




Article

Automatic Identification of Bioprostheses on X-ray Angiographic Sequences of Transcatheter Aortic Valve Implantation Procedures Using Deep Learning

Laura Busto ^{1,*} , César Veiga ^{1,*}, José A. González-Nóvoa ¹ , Marcos Loureiro-Ga ¹ , Víctor Jiménez ², José Antonio Baz ² and Andrés Íñiguez ²

¹ Cardiovascular Research Group, Galicia Sur Health Research Institute (IIS Galicia Sur), 36213 Vigo, Spain; jose.gonzalez@iisgaliciasur.es (J.A.G.-N.); marcos.loureiro@iisgaliciasur.es (M.L.-G.)

² Cardiology Department, Complejo Hospitalario Universitario de Vigo (SERGAS), Álvaro Cunqueiro Hospital, 36213 Vigo, Spain; victor.alfonso.jimenez.diaz@sergas.es (V.J.); jose.antonio.baz.alonso2@sergas.es (J.A.B.); andres.iniguez.romo@sergas.es (A.Í.)

* Correspondence: laura.busto@iisgaliciasur.es (L.B.); cesar.veiga@iisgaliciasur.es (C.V.)

Abstract: Transcatheter aortic valve implantation (TAVI) has become the treatment of choice for patients with severe aortic stenosis and high surgical risk. Angiography has been established as an essential tool in TAVI, as this modality provides real-time images required to support the intervention. The automatic interpretation and parameter extraction on such images can lead to significant improvements and new applications in the procedure that, in most cases, rely on a prior identification of the transcatheter heart valve (THV). In this paper, U-Net architecture is proposed for the automatic segmentation of THV on angiographies, studying the role of its hyperparameters in the quality of the segmentations. Several experiments have been conducted, testing the methodology using multiple configurations and evaluating the results on different types of frames captured during the procedure. The evaluation has been performed in terms of conventional classification metrics, complemented with two new metrics, specifically defined for this problem. Those new metrics provide a more appropriate assessment of the quality of the results, given the class imbalance in the dataset. From an analysis of the evaluation results, it can be concluded that the method provides appropriate segmentation results for this dataset.

Keywords: angiography; deep learning; segmentation; transcatheter aortic valve implantation (TAVI); U-Net



Citation: Busto, L.; Veiga, C.; González-Nóvoa, J.A.; Loureiro-Ga, M.; Jiménez, V.; Baz, J.A.; Íñiguez, A. Automatic Identification of Bioprostheses on X-ray Angiographic Sequences of Transcatheter Aortic Valve Implantation Procedures Using Deep Learning. *Diagnostics* **2022**, *12*, 334. <https://doi.org/10.3390/diagnostics12020334>

Academic Editors: Cees A. Swenne, Laura Burattini and Agnese Sbröllini

Received: 22 December 2021

Accepted: 21 January 2022

Published: 27 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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

Transcatheter aortic valve implantation (TAVI) is the procedure for the replacement of the aortic valve in the heart through blood vessels, using catheters to deliver and deploy the bioprosthesis. Due to its numerous advantages, the procedure has been established during the last decade as the main treatment for most patients presenting symptomatic severe aortic stenosis [1]. It has been proved to be more beneficial than medical treatment for inoperable cases, becoming a better alternative to surgery on intermediate or high risk patients [2]. There is also a consensus in the imaging modalities and protocols involved in all stages of TAVI [3], including several different modalities at pre, peri, and post-procedural states [4]. The standard practice uses intra-procedural angiography and pre-implant three-dimensional imaging sources to ensure transcatheter heart valve (THV) correct placement, assessing the position and functioning immediately after deployment and identifying any intra-procedural complication. Thus, angiography has become an important supporting tool during TAVI due to its ability to produce real-time views [5].

Even if the imaging role in TAVI is clearly defined in the procedural guides, the possibility of extracting additional value from images gathered during and after the procedure for its evaluation and the prediction of long-term outcomes is still an open challenge [5].

Despite the maturity of the modality, the automatic interpretation of these images would boost new applications for angiography prior to and during TAVI, facilitating the standardization of the procedure and potentially limiting its complications. Such applications could include the assessment of the optimal deployment, which can be obtained by measuring the depth of the left ventricular edge of the device in the left ventricular outflow tract (LVOT) relative to the annular plane of the aortic valve [6]. Another possible utility is the implant characterization by extracting parameters as the stent tip deflection, the maximum bioprosthesis extension, and LVOT eccentricity [7]. All those applications rely on a prior requirement, which is the precise and automatic identification of the THV. Thus, detecting and tracking the bioprosthesis in angiographic X-ray imaging sequences has several new practical applications [8]. This automatic detection points directly to the hot topic of Artificial Intelligence (AI) in medical imaging [9]. AI has been successfully applied to several cardiovascular imaging modalities [10,11]; however, its use in percutaneous cardiovascular interventions has scarcely been reported [12].

This work presents a way to improve TAVI by the automatic detection of the bioprosthesis on X-ray angiographic sequences. U-Net architecture is proposed for this purpose, testing different hyperparameter configurations and studying their impact on the segmentation results. Such network is trained and tested using an annotated dataset generated by the research group. The validation of any methodology plays an important role in this realm [13], and the task becomes specially complex in class imbalance problems [14], as the one presented in this work. New metrics have been defined specifically for this imbalanced segmentation, which provide a more proper assessment of the quality of the results.

The next section introduces the required background in terms of imaging particularities; Deep Learning and, particularly, U-Net fundamentals; and the metrics used to assess the performance of this kind of systems. Section 3 describes the proposed methodology and its implementation. In Section 4, results of the performed experiments are provided, and the last section includes the discussion and conclusions.

2. Background

This section introduces the main parts of the proposed approach: angiographic imaging in TAVI, U-Net architecture, and evaluation metrics suitable to assess the methodology.

2.1. Angiographic Imaging in TAVI

TAVI takes place in a catheterization laboratory, such as the one shown in Figure 1, which includes the C-arm that produces the real-time angiographic sequences during the intervention.



Figure 1. Álvaro Cunqueiro Hospital Cath Lab.

During the procedure, the bioprosthesis is introduced compressed in the catheter, using angiographic views it is guided towards the aortic root, and finally placed on the native valve (Figure 2b). Using angiographic imaging, the optimal plane is identified,

which is crucial for the success of the bioprosthesis positioning. Once the device is placed in the appropriate location, it is deployed (Figure 2c), until it is fully expanded (Figure 2d). It is remarkable the great variety of images gathered in different stages of the procedure, as all the examples in Figure 2. An example of an angiographic sequence is provided as a Supplementary Figure S1 (angioseq.gif).

