as medians and interquartile range (IQR). Pearson correlation's test was used to assess the similarity of LA volumes and the LA sphericity

index estimated from model prediction and ground truth. The Student's *t*-test, Chi-square test, or McNemar's test was applied as appropriate.  $p < 0.05$  was considered significant.

### 3. Results

#### 3.1. DOKEN Algorithm Can Robustly Parse Cardiac Geometry

In digital hearts, the novel DOKEN approach separated the PVs and LAA from the left atrial bodies (Figure 5A) with a mean difference for the centroid–boundary distances of  $-0.27$  mm (95% CI:  $-3.87$  to  $3.33$ ;  $r = 0.99$ ;  $p < 0.0001$ ; Figure 5B). Randomly, five shells of seed data was selected from the  $N = 5$  digital atria for tuning, with LA sizes from 71 to 140 mL that cover a broad range of patients [29].



**Figure 5.** Evaluation of the DOKEN algorithm and the DNN performance for CT image segmentation. (A) Examples of digital LA models segmented by the DOKEN algorithm. (B) Bland–Altman plot of centroid–boundary distance of  $N = 6$  digital LA models segmented by DOKEN compared to experts. (C) Examples of patient LA models segmented by the DOKEN algorithm. (D) Bland–Altman plot of centroid–boundary distance of  $N = 100$  patient LA models in the test set segmented by DOKEN compared to experts. (E) Success rate of DOKEN algorithm with different seed cases for SVM training. Refer to panel D for color codes for the plot in panel B.

In the test set ( $N = 100$ ), the performance of the tuned DOKEN algorithm was compared to expert annotations. Figure 5C presents example results on the test set. The DOKEN method produced a mean difference and limits of agreement for the centroid–boundary distance of 1.46 mm (95% CI:  $-5.58$  to  $8.49$ ;  $r = 0.99$ ;  $p < 0.0001$ ; Figure 5D). The success rate of the algorithm’s parsing when adding more seed data for tuning was assessed. As shown in Figure 5E, the success rate increased from 67% (no tuning) to 94% by tuning with  $N = 5$  shells of seed data ( $p = 0.034$ ; McNemar’s test) and then showed only modest changes (consistency) when tuning in 10–30 shells (92–94%), justifying the selection of seed number.

### 3.2. DNN Trained by DOKEN-Labeled Samples Can Accurately Segment CT

Figure 6 shows comparisons between DNN prediction (left) and manually labeled (right) atria from select samples representing the 25th, 50th, and 75th percentile accuracy in the hold-out set ( $N = 100$ ). The Dice score was 96.7% (IQR: 95.3% to 97.7%, Figure 7A), with a median error in surface distance of boundaries of 1.51 mm (IQR: 0.72 to 3.12, Figure 7B) and a mean centroid–boundary distance of 1.16 mm (95% CI:  $-4.57$  to  $6.89$ , Figure 7C), again similar to expert results ( $r = 0.99$ ;  $p < 0.001$ , Figure 7D).

Thus, this approach enabled a  $>10$ -fold reduction in the relative ratio of training to test cases, inverting the ratio of training:test cases to less than 1:5 from a typical ratio of  $>3$ :1.

### 3.3. Analysis of Anatomical Variants

As previously noted, real CT data have more heterogeneity than digital models, such as variation in patient anatomies. Some anatomies could, in fact, be outliers, i.e., their shape does not follow the typical configuration identified in clinical studies. As no pre-screening was performed to eliminate such anatomy variants, it was analyzed if and how variation in anatomies would affect the method’s performance.

