



Improving convolutional neural network learning based on a hierarchical bezier generative model for stenosis detection in X-ray images

Emmanuel Ovalle-Magallanes^a, Juan Gabriel Avina-Cervantes^a, Ivan Cruz-Aceves^{b,*}, Jose Ruiz-Pinales^a

^a Telematics and Digital Signal Processing Research groups (CAs), Engineering Division, Campus Irapuato-Salamanca, University of Guanajuato, Carretera Salamanca-Valle de Santiago km 3.5 + 1.8km, Comunidad de Palo Blanco, Salamanca, 36885 Guanajuato, Mexico

^b CONACYT, Center for Research in Mathematics (CIMAT), A.C., Jalisco S/N, Col. Valenciana, Guanajuato, 36000 Guanajuato, Mexico

ARTICLE INFO

Article history:

Received 17 August 2021

Revised 9 March 2022

Accepted 19 March 2022

Keywords:

Stenosis detection

X-Ray coronary angiography

Attention network

Generative model

ABSTRACT

Background and Objective: Automatic detection of stenosis on X-ray Coronary Angiography (XCA) images may help diagnose early coronary artery disease. Stenosis is manifested by a buildup of plaque in the arteries, decreasing the blood flow to the heart, increasing the risk of a heart attack. Convolutional Neural Networks (CNNs) have been successfully applied to identify pathological, regular, and featured tissues on rich and diverse medical image datasets. Nevertheless, CNNs find operative and performing limitations while working with small and poorly diversified databases. Transfer learning from large natural image datasets (such as ImageNet) has become a de-facto method to improve neural networks performance in the medical image domain.

Methods: This paper proposes a novel Hierarchical Bezier-based Generative Model (HBGM) to improve the CNNs training process to detect stenosis. Herein, artificial image patches are generated to enlarge the original database, speeding up network convergence. The artificial dataset consists of 10,000 images containing 50% stenosis and 50% non-stenosis cases. Besides, a reliable Fréchet Inception Distance (FID) is used to evaluate the generated data quantitatively. Therefore, by using the proposed framework, the network is pre-trained with the artificial datasets and subsequently fine-tuned using the real XCA training dataset. The real dataset consists of 250 XCA image patches, selecting 125 images for stenosis and the remainder for non-stenosis cases. Furthermore, a Convolutional Block Attention Module (CBAM) was included in the network architecture as a self-attention mechanism to improve the efficiency of the network.

Results: The results showed that the pre-trained networks using the proposed generative model outperformed the results concerning training from scratch. Particularly, an accuracy, precision, sensitivity, and F1-score of 0.8934, 0.9031, 0.8746, 0.8880, 0.9111, respectively, were achieved. The generated artificial dataset obtains a mean FID of 84.0886, with more realistic visual XCA images.

Conclusions: Different ResNet architectures for stenosis detection have been evaluated, including attention modules into the network. Numerical results demonstrated that by using the HBGM is obtained a higher performance than training from scratch, even outperforming the ImageNet pre-trained models.

© 2022 Elsevier B.V. All rights reserved.

1. Introduction

Computer-Aided Diagnosis (CAD) systems play a paramount role in medical examination, improving diagnostic accuracy and reducing workloads. In particular, the physician performs an exhaustive visual inspection of the X-ray Coronary Angiography (XCA) images to detect Coronary Artery Stenosis (CAS) [1]. CAS is an abnormal narrowing of the artery that supplies rich-oxygen blood to the

* Corresponding author.

E-mail addresses: e.ovallemagallanes@ugto.mx (E. Ovalle-Magallanes), avina@ugto.mx (J.G. Avina-Cervantes), ivan.cruz@ciamat.mx (I. Cruz-Aceves), pinales@ugto.mx (J. Ruiz-Pinales).

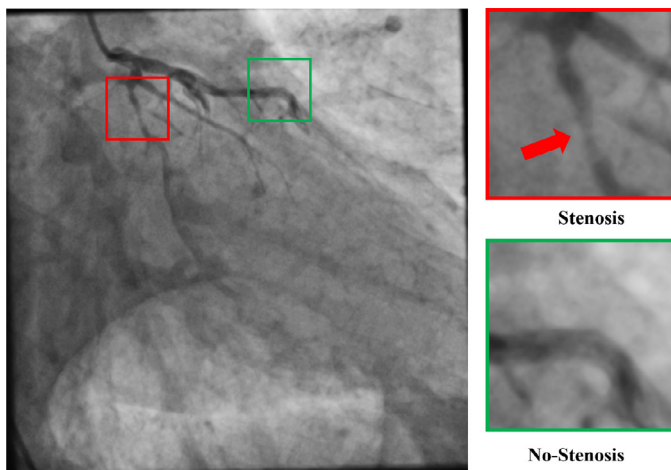


Fig. 1. The XCA test requires injecting a contrast agent into the coronary artery tree, revealing the arterial structure easily observed on X-ray images. A visible stenosis region is marked in a red box, including a red arrow indicating the position of the stenosis and a non-stenosis region in green.

heart due to the accumulation of cholesterol plaque buildups [2]. This condition is a common type of coronary artery disease leading the causes of cardiovascular diseases, estimated at 19.8 million global deaths every year. According to the World Health Organization, it persists as the leading cause of mortality [3]. Therefore, early stenosis detection is essential in cardiology to diagnose and provide suitable medical treatment before permanent heart damage.

Among the medical imaging modalities for diagnosing stenosis and related conditions, such as Coronary Computed Tomography Angiography (CCTA) or cardiac magnetic resonance imaging (CMRI), the XCA remains the gold-standard modality [4]. XCA provides assistance for both stenosis evaluation and artery blockage removal by angioplasty, which CCTA is unable to support. Nevertheless, the automation of this task is considered a challenging problem up to date because of background noise [5], non-coronary vascular structures, and multiple superposed branching points [6], among others. Fig. 1 shows a representative sample of an XCA image, where two regions are highlighted. A stenosis region with a rapid reduction of vessel diameter can be observed, and a non-stenosis region (*i.e.*, uniform tubular shape of the blood vessel).