Nowadays, there are several manufacturers that produce different THV models, with different shapes and sizes. However, all of them share a characteristic radiopaque reticular pattern, which is perceivable in the angiographies, independently of the image generation properties or angulations.

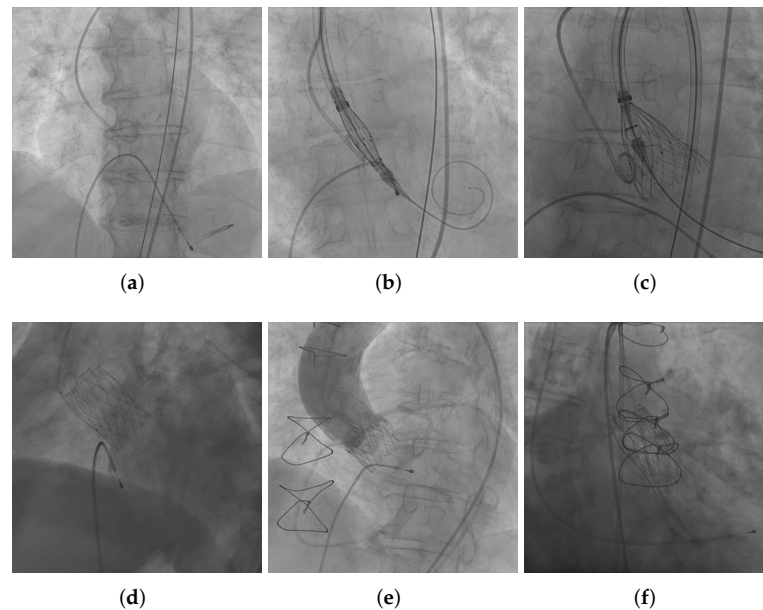


Figure 2. Angiographic frames captured during TAVI. (a) Aortic root previous to the device delivery with a pigtail catheter in the non-coronary sinus, and a temporary pacemaker lead in the right ventricle. (b) THV loaded in the delivery catheter and positioned in the aortic annulus before deployment. (c) First part of the THV deployment in the aortic annulus. (d) THV completely deployed in the aortic annulus. (e) Functioning assessment of a deployed THV using a contrast agent injection. (f) Other elements overlapping the device (previous surgical sutures at the level of the sternum and a temporary pacemaker lead in the right ventricle).

2.2. U-Net

In a Deep Learning application, two important aspects are the network architecture and the training parameters and strategy.

Concerning architectures, U-Net is a convolutional neural network developed for biomedical image segmentation [15]. It relies on the use of data augmentation, which means that the images presented to the network are modified versions of the original ones, generated on the fly by applying elastic transformations. By using data augmentation, the network does not process the same image twice in the different epochs, enlarging the dataset and aiming to teach the network the desired invariance and robustness properties.

With respect to the training, the stage when the network weights are adjusted in each epoch such that a cost function is optimized, several optimizers and cost functions can be employed. One frequently used cost function is cross-entropy, defined as follows for an image binary classification:

$$CE = - \sum_{i=1}^{N \times M} y_i \log(p_i) + (1 - y_i) \log(1 - p_i), \quad (1)$$

where i is an index over all the pixels of an image of size $N \times M$. $y_i \in \{0, 1\}$ corresponds to the pixel classification in the image annotation, background or foreground respectively. p_i is the probability of the pixel i to belong to foreground, predicted by the model. Equation (1) can be written as (2):

$$CE = - \sum_{i=1}^{N \times M} \log(q_i), \quad (2)$$

where q_i is defined as follows:

$$q_i = \begin{cases} p_i & \text{if } y_i = 1 \\ 1 - p_i & \text{otherwise.} \end{cases} \quad (3)$$

In situations of highly imbalanced classes, a well-known alternative to cross-entropy is focal loss [16]. Focal loss expression consists in a modification of cross-entropy with a modulating factor $(1 - q_i)^\gamma$ that decreases as the classification confidence increases, where $\gamma \geq 0$ is a tunable *focussing* parameter. In practice, the following α -balanced variant (4) of the focal loss is used.

$$FL = - \sum_{i=1}^{N \times M} \alpha_i (1 - q_i)^\gamma \log(q_i), \quad (4)$$

where α_i is obtained as follows:

$$\alpha_i = \begin{cases} \alpha & \text{if } y_i = 1 \\ 1 - \alpha & \text{otherwise.} \end{cases} \quad (5)$$

In short, these α and γ parameters repercut in the contribution of each sample to the loss. $\alpha \in [0, 1]$ plays a role in the balance of the classes (background and foreground in binary image segmentation), whereas γ affects the weights of correct and wrong classifications.

2.3. Evaluation Metrics

The evaluation of the results plays a crucial role in general image segmentation problems, and in medical imaging in particular [13]. Most frequent metrics in binary classification consist in a pixel-wise evaluation that requires a gold standard where an expert determines the pixels on each image corresponding to the target to detect, often by manually drawing it. Comparing the gold standard with the results permits quantitative evaluation and statistics. Some widely used classification metrics are accuracy, precision, and recall [17]. Other common methods are the Receiver Operating Characteristic (ROC) [18] curve analysis, particularly the area under the ROC curve (AUROC), and metrics measuring overlapping areas between finite sample sets, such as Dice coefficient [19] and Jaccard index [20].

A different evaluation approach is measuring the distance between sets of points. One of those is Hausdorff distance [21], which is the greatest of all the distances from a point in one set to the closest point in the other set, defined as:

$$HD(\mathbf{G}, \mathbf{S}) = \max_{g \in \mathbf{G}} \left(\min_{s \in \mathbf{S}} (d(g, s)) \right), \quad (6)$$

where \mathbf{G} and \mathbf{S} are gold standard and segmentation result respectively, both binary images, and G and S are the sets of non-zero elements in these images. g and s iterate over the points of such sets and $d(\cdot)$ denotes the Euclidean distance between two points.

3. Materials and Methods

The methodology followed in this work distinguishes four main blocks. First block is image dataset acquisition and annotation. The second block is the hyperparameter tuning of a U-Net architecture, which allows to select the network configuration. In third place,

the training and test of the model using such configuration. Finally, the assessment of the segmentation results obtained by such model. These evaluation results are used by the hyperparameter tuning block for the selection of the configuration that provides the best segmentation results for this dataset among the considered options.

The whole work has been implemented in Python, using Keras [22] and TensorFlow [23] libraries. The training of this kind of networks take advantage of high performance computational infrastructures. In this case, the computations of this work were performed in a GPU cluster owned by the research group, which has three NVIDIA TESLA A100 GPUs.