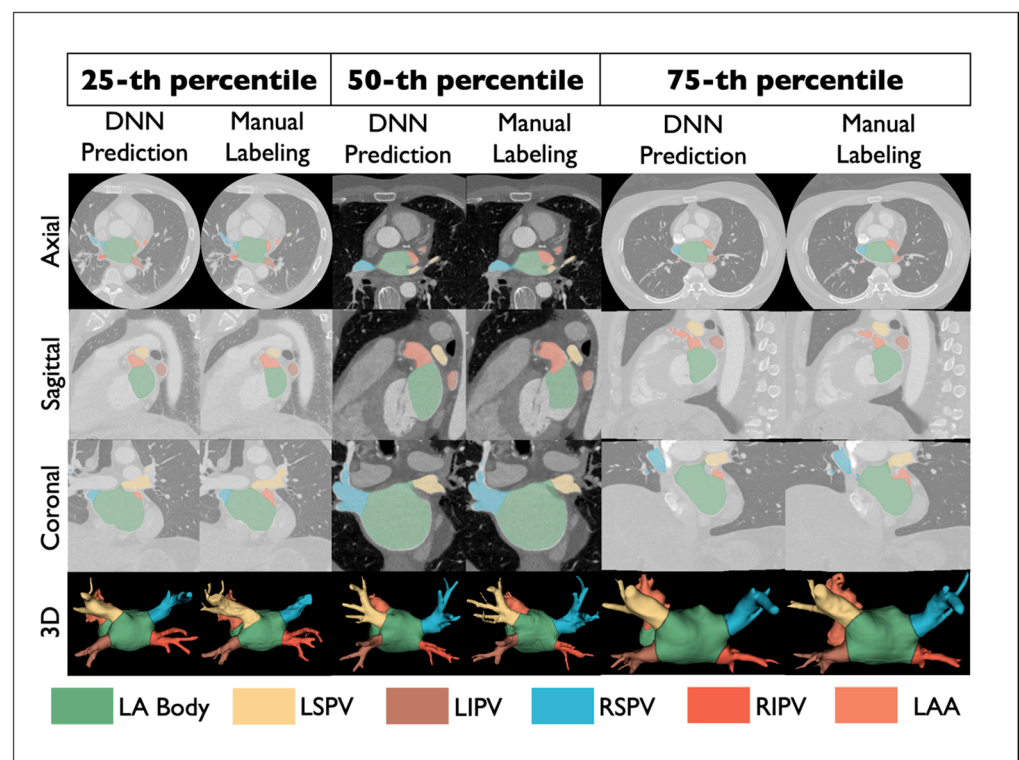
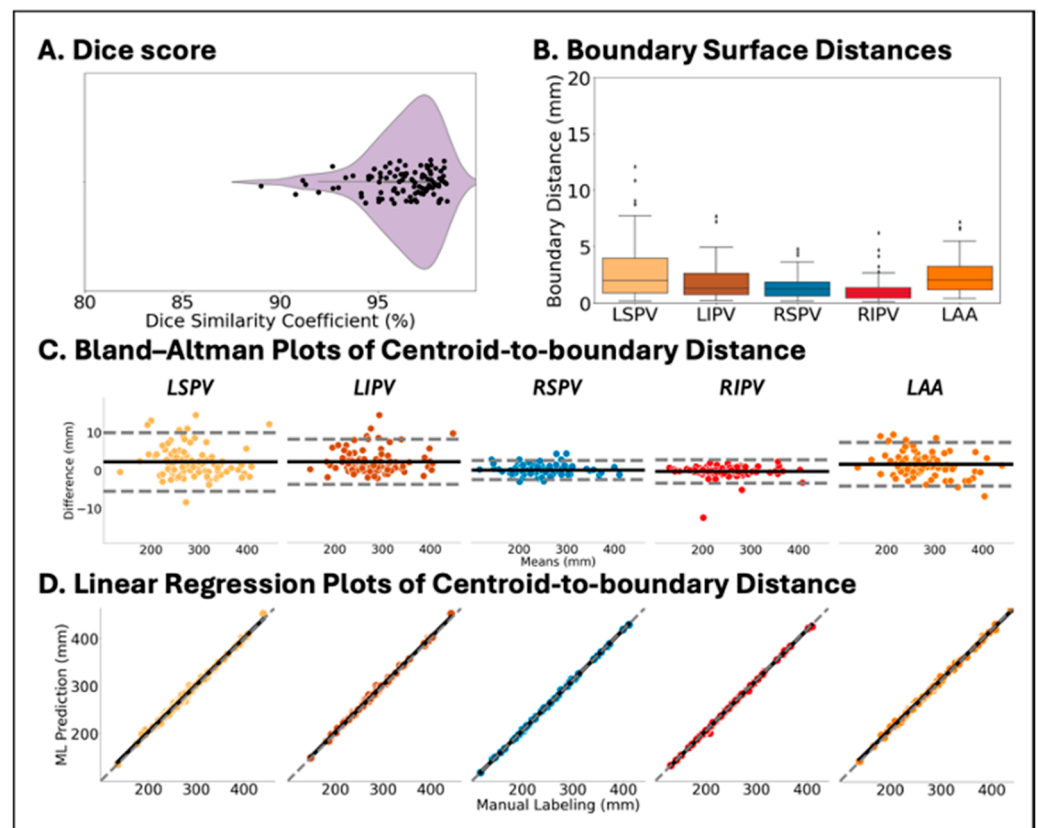


Figure 6. Example results showing DNN segmentation and manual annotation by experts.