Traditional machine learning techniques for stenosis detection in coronary angiograms include three major steps: handcrafted feature extraction, feature selection, and classification. In the first two steps, design decisions are made using expert knowledge about the problem (*i.e.*, blood vessel shape). Wan et al. [7] introduced a framework for an automated quantitative stenosis assessment, including a Hessian-based vessel enhancement and vessel skeleton extraction as feature extraction. Afterward, for each pixel placed on the vessel's skeleton, an improved Measure of Match (MoM) compute the vessel diameter. Finally, the location and degree of stenosis are based on the local extremum of the diameter curve. Kishore and Jayanthi [8] proposed a four-step system combining: pre-processing, artery segmentation, stenosis detection, and grading. Median filtering and a Contrast Limited Adaptive Histogram Equalization (CLAHE) highlight the coronary artery tree from the background. Next, adaptive thresholding and morphological operations generate a refined coronary artery segmentation. Finally, the width of a vessel segment is calculated by adding the intensity values from its left edge to the right edge to determine stenosis candidates. A percentage of stenosis is computed if the evaluated patch is a positive stenosis case. Sameh et al. [9] introduced a narrowed coronary artery detection and classification approach

by joining the Frangi Vesselness filter and Threshold Morphological filter to detect and segment vessels, respectively. Then, a region growing algorithm enhance vessel segmentation results. From this segmented image, the artery lumen diameter is measured and classified according to the percentage of blockage applying the k-Nearest Neighbor (kNN) classifier.

On the other hand, Convolutional Neural Networks (CNNs) extract, select and classify features within the optimization process of the model in an end-to-end manner. CNNs have succeeded in solving complex tasks in the medical imaging domain, such as segmentation, lesion localization, and classification [10]. However, generating annotated data is costly and requires much handy effort from specialists. Methods have appeared to improve CNNs performance addressing insufficient training data, *e.g.*, using small image patches rather than full images for training [11], data augmentation [12], transfer learning [13], Generative Adversarial Networks (GANs) [14], and self-attention mechanisms [15]. Thereunto, patch-based learning strategies crop the entire image into patches to provide more training samples for the CNN. In particular, data augmentation includes a classic combination of affine transformations based on features as geometry, color, and random erasing to produce modified images from the initial dataset [12].

Alternatively, the transfer learning paradigm has shown that it is very efficient to start learning from a pre-trained network despite the dissimilarity between two datasets and respective aimed tasks [13]. Conversely, the generator in a GAN learns the distribution from its training data from which new data can be sampled. This operation can be seen as another type of transformation of the original data. At par with the other methods, the self-attention mechanism recalibrates the intermediate feature maps adaptively by considering robust inter-channel and inter-spatial relations [15,16].

In the context of this research, only a few methods based on Deep Learning for stenosis detection in XCA images have been proposed. Antczak and Liberadzki [17] introduced a patch-based CNN consisting of five layers: four convolutional and one dense layer. This approach incorporates dropout layers after the convolutional layers to avoid overfitting. An artificial dataset generated by random Bezier curves simulating blood vessels was employed during the training process. Ovalle et al. [18] used Transfer Learning evaluating three ImageNet pre-trained networks (VGG16 [19], InceptionV3 [20], and ResNet50 [21]) to deal with the limited amount of training data. The proposed approach includes an optimal selection of layers to be transferred and fine-tuned. Au et al. [22] proposed a two-step method to detect and classify stenosis. First, a localization module obtains a bounding box for prediction. The focused region is processed through a compact CNN containing five convolutional layers, a single global max-pooling layer, and a fully connected layer. Subsequently, data augmentation techniques are applied during the training phase.

In contrast, some approaches work directly with the raw-XCA images and exploit the availability of a rich-featured database. Cong et al. [23] used an Inception-v3 [20] architecture pre-trained on the large ImageNet dataset [24]. The authors maintain the weights idle (frozen) during the optimization process, where the last layer is only trained. During the training, redundant frames of the XCA sequence were utilized to speed-up convergence and improve the network performance. Wu et al. [25] developed a Deconvolutional Single-Shot multi-box Detector based on a VGG19 network model [19], generating the most suitable bounding box for a stenosis region. Additionally, they exploit the potential temporal consistency of consecutive XCA frames. These frames are also required to train and improve detection results. Pang et al. [26] proposed a two-stage object detector based on a ResNet50 [21] architecture. Firstly, the feature maps are extracted, generating candidate boxes. Secondly, the candidate boxes are classified as steno-

sis or non-stenosis. Feature extraction and fusion sequence are introduced to correlate the candidate boxes from consecutive XCA frames to increase classification accuracy. Moon et al. [27] utilized a pre-trained Inception-v3 [20] model to classify stenosis in XCA images. After each inception module, a Convolutional Block Attention Module (CBAM) [15] is included to obtain the channel and spatial information recalibration. The entire XCA image dataset is represented with a subset of images extracted through automatic key-frame detection algorithms. Next, each image feeds the attention CNN model to classify the stenosis cases. For every key-frame, distinct types of augmentation strategies were deployed and evaluated. Danilov et al. [28] trained and tested eight detectors based on different neural network architectures to detect the location of stenotic lesions using angiography imaging series. The use of RFCN ResNet-101 V2 showed a balanced trade-off between real-time performance and accuracy.

These last methods utilize a full angiographic test consisting of thousands of frames and data augmentation techniques. However, one main drawback of data augmentation is that some deformations may directly impact the degree of stenosis, *i.e.*, elastic deformations [29]. In such a way, a stenosis case could no longer contain a blood vessel narrowing.