3.1. Image Acquisition and Annotation

The images used in this work belong to a total of 15 patients suffering from severe aortic stenosis. All of the patients have been implanted an Allegra (New Valve Technology AG, Muri, Switzerland) THV of different sizes: 23 mm, 27 mm, or 31 mm. This prosthesis is a self-expandable valve made of a nitinol stent frame and bovine pericardium leaflets with a supra-annular functioning. These cases are consecutive procedures with such implant that took place from June 2019 to January 2020 at the Department of Cardiology at Álvaro Cunqueiro Hospital, Vigo, Spain. The trial protocol was approved by the Spanish national authorities and ethics committees and by institutional research boards at the participating site. Informed consent was obtained and documented for all patients before conducting any study-related procedures.

The TAVI procedures were performed according to the local protocol at the participating site. TAVI images were obtained by conventional fluoroscopy and cine at 15 and 25 fps on AlluraClarity equipment with ClarityIQ technology (Philips, Amsterdam). During the procedure, unfractionated heparin was used with the goal of an activated clotting time >250 s and the decision to reverse its action with protamine at the end of the procedure were done according to local standard.

Several sequences have been captured from each patient and from different projections. All the sequences contain between 25 and 100 frames, and it is guaranteed that all of them contain at least a complete heartbeat. Each frame is of 512×512 pixels, with a pixel resolution of 8 bits and a pixel spacing of 0.34 mm.

Each of these images has been annotated, generating a gold standard. This is a set of labels of the frames containing the THV structure, an example can be seen in Figure 3a. These labels were manually drawn by the clinical members of the research team, namely, experienced hemodynamic cardiologists, using GNU Image Manipulation Program (GIMP) software. A particular case in the generation of the gold standard occurs in contrast injection frames, where the structure is only visible in part of the frame, as the threads cannot be distinguished where the contrast is injected. The gold standard of these frames considers as foreground only the part where the THV is visible, excluding the contrast pixels, yielding a partial pattern, as shown in Figure 3b.

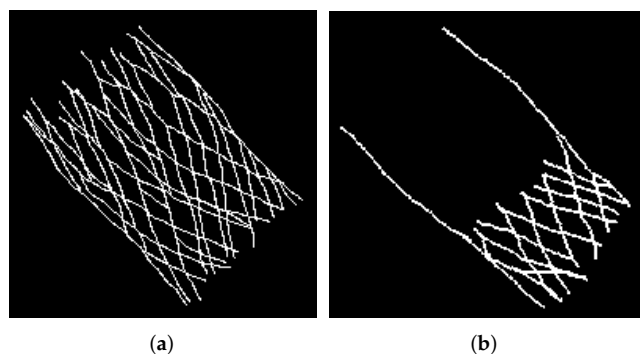


Figure 3. Cropped versions, only the region where the THV is contained, of the gold standard of different frame types. (a) Fully expanded THV. (b) Contrast injection.

3.2. Hyperparameter Tuning

As there are several hyperparameters and operation conditions of the architecture that play an important role in the quality of the segmentation results, a hyperparameter tuning stage is included in the methodology to study their repercussion and select an adequate configuration for this TAVI dataset. These hyperparameters considered for tuning the network are training set size, the use or not of data augmentation, and the cost function used in the model training.

Concerning the cost functions, two different options have been contemplated, namely, binary cross-entropy (2) and focal loss (4). For the selection of the focal loss parameters, the methodology proposed in the original focal loss work [16] in the ablation experiments for RetinaNet detector has been replicated, now using TAVI dataset and a U-Net. The best THV segmentation results have been obtained with $\alpha = 0.25$ and $\gamma = 2$. On the other hand, the tuning stage considered data augmentation during training, applying transformations such as translation, zoom, rotation, and horizontal flip when it is enabled.

Different combinations of such hyperparameters have been tested, as will be detailed in following experiments, allowing to determine a proper configuration for the present problem and dataset by taking into account their evaluation metrics results.

3.3. THV Segmentation

The U-Net implementation used in this work for the segmentation of the THV has 64 filters in the first convolutional layer, as in the original paper [15], and it has been trained employing Adam optimizer [24].

The expected inputs are 512×512 , 8-bit (integer values from 0 to 255), grayscale images and their corresponding labels as binary masks for the training. The outputs are 512×512 matrices representing the predicted probability of each pixel to belong to the foreground. Therefore, all the elements in the output images are within the interval $[0, 1]$, but generally not binary. It is important to remark that these results do not label the frames in classes, they provide probability maps of this classification, and thus are called predictions. Binarized versions of such predictions would be called segmentations.

Figure 4 provides examples of segmentation results, as binarized versions obtained using Otsu's method [25], overlaid on the original frames. The whole sequence from which the frame in Figure 4c was extracted is provided as a Supplementary Figure S2 (seg.gif), with the segmentation results overlaid.

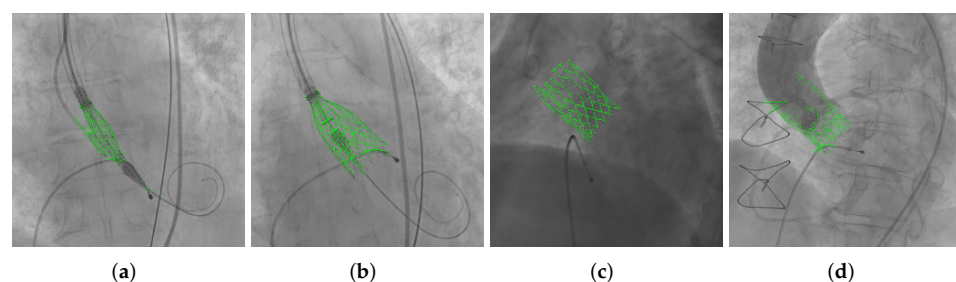


Figure 4. Different types of frames with their segmentation results (overlaid in green) obtained using Otsu's method. (a) THV loaded in the delivery catheter, previous to the deployment. (b) THV deployment. (c) THV totally deployed and expanded. (d) THV with contrast agent injection.

3.4. Evaluation of the Segmentation Results

In the evaluation block, the quality of the results is assessed with different metrics by comparing each frame with the corresponding gold standard.