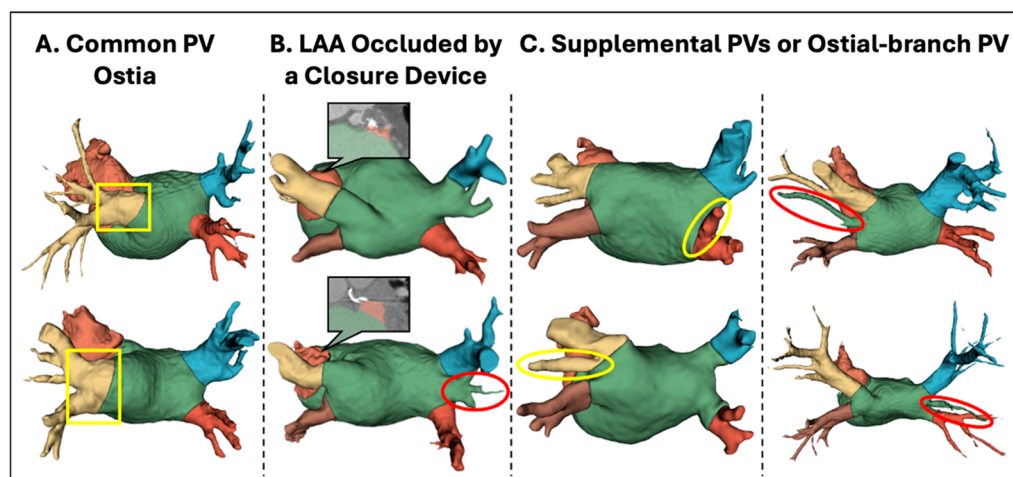




**Figure 7.** Accuracy of CT image segmentation between DNN prediction and expert labeling in the test set ( $N = 100$ ). (A) Violin plot of mean Dice score. (B) Box plot of the surface distance of boundaries of 4 PVs and LAA. (C,D) Bland–Altman and linear regression plots of centroid–boundary distance of 4 PVs and LAA.

Overall, 100% cases with four PV ostia (the most common anatomic configuration, representing 66 cases) were parsed with boundary distances of 1.26 mm (95% CI:  $-5.15$  to  $7.68$ ;  $r = 0.99$ ;  $p < 0.0001$ ). Three main outlier variants were identified (Figure 8): (1) common left PV ostia ( $N = 12$ ), which was successfully parsed despite a lack of specific training on such cases; (2) LAA occlusion by a closure device ( $N = 3$ ), where residual LAA stumps proximal to the occlusion device were correctly identified despite a lack of specific training in such cases; and (3) supplemental PVs or ostial-branch PV, where the DOKEN algorithm was able to segment 19/25 cases.

In summary, 28/34 of identified variants were successfully parsed with anatomic agreement within 1.95 mm (95% CI:  $-6.34$  to  $10.25$ ), which again was in line with expert annotations ( $r = 0.99$ ;  $p < 0.0001$ ), despite lack of specific training for variants. In the remaining six cases, errors arose mostly from missing PVs or branches relative to the four-PV digital model, which could be addressed by geometric models that adapt to a range of PVs.



**Figure 8.** Robust segmentation performance of anatomical variants by the DOKEN algorithm. Three main variants were identified: (A) common left PV ostia ( $N = 12$ ), (B) LAA occlusion by a closure device ( $N = 3$ ), and (C) supplemental PVs or ostial-branch PV ( $N = 25$ ). DOKEN successfully parsed 28/34 of the identified variants (boxed/circled in yellow). However, it missed some extra PVs or branches in the remaining cases (circled in red).

#### 4. Discussion

Domain knowledge encoding of atrial geometry was able to accelerate a DNN for the segmentation of CT images and enable its training on very small datasets. In this study, the training-to-testing ratio was  $<1$  training to 5 test, which indicates a far lower need for training than the conventional published ratios of  $>3:1$  for ML [8,24,26]. This approach was then tested in a hold-out test set, in which the model accelerated segmentation while maintaining similar accuracy to experts. This novel approach could broaden the ease of access and accuracy of AF ablation. More broadly, this approach has analogies to natural intelligence, which has the potential to reduce the need for large, annotated datasets to train ML and could be applied for diverse applications in imaging as well as 3D printing. A simple post-processing step involving a 3D smoothing operation such as a Taubin filter [30] could extend the proposed work for 3D printing applications (illustrated in Supplemental Figure S1).