Nowadays, only a handful of blood vessel generation methods have been explored in the literature. For instance, Keeland et al. [30] introduced a simulated annealing approach to generate 3D coronary arterial trees. Such a method requires prior information about the tissue structure and entry-point positions of the largest arteries. Jaquet et al. [31] proposed an algorithm for generating cardiac vascular trees, emerging from actual vascular tree models segmented from CTA images. Tetteh et al. [32] presented a neural network to generate 3-D angiographic volumes based on a computational angiogenesis model constrained to physiological and topological structure to simulate vascular tree growth. Antczak and Liberadzki [17] created and simulated real XCA patches based on the assumption that a Bezier curve involving additive random noise can parametrize a blood vessel. However, the curves are drawn with independent parameters, generating patches with non-vascular structures and non-visible stenosis areas due to curves overlapping. Thus, bifurcation structures are not guaranteed.

Nevertheless, studying bifurcations in the coronary vascular tree is crucial to accurately classify and localize coronary bifurcation lesions [6,33–35]. For such a reason, it is paramount to model this specific type of vascular structure. This study proposes a hierarchical generative model that creates faithfully looking data, producing images including complex blood vessel structures. The main blood vessel can include multiple bifurcations and stenosis regions, generating realistic coronary arterial tree patches. Previously, Antczak and Liberadzki [17] addressed a similar problem while working with small databases for stenosis detection in XCA images.

Additionally, the ground truth for segmentation and stenosis localization labels are generated during the same pipeline of the proposed generative model. Thus, a multi-purpose dataset is provided. The evaluation of these two extra datasets for blood vessel segmentation and stenosis localization is out of the scope of this work.

Furthermore, the Fréchet Inception Distance (FID) [36] was successfully applied to obtain a quantitative similarity metric between real and generated data. These diversified artificial data can hereafter be employed to pre-train a CNN for stenosis detection. Consecutively, the network weights pre-learned on the artificial dataset are thereby efficiently retrained using the real XCA image patches.

The ResNet [21] family architecture was selected to identify stenosis and non-stenosis cases in XCA images in this research. The ResNet model introduces residual mapping and shortcut connections to avoid degradation (of training accuracy) even though

the architecture is getting deeper. In particular, the ResNet-18, 34, and 50 have shown a good trade-off between depth and performance when small medical imaging datasets are evaluated [26,37–39]. Furthermore, a CBAM is incorporated into the network architecture to exploit the attention mechanism and enhance generalization when the network is trained from scratch.

This paper aims to deal with the lack of annotated data for accurate stenosis detection in XCA images by generating synthetic images that model real coronary vascular structure regions. The significant contributions of this paper are as follows:

1. A robust Bezier-based generative model generates faithfully image patches with blood vessel structures, including bifurcations and stenosis areas.
2. A quantitative measure of the generative model is obtained using the Fréchet Inception Distance.
3. A self-attention mechanism is incorporated within a ResNet network for stenosis detection.

2. Methods

An extensive training dataset is required to train a CNN correctly. For such a reason, in this study, a Hierarchical Bezier-based Generative Model (HBGM) is proposed to create a large synthetic dataset to perform the pre-training step. The generative model can generate blood vessels of distinct diameters, vessel bifurcations, and different grades, lengths, and localization of stenosis cases. Therefore, each generated image contains p -parents blood vessels, each with c -children blood vessels. Suppose the number of children of each parent is set to zero; the stenosis position and length remain constant. In that case, the generative model behaves as introduced by Antczak and Liberadzki [17]. Fig. 2 shows the general pipeline of the HBGM and the training process.

2.1. Hierarchical bezier-based generative model

A Hierarchical Bezier-based Generative Model (HBGM) is presented in this work. Small regions of XCA artery blood vessels are modeled as gray-scale images representing a set of curves of several lengths, drawn on a gradient background, and noise-distorted [17]. Notice that no previous information of real XCA is used. However, two constraints are employed to accept a generated patch. First, the ratio of blood vessel pixels concerning the size of the image must be above a threshold. Secondly, for patches where stenosis is created, the ratio of stenosis blood vessel pixels must be greater than a fixed value. As such, the proposed generative model creates images including blood vessel structures with stenosis regions, containing 10k images, 50% with stenosis, and 50% with no-stenosis cases. The generative process is divided into three steps: drawing area and gradient background generation, Bezier curve parametrization, and Bezier curve drawing.

Drawing area and background generation

Let (w, h) be the size of the patch to be generated, with a given width and height, respectively. A white drawing area or canvas DA of size $(3w, 3h)$ is created, such that the final patch (Region of Interest: ROI) is placed at the central position of the canvas. The canvas is formally defined as

$$DA = 255 \cdot \mathbf{J}_{3w, 3h}, \quad (1)$$

where \mathbf{J} is an all-ones matrix. Let us define U as an uniformly distributed random variable in the interval $[0,1]$, then an uniformly distributed function in the interval $[a, b]$, here denoted as $rand(a, b)$, can be given by