As the foreground in the gold standard consists of less than 1% of the image pixels, THV identification is a class imbalance problem. Therefore, conventional classification metrics used in supervised learning, based on the number of true/false positives and negatives, are not fully suitable on this problem [26]. Aiming to find a better way to assess the results, new metrics are required and will be defined. The first proposed metric, denoted

as D_1 , measures the average minimum Euclidean distance (in pixels) between the non-null elements in the gold standard and the closest non-null elements in the segmentations. It can be computed as follows:

$$D_1(\mathbf{G}, \mathbf{S}) = \frac{1}{K} \sum_{g \in G} \min_{s \in S} d(g, s), \quad (7)$$

where G and S are the sets of non-zero elements in the gold standard (\mathbf{G}) and a segmentation (\mathbf{S}), respectively. g and s iterate over the points of such sets and $d(\cdot)$ denotes the Euclidean distance, measured in pixels, between two points. $\frac{1}{K}$ is a normalization term, being K the size of G , and this is the number of foreground pixels in the gold standard. In case of optimal segmentation, D_1 would be zero. However, notice that in an extreme case where the segmentation is an all-ones matrix, the distance would still be zero although the segmentation would be incorrect. In order to penalize this effect, a second metric D_2 is defined. D_2 , expressed as (8), measures the noise by aggregating the probability in all the prediction pixels not present in the ground truth.

$$D_2(\mathbf{G}, \mathbf{P}) = \sum_{g' \in G^c} \mathbf{P}_{g'}, \quad (8)$$

where G is the set of non-zero elements in the gold standard frame \mathbf{G} , and G^c is the complement of such set. Therefore, G^c is the set of zero elements in \mathbf{G} , and g' iterates over those points. $\mathbf{P}_{g'}$ is the element of the prediction frame \mathbf{P} indexed by g' . D_2 is inspired on false positive ratio, which is defined for binary classifications, whereas D_2 is computed on probability maps, not binary images. The combination of these two metrics D_1 and D_2 could show deeper insights about the results than other conventional metrics.

Some evaluation metrics used in this work require binary images as inputs, namely, two classes corresponding to background and foreground (the THV, in this case) in the scene. As the U-Net outputs provide probability maps, it is necessary to post-process these predictions to get binary images before computing the metrics. Such binarization can be performed in different ways and, consequently, provide different results. The first option considered was a fixed threshold value of 0.5. This approach has been tested, obtaining in general a number of foreground pixels under the average value known from the gold standard, as the probability values are often below such threshold. Considering other possibilities, one widely accepted approach is Otsu's [25] method, which is used to perform automatic image thresholding. An alternative binarization proposed in this work is based on the prior knowledge of the foreground pixels (PKFP). The number of foreground pixels can be estimated as the mean number of foreground pixels in the set of gold standard images, obtaining 2346.04 ± 371.45 in mean and standard deviation from the frames where the THV is fully expanded. This estimation allows to build a binarization method by setting a threshold that assigns to foreground class the 2500 pixels with the highest probability values in the predictions.

4. Experiments and Results

Several experiments have been conducted to evaluate the performance of the proposed methodology, as the evaluation of different U-Net configurations, the assessment of the prediction results for the model trained with the best configuration obtained in the previous experiment.

This work includes 15 patients, 50 different angiographic sequences and a total of 2827 frames. The mean age of the whole population was 84.39 ± 4.02 years and an 83.33% were females.

4.1. Hyperparameter Tuning for the Model Selection

In order to identify an appropriate U-Net configuration, several architecture implementations and operating conditions were considered and evaluated. The best model obtained in this evaluation will be used in a posterior deeper analysis.

To perform this task, the dataset was split into two sets, training and test, with a ratio of 80/20. Such sets were generated by random selections, ensuring that the frames used to test the model did not belong to the patients used during training. The training set consisted of 2026 frames where the THV was completely expanded and 267 frames containing the aortic root previous to the device deployment, without any THV. Therefore, the split generated a training set constituted by a total of 2293 frame-label pairs extracted from 40 angiographic sequences and, on the other hand, a test set made of 534 images, obtained from 10 sequences.

Training set size is one of the parameters which could have important effects on the training process. To study its repercussion on performance, different subsets were generated from the training set with 600, 300, 150, and 75 images each. The subsets were generated with random selections from bigger subsets, such that each of them is contained in the subsets of greater sizes.

Experiments were conducted with binary cross-entropy (1) or focal loss (4) as cost functions, as well as using or not data augmentation, yielding four possible configurations for each dataset. As there are four different datasets, it makes a total of 16 models studied in this evaluation. All these models were trained for 30 epochs, using a batch size of 2 and Adam optimizer with a learning rate of 10^{-4} . Figure 5 shows cost function evolution with training epochs of all the models used in this work, this being the 16 models considered in this evaluation and another one (black curve), which will be detailed in later sections. In all cases, it is observable the asymptotical behavior of the cost function. It can also be seen that focal loss values are always lower than binary cross-entropy, even from the first epoch.

Once the 16 models were trained, they were tested with the same test set, in order to obtain comparable results. This subset of images to test the models includes the 460 with fully expanded THV frames, extracted from the test set. The performance was evaluated in terms of D_1 (7) and D_2 (8), both metrics were computed for all the frames in the subset and expressed in mean and standard deviation values. All the results obtained in this evaluation of the U-Net configuration are provided in Table 1, where each row gathers the results of each tested configuration.

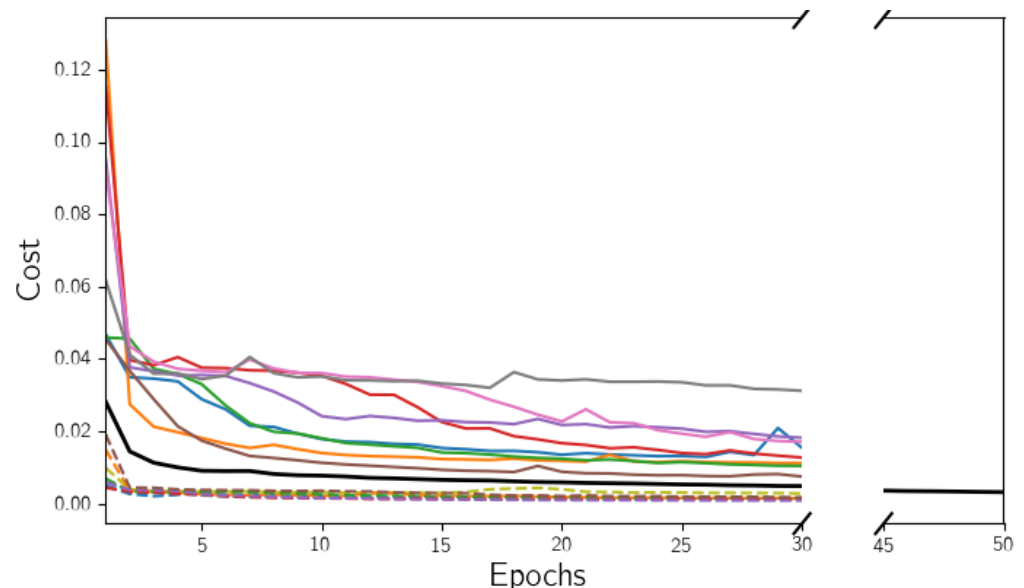


Figure 5. Cost function evolution with training epochs of all the models. Continuous lines correspond with the models using binary cross-entropy as cost function and dashed lines, with focal loss models. The black curve is the cost function of the best model, used in the evaluation in Section 4.2.