##### 4.1. DNN Segmentation of Cardiac CT Images

Cardiac CT is increasing used [24,26,31] to guide ablation for AF and to predict clinical endpoints such as the risk of AF recurrence [32,33]. However, the segmentation of these large 70–200 MB datasets manually by experts takes tens of minutes [6–8,24] and 4.4–10 min even with latest commercial software such as the CARTO Segmentation Module version 6 (Biosense Webster, Irvine, CA, USA) [34,35]. The present approach greatly accelerates these reports while retaining high accuracy for routine and variant anatomy while achieving competitive accuracy (93.5–96.7%) with previous work (e.g., 91–97% [25] and 93.4% [24]). This study involved a dataset of  $N = 120$  patients at a single center. The future extension of this work should expand the study cohort with data from multiple institutions, and the labeling should be further refined using a fusion of annotations from multiple experts and addressing discrepancies by an adjudication committee. One such example is demonstrated in our previous work [36], where we used an independent external dataset to test the performance of the algorithm.

The approach also circumvents the limitation that most CT studies that segmented the LA often did not specifically segment the PVs and LAA [24,26]. Similarly, software tools such as SimVascular (v.2023, <https://github.com/SimVascular>, accessed on 14 May 2024) provide automatic segmentation, which uses an ML model (CNN) that was trained using a public dataset MM-WHS [7], which only focuses on labels for the chambers but not

specifically for the complex substructures such as narrow veins (PVs) and the anisotropic-shaped LAA, which are critical for AF ablation. The DOKEN algorithm, on the other hand, offers a scalable solution to segment complex structures in large medical databases. Further, the DOKEN algorithm's goal is focused on segmenting intricate cardiac structures and is not intended to be an alternative for advanced tools like SimVascular, which can perform high-fidelity simulations.

Another limitation is the size of publicly available labeled datasets, which are often small, typically provide test cohorts of <40 cases [6–8], and may create overfitted ML models that generalize poorly [37]. The DOKEN algorithm enabled training from smaller datasets, inverting the typical ratio of training:test cases and reducing the relative size of training to test cases by 10-fold. This “inversed training–test ratio” paradigm has recently been applied in domains outside medicine such as for Amazon co-purchasing product predictions [38]. Other cardiac imaging applications include the segmentation of magnetic resonance imaging (MRI) data to boost ML by reducing the need for large training data sets.

#### 4.2. Challenges in Machine Learning

LeCun et al. and others have stated that difficulties in obtaining large training datasets are among the greatest challenges to machine learning [39]. Obtaining such data is particularly challenging in medicine [40], healthcare [41], and biosciences [42] due to privacy and regulatory requirements. The mathematical encoding of domain knowledge, which emulates some features of natural intelligence, may be a useful approach to address such limitations.

Domain knowledge can be applied in diverse ways. Databases and anatomic atlases have long been used for image segmentation [43,44] but do not encode knowledge principles in a fashion that could be generalized by learning algorithms. Indeed, Trutti et al. [44] pointed out that atlases may identify only a fraction of important structures (7% of 455 subcortical nuclei in the brain), and it is not clear how such “flat” data could be used to identify variants, as we demonstrated. Encoding anatomical knowledge also de-emphasizes low-level details while maintaining high-level abstract information, which may be central to human cognition [12]. The extent of detail required for mathematically encoding is unclear and should be defined for separate applications. Domain knowledge encoding need not be restricted to anatomy and could be applied to processes such as cellular metabolism and physician diagnostic patterns or reports [15].

Alternative approaches are being studied to circumvent large training datasets. Synthetic data may be generated in large quantities to mitigate a lack of actual training data [45], but while they may appear very realistic, they may lack diversity or even introduce bias due to the overfitting [46]. Data augmentation is a widely used approach to training ML on altered versions of the input data to increase the size of the training set [47] but does not capture variations in larger real data [48].