$$rand(a, b) = a + (b - a) \cdot U. \quad (2)$$

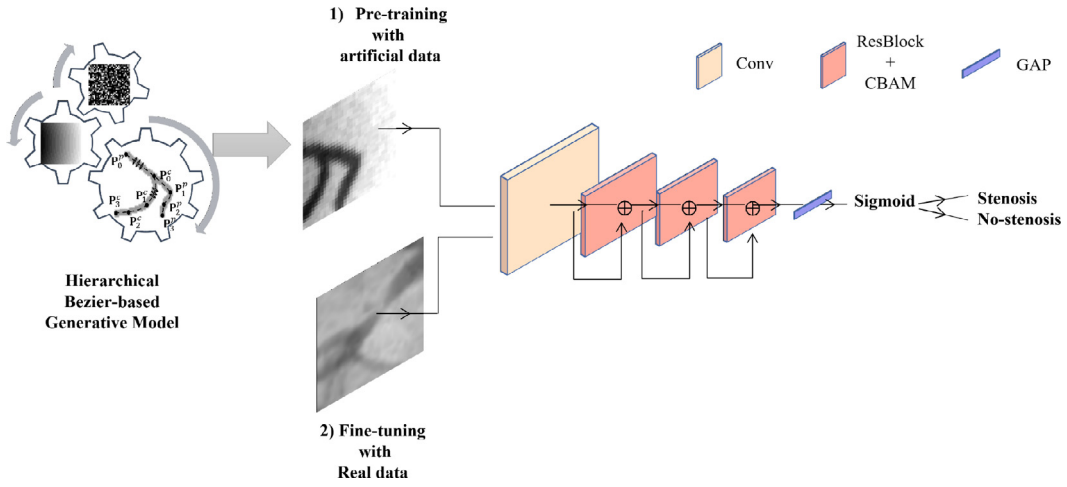


Fig. 2. A Hierarchical Bezier-based Generative model creates images including different blood vessel structures: multiple blood vessel widths, bifurcations, and stenosis. This synthetic dataset feeds a CNN to perform a pre-training step to improve the stenosis detection of real XCA image patches.

The radial gradient background is generated, centered at $c_g = (rand(-w, 2w), rand(-h, 2h))$. Thus, the intensity of each pixel at the ROI is given by

$$ROI(u) = \alpha \cdot (1 - d(u, c_g)) + \beta \cdot d(u, c_g), \quad |u = \{(i, j)\} \quad \forall i \in [0, w], j \in [0, h], \quad (3)$$

where $\alpha = rand(0, a)$ and $\beta = rand(a, b)$ with $0 < a < b \leq 255$ are the lower and upper pixel intensities. $d(u, c_g)$ is a distance between the position $u(i, j)$ and the gradient center given by

$$d(u, c_g) = \frac{1}{2} \frac{|u - c_g|}{\sqrt{w^2 + h^2}}, \quad d \in (0, 1). \quad (4)$$

Bezier curve parametrization

Under the prior assumption that a Bezier curve can parameterize a vessel, an arterial vessel structure can be defined as a central (parent) Bezier curve $B^{(p)}$. Then this parent curve can hold a subset of c Bezier curve children $B^{(c)} = \{B_1^{(c)}, B_2^{(c)}, \dots, B_c^{(c)}\}$. Moreover, it holds that any child's width is smaller than their respective parent's width. In such a way, a more complex vascular model can be accomplished. Thus, any Bezier curves are constructed by their cubic form as

$$B(t) = (1-t)^3 P_0 + 3(1-t)^2 t P_1 + 3(1-t) t^2 P_2 + t^3 P_3, \quad t \in [0, 1], \quad (5)$$

where $P = [P_0, P_1, P_2, P_3]^T$ is a set of control points. The first and the last control points are always the curve's endpoints.

The parent control points are randomly chosen to lie inside the canvas region, given by

$$P^{(p)} = \begin{bmatrix} rand(-w, 2w) & 0 \\ rand(0, w) & rand(0, h) \\ rand(0, w) & rand(0, h) \\ rand(-w, 2w) & h \end{bmatrix}, \quad (6)$$

where the control points $P_1^{(p)}$ and $P_2^{(p)}$ determining the shape of the curve are restricted to lie into the ROI. Thus, it follows that the control point $P_0^{(c)}$ of any child curve is subject to lie on the parent curve $B^{(p)}$, such as

$$P_0^{(c)} = [rand(B_0^{(p)}, B_{ts}^{(p)})], \quad (7)$$

where ts is the number of points that a Bezier curve was discretized (e.g., if $ts = 100$, the curve is constructed by the values $t = [0, 0.01, 0.02, \dots, 0.98, 0.99, 1]$. Fig. 3 shows one parent with one child Bezier curve, respectively.

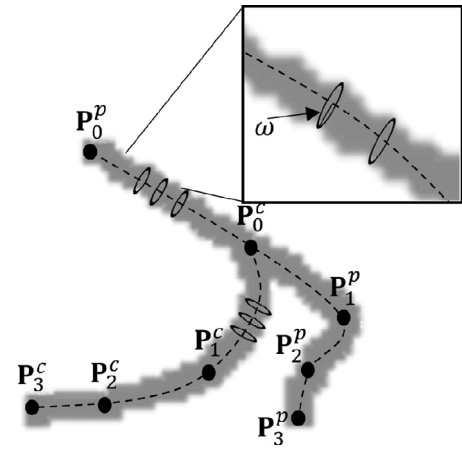


Fig. 3. Cubic Bezier curve example. Four control points define this Bezier curve. Points P_0^p and P_3^p are the ends of the curve, and points P_1^p and P_2^p determine the shape of the parent curve. Point P_0^c are the points of the child curve.

Bezier curve drawing

The widths for each Bezier curve are given by

$$W(k) = \begin{cases} \omega \cdot \max(0.3, 1 - 10 \cdot \mathcal{N}(k, \mu, \sigma^2)) & \text{if stenosis (+)} \\ \omega & \text{otherwise,} \end{cases} \quad (8)$$

where $\mathcal{N}(t, \mu, \sigma^2)$ is a normal random variable with mean μ and variance σ^2 for a real number $k \in [0, ts]$ and ω is a basic width (given in pixels). Notice that if a curve, ergo, a blood vessel is constructed with a stenosis region, its width is affected by a stenosis factor. Therefore, μ controls the position of the stenosis region center and σ^2 the stenosis' length, as shown in Fig. 4.