Table 1. Results of the 16 models considered in the evaluation of the U-Net configuration.

Training Set Size	Data Aug. (DA)	Loss Function	Best Epoch	Best Loss Value	D_1 (Otsu's Binarization)	D_1 (PKFP Binarization)	D_2
75	0	CE	30	0.0171	21.87 ± 26.43	6.03 ± 3.79	108.79 ± 81.54
		FL	6	0.0028	0.16 ± 0.55	7.02 ± 3.75	9314.98 ± 3149.51
	1	CE	30	0.0312	118.38 ± 218.96	65.96 ± 56.89	1082.49 ± 666.88
		FL	30	0.0028	0.093 ± 0.40	13.95 ± 8.07	11415.52 ± 3787.17
150	0	CE	30	0.0128	2.62 ± 3.69	1.75 ± 2.11	556.16 ± 232.48
		FL	30	0.0015	0.33 ± 0.56	2.54 ± 1.45	5041.36 ± 1334.57
	1	CE	30	0.0182	2.96 ± 4.87	7.82 ± 8.99	771.07 ± 313.18
		FL	30	0.0016	0.21 ± 0.42	2.66 ± 1.73	2983.14 ± 1011.65
300	0	CE	30	0.0085	3.65 ± 3.15	1.02 ± 1.07	92.81 ± 62.13
		FL	29	0.0010	0.79 ± 2.03	1.53 ± 2.49	1040.91 ± 220.49
	1	CE	26	0.0129	10.16 ± 17.12	9.92 ± 13.54	451.25 ± 161.92
		FL	30	0.0012	0.09 ± 0.18	0.95 ± 0.97	1804.71 ± 353.82
600	0	CE	30	0.0075	1.36 ± 1.56	0.63 ± 0.73	244.23 ± 90.97
		FL	30	0.0009	0.51 ± 3.40	1.33 ± 4.63	1025.85 ± 170.54
	1	CE	29	0.0112	2.28 ± 4.82	3.03 ± 5.02	550.40 ± 176.58
		FL	30	0.0012	0.22 ± 0.37	1.19 ± 1.14	1374.78 ± 233.04

In this table, both cross-entropy (CE) and focal loss (FL) are considered, and the activation or not of data augmentation is expressed with a binary variable, $DA \in \{0, 1\}$. Two measurements of D_1 are provided for each configuration, one where the binarization was obtained by Otsu's method and the other, using PKFP binarization. The table also includes the best loss value achieved during the training of each model and the number of the training epoch where such value has been obtained. Loss values are not comparable for different cost functions since their typical values are not in the same range. However, analyzing separately the configurations using each cost function, it can be observed that there is a relation between loss values and the other metrics, as in general the models with the lowest loss values obtain the best results in terms of D_1 and D_2 .

When a model obtains a high D_1 mean value, in many cases it is because the segmentations have few foreground pixels, and therefore tend to get low D_2 in average, as happens with the model trained with the dataset of 75 frames, data augmentation disabled, with cross-entropy as cost function, and using Otsu's method for the binarization (first row in Table 1). On the other hand, a low D_1 value can be due to a high number of foreground pixels in segmentations, yielding high D_2 values as the model trained with the dataset of 75 frames, data augmentation enabled, with focal loss as cost function, and using Otsu's method for the binarization (fourth row).

Figure 6 shows graphically the results from Table 1, both axes are in logarithmic scale for representation purposes. Each dot corresponds to a model and the coordinates are their D_1 and D_2 obtained mean values. Each line represents the evolution in performance of each configuration when the training set size increases, represented with an increase in the marker size. Continuous lines were obtained by computing D_1 on predictions binarized using Otsu's method, and discontinuous lines correspond to PKFP binarization. D_2 is the same, as it is computed over the predictions and not the binary segmentations, therefore the difference between continuous and discontinuous analogous lines is a shift in D_1 .

Regarding the results, in all cases the performance in terms of these metrics improves with training set size. It can be observed that in general the use of data augmentation does not provide an improvement in this problem. In addition, it is noticeable that binary cross-entropy configurations performed better than focal loss ones with this dataset.

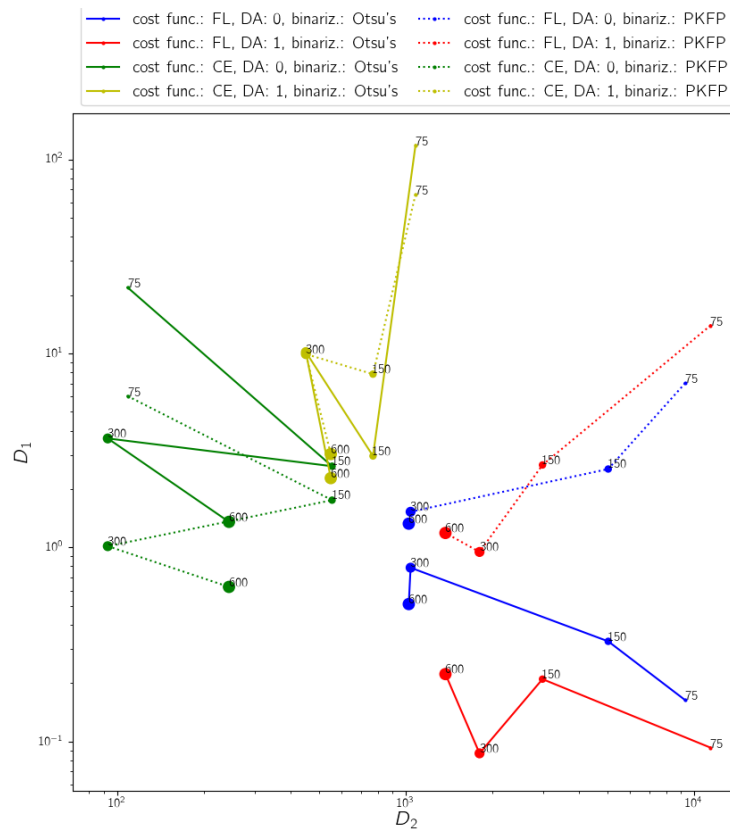


Figure 6. Graphical representation of the results in terms of D_1 and D_2 for each of the 16 models in Table 1. Each color corresponds to each of the four tested configurations. Blue: FL, $DA = 0$; Red: FL, $DA = 1$; Green: CE, $DA = 0$; Yellow: CE, $DA = 1$. Marker size refers to the number of images used to train each model (from 75 to 600). Continuous line D_1 values were computed using Otsu’s method, whereas discontinuous lines were obtained using PKFP binarization.