### 5. Conclusions

The novel domain knowledge-encoding algorithm was able to perform the segmentation of six substructures of the LA, reducing the need for large training data sets. The training set had as few as 20 samples, and the hold-out test set included hundreds of patients. The combination of domain knowledge encoding and machine learning approaches could reduce the dependence of ML on large training datasets and could potentially be applied to AF ablation procedures and extended in the future to other imaging, 3D printing, and data science applications.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/diagnostics14141538/s1>, Figure S1: Demonstration of a potential application of our DOKEN algorithm in 3D printing.

**Author Contributions:** Conceptualization, P.G., R.F., B.D., F.V.Y.T., J.Z., M.Z. and S.M.N.; Methodology, R.F., J.Z., M.Z. and S.M.N.; Software, R.F.; Validation, P.C. and T.B.; Formal analysis, R.F. and P.C.; Investigation, B.D., F.V.Y.T., A.J.R., S.R.-C., S.S., P.C., M.R., J.Z., F.H., M.Z. and S.M.N.; Data curation, P.G., B.D., F.V.Y.T. and S.M.N.; Writing—original draft, P.G. and R.F.; Writing—review & editing, P.G., R.F., B.D., F.V.Y.T., A.J.R., S.S., T.B., M.R., J.Z., F.H. and S.M.N.; Supervision, J.Z., M.Z. and S.M.N.; Project administration, S.M.N.; Funding acquisition, S.M.N. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by grants from the National Institutes of Health under award numbers R01 HL149134 and R01 HL83359.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and approved by the review board of Stanford University Human Subjects Protection Committee (NCT02997254, 2016).

**Informed Consent Statement:** Written informed consent was obtained from all participants involved in the study.

**Data Availability Statement:** The code and networks with trained weights will be released upon acceptance. The 3D digital heart models are publicly available at [https://zenodo.org/record/4309958#YdI0JRPMJqs](https://zenodo.org/record/4309958#.YdI0JRPMJqs), accessed on 14 May 2024. The CT dataset used in the study is not currently permitted for public release due to the sensitive nature of patient data.

**Acknowledgments:** Narayan reports grant support from the National Institutes of Health (R01 HL149134 and R01 HL83359) and from the Laurie C. McGrath Foundation, consulting from Uptodate Inc., and TDK Inc., intellectual property owned by University of California Regents and Stanford University. Tjong: Consulting honoraria to institution from Abbott, Boston Scientific, Daiichi Sankyo; no personal gain. Rogers: grants from NIH (F32HL144101), NIH LRP, and Stanford SSPS. Clopton: consulting at the American College of Cardiology. Rodrigo: equity interests in Corify Health Care. Ganesan reports intellectual property owned by Florida Atlantic University and NIH. All other authors have reported that they have no relationships relevant to the contents of this paper to disclose.

**Conflicts of Interest:** SN reports grant support from the National Institutes of Health (R01 HL149134 and R01 HL83359), consulting from Uptodate Inc., and TDK Inc., intellectual property owned by University of California Regents and Stanford University. FT: Consulting honoraria to institution from Abbott, Boston Scientific, Daiichi Sankyo; no personal gain. AR: grants from NIH (F32HL144101), NIH LRP, and Stanford SSPS. PC: consulting at American College of Cardiology. MR: equity interests in Corify Health Care. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

1. Boykov, Y.; Funka-Lea, G. Graph cuts and efficient ND image segmentation. *Int. J. Comput. Vis.* **2006**, *70*, 109–131. [[CrossRef](#)]
2. Preetha, M.M.S.J.; Suresh, L.P.; Bosco, M.J. Image segmentation using seeded region growing. In Proceedings of the 2012 International Conference on Computing, Electronics and Electrical Technologies (ICCEET), Nagercoil, India, 21–22 March 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 576–583.
3. Rodriguez, M.D.; Shah, M. Detecting and segmenting humans in crowded scenes. In Proceedings of the 15th ACM international conference on Multimedia, Augsburg, Germany, 24–29 September 2007; pp. 353–356.
4. Rajaraman, S.; Siegelman, J.; Alderson, P.O.; Folio, L.S.; Folio, L.R.; Antani, S.K. Iteratively pruned deep learning ensembles for COVID-19 detection in chest X-rays. *IEEE Access* **2020**, *8*, 115041–115050. [[CrossRef](#)] [[PubMed](#)]
5. Ganesan, P.; Rajaraman, S.; Long, R.; Ghoraani, B.; Antani, S. Assessment of data augmentation strategies toward performance improvement of abnormality classification in chest radiographs. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; IEEE: Piscataway, NJ, USA; pp. 841–844.
6. Tobon-Gomez, C.; Geers, A.J.; Peters, J.; Weese, J.; Pinto, K.; Karim, R.; Ammar, M.; Daoudi, A.; Margeta, J.; Sandoval, Z.; et al. Benchmark for algorithms segmenting the left atrium from 3D CT and MRI datasets. *IEEE Trans. Med. Imaging* **2015**, *34*, 1460–1473. [[CrossRef](#)] [[PubMed](#)]
7. Zhuang, X.; Li, L.; Payer, C.; Stern, D.; Urschler, M.; Heinrich, M.P.; Oster, J.; Wang, C.; Smedby, Ö.; Bian, C.; et al. Evaluation of algorithms for multi-modality whole heart segmentation: An open-access grand challenge. *Med. Image Anal.* **2019**, *58*, 101537. [[CrossRef](#)] [[PubMed](#)]