During the process involved in the image acquisition and transmission through analogy circuitry in conventional X-ray techniques, the Additive White Gaussian Noise (AWGN) is the most prevalent. Moreover, X-ray imaging systems manifest blur that reduces the sharpness of image edges and the overall contrast. The image can also exhibit the effects of Poisson, Impulsive, and Quantization noises. However, these are rare occurrences related to faulty device manufacturing [5,40,41]. Therefore, AWGN and Gaussian blur are applied to the generated image to simulate the image acquisition process. Algorithm 1 summarizes the proposed generative framework.

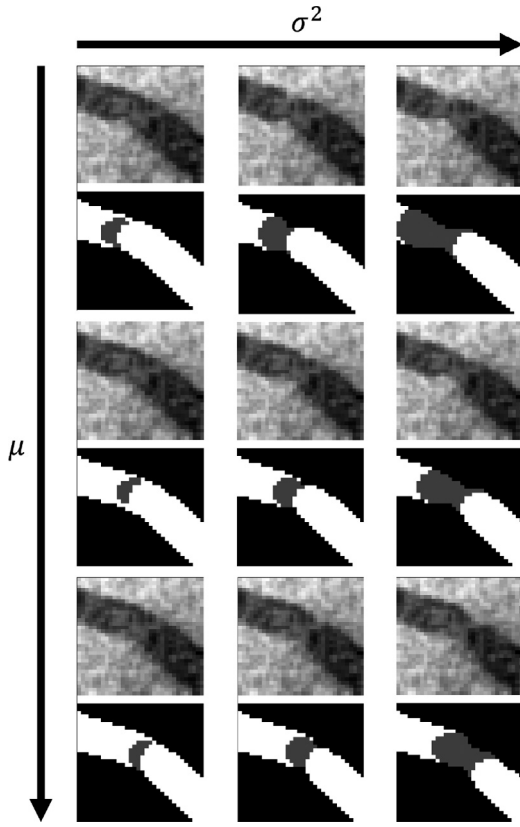


Fig. 4. Parameters μ and σ^2 affect the stenosis position and length. From top to bottom: $\mu = 40, 50, 60$ and from left to right: $\sigma^2 = 2, 4, 8$. The generated image is shown with the corresponding segmentation and stenosis location ground-truth.

Algorithm 1: Hierarchical Bezier-based Generative Patch Model.

Data: Patch size (w, h), number of patches N , background limit intensities (a, b), parent vessel basic width ω , number of child's vessels C , stenosis case (True or False), stenosis position μ , stenosis length σ .

Result: Artificial XCA dataset

```

1 for  $n \leftarrow 0$  to  $N$  do
2   Create a canvas  $DA$  as (1);
3   Draw the gradient background ROI such as (3);
4   Generate a parent Bezier curve  $B^{(p)}$  following (5) with
   control points given by (6);
5   Draw over the ROI the curve  $B^{(p)}$  with a width given by
   (8);
6   for  $c \leftarrow 0$  to  $C$  do
7     Generate each child Bezier curve  $B_c^{(c)}$  following subject
     to the control point  $P_0^{(c)}$  lie in  $B^{(p)}$ ;
8     Draw over the ROI the curve  $B_c^{(c)}$  with a width given
     by (8);
9   end
10  Add white noise in the ROI image;
11  Add Gaussian Blur in the ROI image;
12 end

```

2.2. Generative model performance measure

The objective of a generative model is to produce data that matches the observed (real) data. So, some widespread used metrics in GANs [42] can be exploited to measure the distance between the probability distributions of observing real-world data

$\mathbb{P}_r(\cdot)$ and the generative model $\mathbb{P}_g(\cdot)$. The more relevant are the Average Log-likelihood [14], the Wasserstein Critic (WC) [43], the Inception Score (IS) [44], and the Fréchet Inception Distance (FID) [36]. Each one has its drawbacks. For instance, Average Log-likelihood fall requires a vast number of samples to approximate the true log-likelihood. It also fails when the data dimensionality is high. The Wasserstein Critic distance is not a smooth function, requiring high processing time to be computed. The Inception Score is an asymmetric measure, only considers \mathbb{P}_g and ignores \mathbb{P}_r . The Fréchet Inception Distance assumes that features are of Gaussian distribution, which is often not guaranteed; however, it performs well in terms of discriminability, robustness, and computational efficiency [45–47]. Therefore, the FID is selected as a metric for the generative model, given that CNN conducts the classification process.

The FID requires a feature function ϕ (by default, the activation is from the penultimate pooling layer of a pre-trained Inception-v3 model) to evaluate the similarity of real data and generated data. However, ϕ can use any pre-trained model. In this work, ϕ is the feature vector obtained by applying a global average pooling over the third residual block of the ResNet18 (pre-trained on the ImageNet dataset). Therefore, an image is embedded into a vector with 256 features. These output vectors are summarized as a continuous multivariate Gaussian; the mean and covariance are estimated for the real XCA and the generated datasets. Thus, the FID is given by

$$\text{FID}(\mathbb{P}_r, \mathbb{P}_g) = \|\mu_{\phi_r} - \mu_{\phi_g}\|_2^2 + \text{Tr} \left(\sum_{\phi_r} + \sum_{\phi_g} - 2 \left(\sum_{\phi_r} \sum_{\phi_g} \right)^{1/2} \right), \quad (9)$$

where ϕ_r and ϕ_g are the embedding feature vectors of the real and artificial images, with their respective means μ_{ϕ_r} , μ_{ϕ_g} and empirical covariance matrices Σ_{ϕ_r} , Σ_{ϕ_g} . Tr is the trace of the matrix. Accordingly, a lower FID indicates a better-looking image patch; conversely, a higher score indicates a poor-looking artificial patch.