The values of D_1 and other metrics are importantly affected by the binarization method. The suitability of such binarization on the predictions of a model is related to the cost function used during the training, as the predictions obtained by each cost function have their own distinctive properties. It has been observed that cross-entropy models tend to obtain incomplete patterns, whereas focal loss models usually produce noisy predictions. Directing attention to focal loss lines in the graph, Otsu’s binarization obtains lower D_1 values and appears to perform better, which is also noticeable in the segmentation examples in Figure 7a,b. However, this is because the binarization assigns many pixels to foreground, reducing D_1 but, at the same time, obtaining false positives. For focal loss models, PKFP actually obtains binarizations more similar to the ground truth than Otsu’s method, as can be seen in Figure 7c,d. On the other hand, as cross-entropy often gets false negatives, excluding pixels in the pattern, forcing a number of foreground pixels in the binarizations that might be too many, yielding binarizations with many false positives, as shown in Figure 7c. Otsu’s method performs better than PKFP in the segmentations of models trained using cross-entropy.

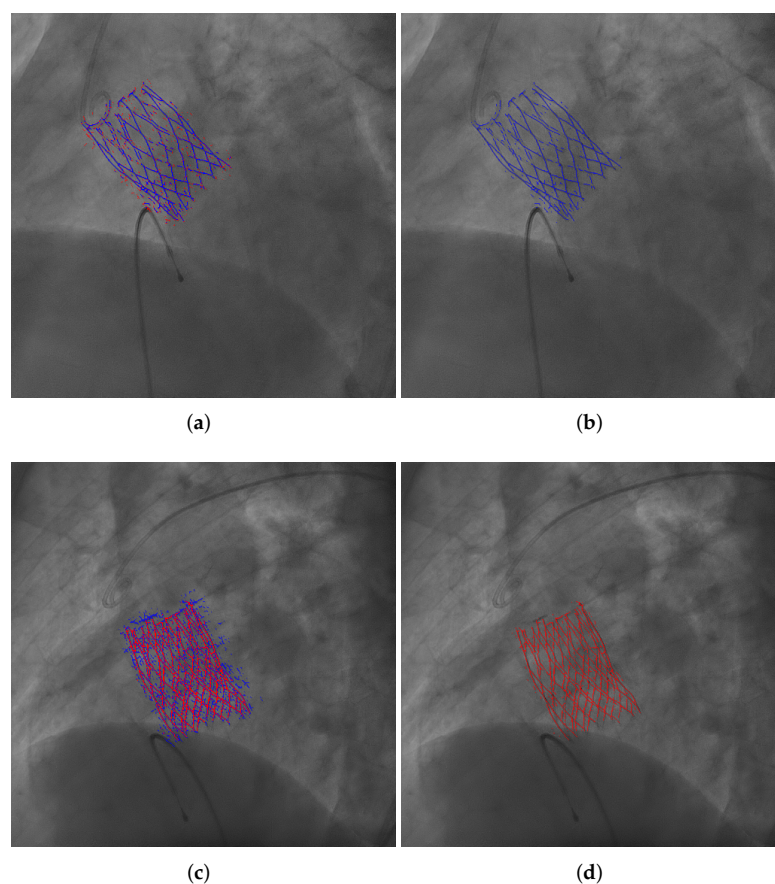


Figure 7. Examples of segmentation results obtained by different models, using Otsu’s method (blue) and PKFP (red) binarizations, overlaid on the original angiographic frames. **(a)** Segmentation obtained by the model trained with 600 frames, using cross-entropy and data augmentation disabled (13th row in Table 1), binarized using Otsu’s method and PKFP. **(b)** Same as **(a)**, but with only Otsu’s method result overlaid. **(c)** Segmentation obtained by the model trained with 600 frames, using focal loss and data augmentation enabled (16th row in Table 1), binarized using Otsu’s method and PKFP. **(d)** Same as **(c)**, but with only PKFP result overlaid.

Regarding the results in terms of D_1 and D_2 in the performed experiments (results provided in Table 1), and normalizing by the worst mean result obtained for each of the metrics, the best configuration would be the closest to the origin. In this case, the best tested configuration with this dataset is using binary cross-entropy as cost function and disabling data augmentation. In addition, the model segmentation performance improves as the number of training samples grows.

4.2. Evaluation of the Results of the Selected Model

The best parameter configuration obtained in the previous section was used to train a new model, now using the whole training set, which is a total of 2293 frames. In addition, the number of training epochs was increased to 50 to get better parameter estimations. The black curve in Figure 5 shows the evolution of the cost function values obtained in each training epoch.

Several evaluations and experiments were conducted, assessing the model on different subsets of images, all of them belonging to the test set, not included in the training. Table 2 gathers the results of all the performed experiments, where each column in the table corresponds to an evaluation. The metrics are computed over several frames and the results are expressed in mean and standard deviation for different image sets. For metrics requiring binary inputs, the results are computed on images binarized using both Otsu’s

method and PKFP. The number of test frames used in each experiment is also specified in the table.

Table 2. Results of the selected model in the experiments.

Metric	Bin. Method	Exp. 1 (460 Frames)	Exp. 2 (65 Frames)	Exp. 3 (3 Frames)	Exp. 4 (6 Frames)
D_1	Otsu's	0.91 ± 0.86	–	67.95 ± 75.11	10.21 ± 8.73
	PKFP	0.30 ± 0.38	–	52.94 ± 62.72	3.92 ± 4.16
D_2	–	138.48 ± 75.76	2.40 ± 9.21	204.45 ± 19.83	369.83 ± 48.50
Accuracy	Otsu's	0.99 ± 0.01	0.99 ± 0.01	0.99 ± 0.01	0.99 ± 0.01
	PKFP	0.99 ± 0.01	0.99 ± 0.01	0.99 ± 0.01	0.99 ± 0.01
Precision	Otsu's	0.73 ± 0.05	0	0.51 ± 0.06	0.46 ± 0.06
	PKFP	0.63 ± 0.07	0	0.43 ± 0.06	0.34 ± 0.04
Recall	Otsu's	0.87 ± 0.04	0	0.70 ± 0.09	0.75 ± 0.04
	PKFP	0.97 ± 0.02	0	0.80 ± 0.09	0.93 ± 0.03
Dice Coefficient	Otsu's	0.15 ± 0.03	0	0.29 ± 0.05	0.33 ± 0.04
	PKFP	0.22 ± 0.06	0	0.36 ± 0.05	0.47 ± 0.06
Jaccard Index	Otsu's	0.74 ± 0.04	0	0.55 ± 0.06	0.50 ± 0.05
	PKFP	0.64 ± 0.07	0	0.47 ± 0.06	0.36 ± 0.03
AUROC	Otsu's	0.94 ± 0.02	–	0.85 ± 0.05	0.89 ± 0.02
	PKFP	0.98 ± 0.01	–	0.90 ± 0.04	0.96 ± 0.01

The first experiment evaluates the U-Net model on frames where the THV is fully expanded, computing the metrics on the 460 frames in the test set; an example of the obtained results is shown in Figure 4c. Comparing these results with the ones obtained by the best configuration in the previous section, which was tested on the same set of images, it can be observed an improvement in D_1 and D_2 , an effect of the increase in the training set size and the number of training epochs.