8. Xu, X.; Wang, T.; Zhuang, J.; Yuan, H.; Huang, M.; Cen, J.; Jia, Q.; Dong, Y.; Shi, Y. Imagechd: A 3d computed tomography image dataset for classification of congenital heart disease. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Lima, Peru, 4–8 October 2020; Springer: Berlin/Heidelberg, Germany; pp. 77–87.
9. Gianfrancesco, M.A.; Tamang, S.; Yazdany, J.; Schmajuk, G. Potential biases in machine learning algorithms using electronic health record data. *JAMA Intern. Med.* **2018**, *178*, 1544–1547. [[CrossRef](#)] [[PubMed](#)]
10. Futoma, J.; Simons, M.; Panch, T.; Doshi-Velez, F.; Celi, L.A. The myth of generalisability in clinical research and machine learning in health care. *Lancet Digit. Health* **2020**, *2*, e489–e492. [[CrossRef](#)] [[PubMed](#)]
11. Markman, E.M. *Categorization and Naming in Children: Problems of Induction*; MIT Press: Cambridge, MA, USA, 1989.
12. Van Gerven, M. Computational foundations of natural intelligence. *Front. Comput. Neurosci.* **2017**, *11*, 112. [[CrossRef](#)]
13. Lake, B.M.; Salakhutdinov, R.; Tenenbaum, J.B. Human-level concept learning through probabilistic program induction. *Science* **2015**, *350*, 1332–1338. [[CrossRef](#)] [[PubMed](#)]
14. Liu, L.; Wolterink, J.M.; Brune, C.; Veldhuis, R.N. Anatomy-aided deep learning for medical image segmentation: A review. *Phys. Med. Biol.* **2021**, *66*, 11TR01. [[CrossRef](#)]
15. Xie, X.; Niu, J.; Liu, X.; Chen, Z.; Tang, S.; Yu, S. A survey on incorporating domain knowledge into deep learning for medical image analysis. *Med. Image Anal.* **2021**, *69*, 101985. [[CrossRef](#)]
16. Nagel, C.; Schuler, S.; Dössel, O.; Loewe, A. A bi-atrial statistical shape model for large-scale in silico studies of human atria: Model development and application to ECG simulations. *Med. Image Anal.* **2021**, *74*, 102210. [[CrossRef](#)] [[PubMed](#)]
17. Soille, P. *Erosion and Dilation, Morphological Image Analysis*; Springer: Berlin, Germany, 2004.
18. Krissian, K.; Carreira, J.M.; Esclarin, J.; Maynar, M. Semi-automatic segmentation and detection of aorta dissection wall in MDCT angiography. *Med. Image Anal.* **2014**, *18*, 83–102. [[CrossRef](#)] [[PubMed](#)]
19. Razeghi, O.; Sim, I.; Roney, C.H.; Karim, R.; Chubb, H.; Whitaker, J.; O’Neill, L.; Mukherjee, R.; Wright, M.; O’neill, M.; et al. Fully automatic atrial fibrosis assessment using a multilabel convolutional neural network. *Circ. Cardiovasc. Imaging* **2020**, *13*, e011512. [[CrossRef](#)] [[PubMed](#)]
20. Piccinelli, M.; Veneziani, A.; Steinman, D.A.; Remuzzi, A.; Antiga, L. A framework for geometric analysis of vascular structures: Application to cerebral aneurysms. *IEEE Trans. Med. Imaging* **2009**, *28*, 1141–1155. [[CrossRef](#)] [[PubMed](#)]
21. Isensee, F.; Jaeger, P.F.; Kohl, S.A.; Petersen, J.; Maier-Hein, K.H. nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **2021**, *18*, 203–211. [[CrossRef](#)] [[PubMed](#)]