2.3. Convolutional neural network configuration

In this work, the family of residual networks was selected to find a good trade-off between the depth and performance of the network when small medical imaging datasets are evaluated. In particular, the residual network for the ImageNet Large Scale Visual Recognition Challenge (of depth 18, 34, and 50) was used. These network architectures are composed of one (plain) convolutional block followed by four residual blocks. Mathematically, a residual block can be formulated as follows

$$\mathbf{y} = \mathcal{F}(\mathbf{x}) + \mathbf{x}, \quad (10)$$

where \mathbf{x} and \mathbf{y} are the input and output feature maps, respectively. $\mathcal{F}(\mathbf{x})$ is the residual mapping to be learned. Residual blocks prevent the gradients from vanishing or exploding in deep neural network architectures. The first convolution block consists of one convolutional layer with 64 kernels of size 7×7 with a stride of two. On the other hand, each residual block contains convolution layers with 3×3 filters. As summarized in Table 1, the number of filters applied depends on the network architecture. Moreover, after each convolution layer, a Batch Normalization (BN) [48] and a Rectified Linear Unit (ReLU) layer [49] are set. Besides, Table 1 shows that from a 32×32 XCA image patch, the output feature map from the third residual block is 2×2 , and the kernels from the last residual blocks are larger than this feature map. For such a reason, this last residual block is omitted for the stenosis detection task.

Furthermore, the first two residual blocks include a CBAM with 3×3 filters in the spatial attention sub-module. Thus, the atten-

Table 1

Architectural configuration for different state-of-the-art ResNet (for ImageNet classification) with an input XCA image patch size of 32×32 .

Layer	Feature Map Size	Feature Map depth		
		18-layer	34-layer	50-layer
Conv1	16×16	64	64	64
ResBlock1	8×8	64	64	256
ResBlock2	4×4	128	128	512
ResBlock3	2×2	256	256	1024
ResBlock4	1×1	512	512	2048

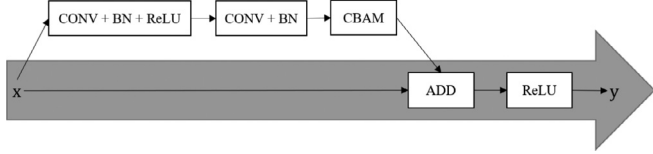


Fig. 5. The overall CBAM added into the first two residual blocks. Hence, the CBAM was set after the second convolutional layer.

tion module is set after the second convolution layer, as shown in Fig. 5.

The Convolutional Block Attention Module consists of two sub-attention modules in parallel: spatial attention and a channel attention module. Given an intermediate feature map, $\mathbf{F} \in \mathbb{R}^{C \times H \times W}$ as input, a refined feature map \mathbf{F}' (with the same dimension as the input) is computed such that

$$\mathbf{F}' = \mathbf{M}_c(\mathbf{F}) \otimes \mathbf{F}, \quad (11)$$

$$\mathbf{F}'' = \mathbf{M}_s(\mathbf{F}') \otimes \mathbf{F}', \quad (12)$$

where $\mathbf{M}_c \in \mathbb{R}^{C \times 1 \times 1}$ infers a 1D channel attention map \mathbf{F}' and $\mathbf{M}_s \in \mathbb{R}^{1 \times H \times W}$ a 2D spatial attention map \mathbf{F}'' . During the computation of the Hadamard product (\otimes), the channel attention values are broadcasted (copied) along the spatial dimensions and vice-versa. The channel attention uses both max-pooling (MaxPool) and average-pooling (AvgPool) outputs of given intermediate feature map, with a shared network MLP. Hence, the channel attention is computed as

$$\mathbf{M}_c(\mathbf{F}) = \sigma(\text{MLP}(\text{AvgPool}(\mathbf{F})) + \text{MLP}(\text{MaxPool}(\mathbf{F}))), \quad (13)$$

where σ denotes the sigmoid function. Similarly, the spatial attention uses the outputs of the max-pooling and the average-pooling along the channel axis and forwards them to a convolution layer such that

$$\mathbf{M}_s(\mathbf{F}') = \sigma(\text{Conv}(\text{AvgPool}(\mathbf{F}'), \text{MaxPool}(\mathbf{F}'))), \quad (14)$$

where σ denotes the sigmoid function, and Conv represents a convolution operation. In this case, 3×3 filters were employed in the spatial attention sub-module. Fig. 6 shows the overall structure of a CBAM.

2.4. Evaluation metrics

Five binary evaluation metrics are computed for comparison purposes. Specifically, Accuracy (ACC), Sensitivity (Se), Specificity (Sp), Precision (Pr), and F1-score (F1). These evaluation metrics are defined as follows

$$\text{ACC} = \frac{TP + TN}{N}, \quad (15)$$

$$\text{Se} = \frac{TP}{TP + FN}, \quad (16)$$

Table 2

Real and synthetic datasets description.

Dataset	# Images	Description
Real	250	32×32 patches of real XCA images
XCA [17]		50–50% stenosis and non-stenosis cases.
BGM [17]	10k	32×32 patches of synthetic images created by the Baseline Generative Model
		50–50% stenosis and non-stenosis cases.
HGBM (Proposed)	10k	32×32 patches of synthetic images created by the Hierarchical Bezier Generative Model
		50–50% stenosis and non-stenosis cases.

$$\text{Sp} = \frac{TN}{TN + FP}, \quad (17)$$

$$\text{Pr} = \frac{TP}{TP + FP}, \quad (18)$$

$$\text{F1} = 2 \frac{\text{Pr} \cdot \text{Se}}{\text{Pr} + \text{Se}}, \quad (19)$$

where N denotes the total number of test images, and TP , TN , FP , and FN denote true positive, true negative, false positive, and false negative. In general, the accuracy specifies the ratio of correct predictions within the total number of XCA test images. Sensitivity denotes how well stenosis cases are correctly identified, while specificity measures correctly classifying non-stenosis samples. Precision measures the ratio between the predicted stenosis cases and the total number of positive predicted patterns. F1-Score is the harmonic mean between precision and sensitivity. This metric is proposed as an alternative to solve the trade-off between specificity and sensitivity. It would be ideal for achieving high scores on the five indicators. A higher value in a range of $[0, 1]$ represents a better prediction.