The second experiment analyzes the model performance on 65 frames belonging to the test set, containing the aortic root before the deployment, where the THV is not present. D_1 cannot be obtained, as there is not any pattern to identify. Furthermore, AUROC is not defined for ground truths where only one class is present, and precision, recall, Dice coefficient, and Jaccard index are equal to 0, independently of the correctness of the segmentation. D_2 is specially insightful in this scenario, obtaining a rather low value, indicating that the model performs well in these frames. It should be noted that, even though it is not reflected in the metrics, PKFP binarization is not opportune in this set of images, as the optimum segmentation would not include any foreground pixels.

The third experiment evaluates the methodology performance to identify the THV during its deployment using three frames, obtaining segmentations as the example in Figure 4b. Results in terms of some metrics are significantly worse than previous experiments. It must be taken into account that the device shape during deployment is not that similar to the bioprosthesis fully deployed and the model has not been trained to identify those different shapes. This fact makes these images harder to segment and the results are often incomplete.

The fourth experiment addresses the effect of injecting contrast, evaluating six frames, an example of a segmentation result is shown in Figure 4d. Even though the model identifies part of the THV correctly, D_1 and D_2 are impaired due to the contrast part, where significant errors occur.

All the experiments achieve high accuracy values. This occurs because almost all background pixels are correctly classified which, given the strong class imbalance, entails

a great part of the frame. All the experiments show a poor performance in terms of Dice coefficient, which can be explained by the fact that this metric measures the overlapping between areas, but the THV in the frames is a thread-like structure. The experiments have shown the important impact of the binarization method on the results, obtaining noticeable differences in all cases, except accuracy. Precision and Jaccard index get better results using Otsu's method, whereas PKFP binarization improves recall and Dice coefficient results.

5. Discussion

The automatic interpretation of angiographic images, if obtained in an unattended manner, can provide valuable additional information during TAVI. The extraction of this information usually may rely on the automatic detection of the bioprosthesis in the image sequence. This work presents a fully automatic way of identifying and segmenting THV on angiographic sequences, training a U-Net and studying different configurations aiming to improve the segmentation performance. Given an annotated dataset of TAVI angiographies, including frames of different stages of the procedure, the method has demonstrated its performance with AUROC values between 0.85 and 0.98 in the experiments.

Validation is of main importance in this kind of applications. Finding the right metrics and assessing the quality of the results can be more difficult than obtaining the solution itself. In this work, standard metrics based on pixel-wise evaluation are used, and new metrics are specifically proposed for this problem. The defined D_1 and D_2 measure more appropriately the similarity between segmentations and the gold standard. The binarization also has an impact on validation, as several metrics require binary inputs. Using a threshold of 0.5 did not perform correctly; thus, other schemas have been considered, analyzing their suitability for the problem. Both Otsu's method and the defined PKFP binarization have demonstrated their validity; however, it has been observed a relation with the cost function used during the model training: Otsu's method is the preferred option for cross-entropy, whereas focal loss models require PKFP binarization. Notice that such PKFP binarization requires a priori knowledge of the THV and it is only valid when the device is present, not being suitable on angiographic frames previous to the delivery.

Concerning the benefits for the clinical practice, the main contribution of this work relies on the possibility of using TAVI segmentation on angiographic sequences, automatically delineating the THV. This work is a first step that would permit the extraction of quantitative features from the images, which could provide several advantages as reducing subjectivity of the physicians during the procedure to get the correct placement of the bioprosthesis or check the device expansion in real-time. The latter may help physicians to decide objectively the need for conducting different measures during TAVI to obtain optimal results (balloon post dilatation due to THV infra expansion or significant paravalvular leakage), and ensure better long-term outcome. Furthermore, the information that could be extracted from the sequences, if obtained in real time, could lead to the automation of the procedure. However, this is a major challenge to be explored, beyond the scope of this paper.

The methodology presented in this work has proved success in the automatic identification of Allegra THV on X-ray angiographic sequences. However, it must be taken into account that this method could also be used to identify other THV models and devices from other manufacturers. Even more, the methodology could be applied in the general problem of identification of ridge structures in X-ray sequences, extending significantly the field of application.

6. Conclusions

This paper presents an automatic method for detecting the THV prosthesis on angiographic sequences using deep learning techniques, without the need of any human interaction. The methodology relies on a U-Net architecture, including a hyperparameter tuning stage for the selection of a proper configuration, and an evaluation of the obtained segmentation results. Such evaluation considers different conventional metrics as well

as two new ones, defined for this specific purpose, in addition to the study of the impact of different binarization schemas in the metrics results. The models are trained using an annotated dataset and tested in different kind of frames captured during TAVI procedure, showing satisfactory results.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/diagnostics12020334/s1>, Figure S1: angioseq.gif. Figure S2: seg.gif.

Author Contributions: Conceptualization, L.B. and C.V.; Data curation, L.B. and C.V.; Formal analysis, L.B. and C.V.; Funding acquisition, C.V. and A.Í.; Investigation, L.B., C.V., J.A.G.-N., M.L.-G., V.J., J.A.B. and A.Í.; Methodology, L.B. and C.V.; Project administration, C.V. and A.Í.; Software, L.B.; Supervision, C.V.; Validation, L.B., C.V., J.A.G.-N., M.L.-G., V.J., J.A.B. and A.Í.; Visualization, L.B.; Writing—original draft, L.B. and C.V.; Writing—review & editing, L.B., C.V., J.A.G.-N., M.L.-G., V.J., J.A.B. and A.Í. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by the Aid for the promotion of youth employment and implementation of the Youth Guarantee in R+D+i of the Agencia Estatal de Investigación, co-financed by the European Social Fund “The ESF invests in your future” (Code PEJ2018-005278-A), and by Axencia Galega de Innovación (GAIN) (Code Number IN845D-2020/29 and IN607B-2021/18).

Institutional Review Board Statement: The study was approved by Regional Ethics Committee for Research (CEIm-G), code 2019/317.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data not publicly available due to privacy restrictions, but available from the corresponding author C.V., upon reasonable request.