22. Ketkar, N. Stochastic gradient descent. In *Deep Learning with Python*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 113–132.
23. Fedorov, A.; Beichel, R.; Kalpathy-Cramer, J.; Finet, J.; Fillion-Robin, J.C.; Pujol, S.; Bauer, C.; Jennings, D.; Fennessy, F.; Sonka, M.; et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn. Reson. Imaging* **2012**, *30*, 1323–1341. [[CrossRef](#)] [[PubMed](#)]
24. Baskaran, L.; Maliakal, G.; Al’Aref, S.J.; Singh, G.; Xu, Z.; Michalak, K.; Dolan, K.; Gianni, U.; van Rosendael, A.; van den Hoogen, I.; et al. Identification and quantification of cardiovascular structures from CCTA: An end-to-end, rapid, pixel-wise, deep-learning method. *Cardiovasc. Imaging* **2020**, *13*, 1163–1171.
25. Xu, H.; Niederer, S.A.; Williams, S.E.; Newby, D.E.; Williams, M.C.; Young, A.A. Whole Heart Anatomical Refinement from CCTA Using Extrapolation and Parcellation. In Proceedings of the International Conference on Functional Imaging and Modeling of the Heart, Stanford, CA, USA, 21–25 June 2021; Springer: Berlin/Heidelberg, Germany, 2021; pp. 63–70.
26. Chen, H.H.; Liu, C.M.; Chang, S.L.; Chang, P.Y.C.; Chen, W.S.; Pan, Y.M.; Fang, S.T.; Zhan, S.Q.; Chuang, C.M.; Lin, Y.J.; et al. Automated extraction of left atrial volumes from two-dimensional computer tomography images using a deep learning technique. *Int. J. Cardiol.* **2020**, *316*, 272–278. [[CrossRef](#)] [[PubMed](#)]
27. Xie, W.; Yao, Z.; Ji, E.; Qiu, H.; Chen, Z.; Guo, H.; Zhuang, J.; Jia, Q.; Huang, M. Artificial intelligence-based computed tomography processing framework for surgical telementoring of congenital heart disease. *ACM J. Emerg. Technol. Comput. Syst. (JETC)* **2021**, *17*, 1–24. [[CrossRef](#)]
28. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; Springer: Berlin/Heidelberg, Germany, 2014; pp. 740–755.
29. Sangsriwong, M.; Cismaru, G.; Puiu, M.; Simu, G.; Istratoaie, S.; Muresan, L.; Gusetu, G.; Cismaru, A.; Pop, D.; Zdregnea, D.; et al. Formula to estimate left atrial volume using antero-posterior diameter in patients with catheter ablation of atrial fibrillation. *Medicine* **2021**, *100*, e26513. [[CrossRef](#)]
30. Taubin, G. Curve and surface smoothing without shrinkage. In Proceedings of the IEEE International Conference on Computer Vision, Cambridge, MA, USA, 20–23 June 1995; IEEE: Piscataway, NJ, USA, 1995; pp. 852–857.
31. Chen, C.; Qin, C.; Qiu, H.; Tarroni, G.; Duan, J.; Bai, W.; Rueckert, D. Deep learning for cardiac image segmentation: A review. *Front. Cardiovasc. Med.* **2020**, *7*, 25. [[CrossRef](#)] [[PubMed](#)]
32. Nakamori, S.; Ngo, L.H.; Tugal, D.; Manning, W.J.; Nezafat, R. Incremental value of left atrial geometric remodeling in predicting late atrial fibrillation recurrence after pulmonary vein isolation: A cardiovascular magnetic resonance study. *J. Am. Heart Assoc.* **2018**, *7*, e009793. [[CrossRef](#)] [[PubMed](#)]