3. Results

This study presents an automatic detection approach of stenosis in coronary angiograms. Moreover, the HBGM for generating artificial image patches is presented to improve the detection performance. All computation was performed on a Cloud-Platform with an Intel (R) Xeon (R), 12 GB of RAM, a 2.00 GHz dual processor, and a NVIDIA Tesla P100 PCIe 16 GB. The algorithms were implemented in Pytorch 1.8 library and Python 3.7.10. The source code is available at <https://github.com/eovallemagallanes/HBGM-Stenosis>.

3.1. Dataset

The CNNs configurations were evaluated with the stenosis detection dataset introduced by Antczak and Liberadzki [17], available at the authors GitHub site [50]. The dataset consisted of 250 XCA patches of 32×32 pixels each, with half of them representing stenosis cases.

Moreover, the authors presented an artificial dataset created by drawing random Bezier curves, thenceforth Baseline Generative Model (BGM). The dataset consists of 10k images of 32×32 to perform an earlier training stage to improve the stenosis detection performance. Additionally, the proposed HBGM created an artificial dataset with 10k images, including 50–50% cases of non-stenosis and stenosis, respectively. Table 2 summarize information about the different datasets.

Fig. 7 shows a sample of the real and artificial images. Hence, the baseline generative model creates blood vessel intersections but not bifurcations, as the proposed generative model. Moreover, each parent blood vessel can contain c-bifurcations, being able to

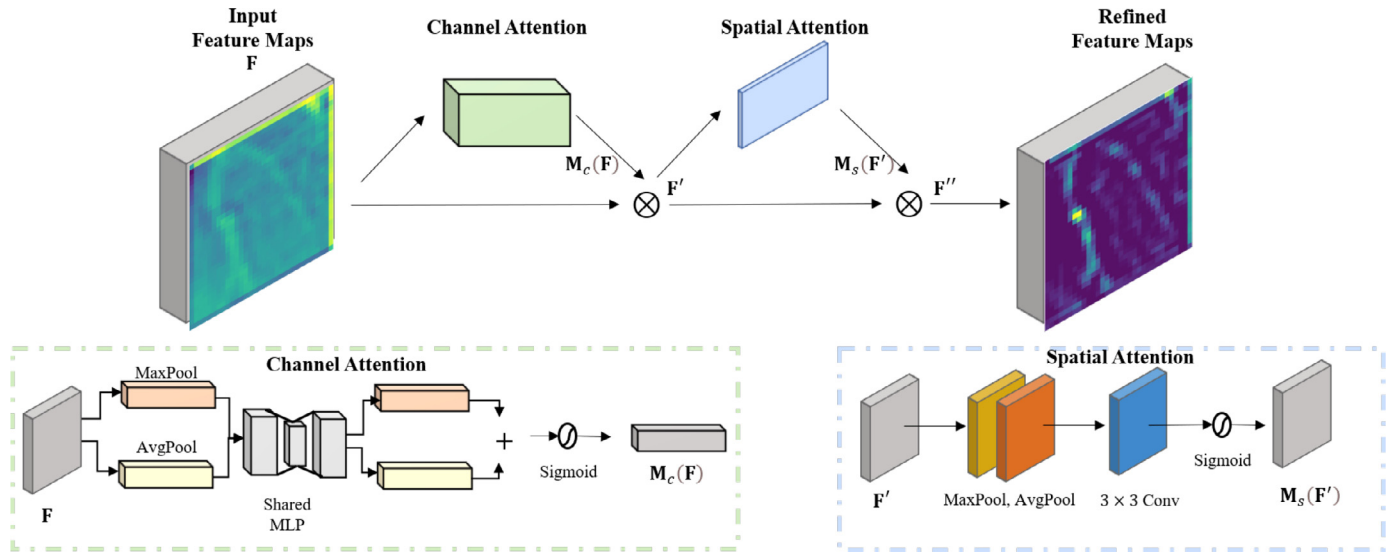


Fig. 6. The Convolutional Block Attention Module (CBAM). Each input feature map F is refined with a Channel Attention and a Spatial Attention sub-module.