Acknowledgments: The authors would like to thank the Cardiovascular Research Unit in Álvaro Cunqueiro Hospital for their support on the identification of patients studies. Authors also thank the Centro de Supercomputación de Galicia (CESGA) for providing computational resources required in this work.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Holmes, D.R.; Mack, M.J.; Kaul, S.; Agnihotri, A.; Alexander, K.P.; Bailey, S.R.; Calhoun, J.H.; Carabello, B.A.; Desai, M.Y.; Edwards, F.H.; et al. 2012 ACCF/AATS/SCAI/STS Expert Consensus Document on Transcatheter Aortic Valve Replacement. *J. Am. Coll. Cardiol.* **2012**, *59*, 1200–1254. [[CrossRef](#)] [[PubMed](#)]
- Leon, M.B.; Smith, C.R.; Mack, M.J.; Makkar, R.R.; Svensson, L.G.; Kodali, S.K.; Thourani, V.H.; Tuzcu, E.M.; Miller, D.C.; Herrmann, H.C.; others. Transcatheter or surgical aortic-valve replacement in intermediate-risk patients. *N. Engl. J. Med.* **2016**, *374*, 1609–1620. [[CrossRef](#)] [[PubMed](#)]
- Bloomfield, G.S.; Gillam, L.D.; Hahn, R.T.; Kapadia, S.; Leipsic, J.; Lerakis, S.; Tuzcu, M.; Douglas, P.S. A practical guide to multimodality imaging of transcatheter aortic valve replacement. *JACC Cardiovasc. Imaging* **2012**, *5*, 441–455. [[CrossRef](#)] [[PubMed](#)]
- Blanke, P.; Weir-McCall, J.R.; Achenbach, S.; Delgado, V.; Hausleiter, J.; Jilaihawi, H.; Marwan, M.; Nørgaard, B.L.; Piazza, N.; Schoenhagen, P.; et al. Computed tomography imaging in the context of transcatheter aortic valve implantation (TAVI)/transcatheter aortic valve replacement (TAVR) an expert consensus document of the Society of Cardiovascular Computed Tomography. *JACC Cardiovasc. Imaging* **2019**, *12*, 1–24. [[CrossRef](#)] [[PubMed](#)]
- Kurra, V.; Kapadia, S.R.; Tuzcu, E.M.; Halliburton, S.S.; Svensson, L.; Roselli, E.E.; Schoenhagen, P. Pre-Procedural Imaging of Aortic Root Orientation and Dimensions: Comparison between X-ray Angiographic Planar Imaging and 3-Dimensional Multidetector Row Computed Tomography. *JACC Cardiovasc. Interv.* **2010**, *3*, 105–113. [[CrossRef](#)]
- Binder, R.; Leipsic, J.; Wood, D.; Moore, T.; Toggweiler, S.; Willson, A.; Gurvitch, R.; Freeman, M.; Webb, J. Prediction of Optimal Deployment Projection for Transcatheter Aortic Valve Replacement Angiographic 3-Dimensional Reconstruction of the Aortic Root Versus Multidetector Computed Tomography. *Circ. Cardiovasc. Interv.* **2012**, *5*, 247–252. [[CrossRef](#)]
- Martin, C.; Sun, W. Transcatheter valve underexpansion limits leaflet durability: Implications for valve-in-valve procedures. *Ann. Biomed. Eng.* **2017**, *45*, 394–404. [[CrossRef](#)]
- Wagner, M.G.; Hatt, C.R.; Dunkerley, D.A.P.; Bodart, L.E.; Raval, A.N.; Speidel, M.A. A dynamic model-based approach to motion and deformation tracking of prosthetic valves from biplane x-ray images. *Med. Phys.* **2018**, *45*, 2583–2594. [[CrossRef](#)]
- Arsalan, M.; Owais, M.; Mahmood, T.; Choi, J.; Park, K.R. Artificial intelligence-based diagnosis of cardiac and related diseases. *J. Clin. Med.* **2020**, *9*, 871. [[CrossRef](#)]
- Wang, X.; Wang, F.; Niu, Y. A Convolutional Neural Network Combining Discriminative Dictionary Learning and Sequence Tracking for Left Ventricular Detection. *Sensors* **2021**, *21*, 3693 [[CrossRef](#)]

11. Galea, R.R.; Diosan, L.; Andreica, A.; Popa, L.; Manole, S.; Bálint, Z. Region-of-Interest-Based Cardiac Image Segmentation with Deep Learning. *Appl. Sci.* **2021**, *11*. [[CrossRef](#)]
12. Dey, D.; Slomka, P.J.; Leeson, P.; Comaniciu, D.; Shrestha, S.; Sengupta, P.P.; Marwick, T.H. Artificial intelligence in cardiovascular imaging: JACC state-of-the-art review. *J. Am. Coll. Cardiol.* **2019**, *73*, 1317–1335. [[CrossRef](#)] [[PubMed](#)]
13. Taha, A.A.; Hanbury, A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Med. Imaging* **2015**, *15*, 29. [[CrossRef](#)] [[PubMed](#)]
14. Li, Z.; Kamnitsas, K.; Glocker, B. Analyzing Overfitting Under Class Imbalance in Neural Networks for Image Segmentation. *IEEE Trans. Med. Imaging* **2021**, *40*, 1065–1077. [[CrossRef](#)]
15. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, Munich, Germany, 5–9 October 2015.
16. Lin, T.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2999–3007. [[CrossRef](#)]
17. Moccia, S.; De Momi, E.; El Hadji, S.; Mattos, L.S. Blood vessel segmentation algorithms—Review of methods, datasets and evaluation metrics. *Comput. Methods Programs Biomed.* **2018**, *158*, 71–91. [[CrossRef](#)] [[PubMed](#)]
18. DeLong, E.; DeLong, D.; Clarke-Pearson, D. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* **1988**, *44*, 837–845. [[CrossRef](#)]
19. Dice, L.R. Measures of the amount of ecologic association between species. *Ecology* **1945**, *26*, 297–302. [[CrossRef](#)]
20. Jaccard, P. The distribution of the flora in the alpine zone. 1. *New Phytol.* **1912**, *11*, 37–50. [[CrossRef](#)]
21. Rockafellar, R.T.; Wets, R.J.B. *Variational Analysis*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2009; Volume 317.
22. Chollet, F. Keras. Available online: <https://keras.io> (accessed on 22 December 2021).
23. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. 2015. Available online: [tensorflow.org](https://www.tensorflow.org) (accessed on 22 December 2021).
24. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv1412.6980.
25. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **1979**, *9*, 62–66. [[CrossRef](#)]
26. Luque, A.; Carrasco, A.; Martín, A.; de las Heras, A. The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognit.* **2019**, *91*, 216–231. [[CrossRef](#)]