33. Firouznia, M.; Feeny, A.K.; LaBarbera, M.A.; McHale, M.; Cantlay, C.; Kalfas, N.; Schoenhagen, P.; Saliba, W.; Tchou, P.; Barnard, J.; et al. Machine learning–derived fractal features of shape and texture of the left atrium and pulmonary veins from cardiac computed tomography scans are associated with risk of recurrence of atrial fibrillation postablation. *Circ. Arrhythmia Electrophysiol.* **2021**, *14*, e009265. [[CrossRef](#)] [[PubMed](#)]
34. Tovia-Brodie, O.; Belhassen, B.; Glick, A.; Shmilovich, H.; Aviram, G.; Rosso, R.; Michowitz, Y. Use of new imaging CARTO® segmentation module software to facilitate ablation of ventricular arrhythmias. *J. Cardiovasc. Electrophysiol.* **2017**, *28*, 240–248. [[CrossRef](#)]
35. Tops, L.F.; Bax, J.J.; Zeppenfeld, K.; Jongbloed, M.R.; Lamb, H.J.; van der Wall, E.E.; Schalij, M.J. Fusion of multislice computed tomography imaging with three-dimensional electroanatomic mapping to guide radiofrequency catheter ablation procedures. *Heart Rhythm.* **2005**, *2*, 1076–1081. [[CrossRef](#)]
36. Feng, R.; Deb, B.; Ganesan, P.; Tjong, F.V.; Rogers, A.J.; Ruipérez-Campillo, S.; Somani, S.; Clopton, P.; Baykaner, T.; Rodrigo, M.; et al. Segmenting computed tomograms for cardiac ablation using machine learning leveraged by domain knowledge encoding. *Front. Cardiovasc. Med.* **2023**, *10*, 1189293. [[CrossRef](#)]
37. Ng, A.Y. Preventing “overfitting” of cross-validation data. In Proceedings of the Fourteenth International Conference on Machine Learning (ICML), Nashville, TN, USA, 8–12 July 1997; Volume 97, pp. 245–253.
38. Hu, W.; Fey, M.; Zitnik, M.; Dong, Y.; Ren, H.; Liu, B.; Catasta, M.; Leskovec, J. Open graph benchmark: Datasets for machine learning on graphs. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 22118–22133.
39. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)]
40. Topol, E.J. High-performance medicine: The convergence of human and artificial intelligence. *Nat. Med.* **2019**, *25*, 44–56. [[CrossRef](#)]
41. Esteva, A.; Robicquet, A.; Ramsundar, B.; Kuleshov, V.; DePristo, M.; Chou, K.; Cui, C.; Corrado, G.; Thrun, S.; Dean, J. A guide to deep learning in healthcare. *Nat. Med.* **2019**, *25*, 24–29. [[CrossRef](#)]
42. Sapoval, N.; Aghazadeh, A.; Nute, M.G.; Antunes, D.A.; Balaji, A.; Baraniuk, R.; Barberan, C.J.; Dannenfelser, R.; Dun, C.; Edrisi, M.; et al. Current progress and open challenges for applying deep learning across the biosciences. *Nat. Commun.* **2022**, *13*, 1728. [[CrossRef](#)] [[PubMed](#)]
43. Vakalopoulou, M.; Chassagnon, G.; Bus, N.; Marini, R.; Zacharaki, E.I.; Revel, M.P.; Paragios, N. Atlasnet: Multi-atlas non-linear deep networks for medical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Granada, Spain, 16–20 September 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 658–666.
44. Trutti, A.C.; Fontanesi, L.; Mulder, M.J.; Bazin, P.-L.; Hommel, B.; Forstmann, B.U. A probabilistic atlas of the human ventral tegmental area (VTA) based on 7 Tesla MRI data. *Brain Struct. Funct.* **2021**, *226*, 1155–1167. [[CrossRef](#)] [[PubMed](#)]
45. Chen, R.J.; Lu, M.Y.; Chen, T.Y.; Williamson, D.F.; Mahmood, F. Synthetic data in machine learning for medicine and healthcare. *Nat. Biomed. Eng.* **2021**, *5*, 493–497. [[CrossRef](#)] [[PubMed](#)]
46. Bhanot, K.; Qi, M.; Erickson, J.S.; Guyon, I.; Bennett, K.P. The problem of fairness in synthetic healthcare data. *Entropy* **2021**, *23*, 1165. [[CrossRef](#)]
47. Shorten, C.; Khoshgoftaar, T.M. A survey on image data augmentation for deep learning. *J. Big Data* **2019**, *6*, 60. [[CrossRef](#)]
48. Shen, J.; Dudley, J.; Kristensson, P.O. The imaginative generative adversarial network: Automatic data augmentation for dynamic skeleton-based hand gesture and human action recognition. In Proceedings of the 2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021), Jodhpur, India, 15–18 December 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1–8.

**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.