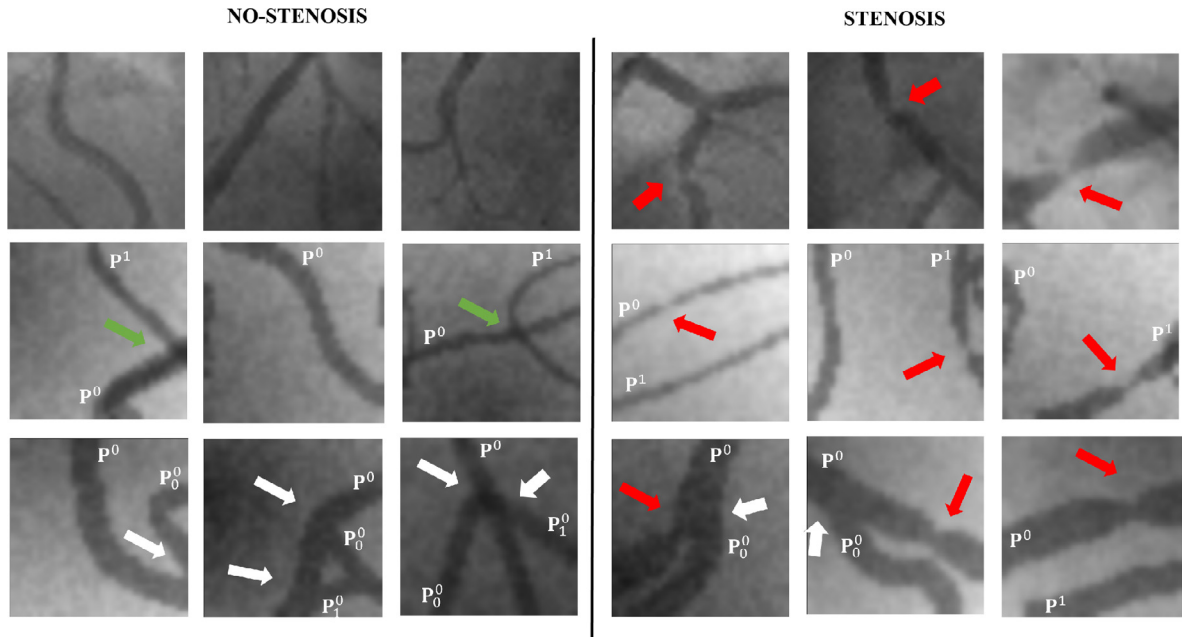


Fig. 7. Generated XCA image patches with non-stenosis and stenosis. First row: real patches, second row: Baseline generative model patches, and third-row: proposed generative model patches. The proposed generative model creates patches with blood vessel bifurcations, with a more clearly stenosis area and blood vessels with different weights. The red arrows indicate the stenosis case; the white ones, the bifurcation points, and the green arrows, the intersections of two-parent curves. Besides, P^i represents the i -th parent curve and P_j^i the j -th child curve of the i -th parent, respectively.

model different stenosis percentages and blood vessel widths (see the third row in Fig. 7).

The HBGM is governed by a set of free parameters, summarized in Table 3.

As expected, the FID retrieves a qualitative similarity measure between real and artificial images. Accordingly, in evaluating each image of the BGM [17] is obtained a mean FID of 92.7967, a minimum and maximum score of 57.4544 and 376.6007, respectively. Fig. 8a shows the distribution of the similarity scores given by the FID, showing a lower and upper quartile of 73.8502 and 94.0127, respectively. Also, 1159 of 10,000 samples have high FID values, indicating images with a higher dissimilarity, which are shown as upper outliers in the box-plot.

Further, the proposed HBGM reached a mean FID of 84.0886, a minimum and maximum score of 63.7430 and 97.4983, respectively. Fig. 8b shows the corresponding distribution of the FID, with

a lower and upper quartile of 80.2559 and 88.6376, respectively. Therefore, the proposed HBGM obtains a lower average FID, indicating that more realistic visual images were generated.

Table 3

Generative model parameters. A total of 10k images was generated, 50% with stenosis cases and the remainder with non-stenosis.

Parameter	Description	Value(s)
(w, h)	Patch size	(32×32)
a	Gradient background lower intensity limit	$rand(25, 50)$
b	Gradient background upper-intensity limit	$rand(85, 105)$
ts	Number of points for each curve	100
ω	Basic width in pixels	$rand(1, 4)$
P	Number of parent's curves	$rand(1, 3)$
C	Number of child's curves	$rand(0, 3)$
μ	Stenosis position	$rand(\frac{1}{4}ts, \frac{3}{4}ts)$
σ	Stenosis length	$rand(4, 8)$

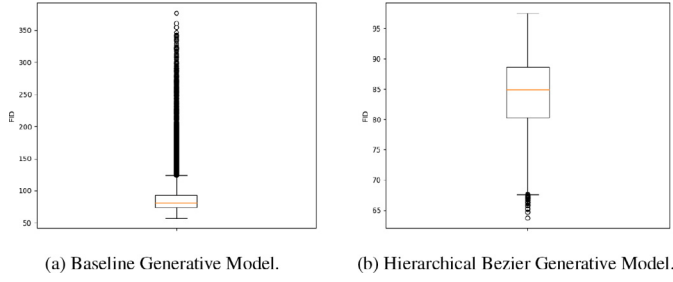


Fig. 8. Fréchet Inception Distance (FID) distributions. (a) Baseline and (b) Proposed Generative Models.

3.2. Training step

Since it is a binary classification problem, *e.g.*, stenosis or non-stenosis, the typical loss function is the binary cross-entropy defined as follows

$$L(y, \hat{y}) = -\frac{1}{N} \sum_{n=1}^N (y_n \log(\hat{y}) + (1 - y_n) \log(1 - \hat{y}_n)), \quad (20)$$

where y is the ground-truth label of a given input that belongs to 0 or 1 (non-stenosis or stenosis), \hat{y} is the predicted probability of the image being a stenosis case for all N images. The binary cross-entropy loss function is minimized by employing the Stochastic Gradient Descent with Momentum (SGDM) optimizer [51], which is mathematically defined as

$$\begin{aligned} v_t &= \gamma v_{t-1} + \eta \nabla_{\theta} L(\theta) \\ \theta &= \theta - v_t, \end{aligned} \quad (21)$$

where $\nabla_{\theta} L(\theta)$ is the gradient of the loss function with respect to the weights θ , v_t is an update vector at epoch t , η is the learning rate, and γ is the momentum.

The evaluated networks were trained by using four different strategies:

1. Training from scratch: only the real XCA dataset was used to optimize a random weight initialized network.
2. Pre-training on ImageNet: a pre-trained network on the ImageNet dataset was fine-tuned using only the real XCA dataset.
3. Fine-tuning on BGM: the network was previously trained by employing the baseline of the artificial dataset and then fine-tuned with the real XCA images.
4. Fine-tuning on HBGM: the network was previously trained by employing the proposed artificial dataset and then fine-tuned with the real XCA images.

Notice that the ImageNet initialization only applies for ResNets without attention modules. Also, the plain CNN was trained from scratch and pre-trained with both generative models datasets. The pre-training step using the artificial datasets was carried out during 50 epochs using an initial learning rate $\eta = e^{-2}$. Subsequently, the fine-tuning step goes through 200 epochs with a learning rate $\eta = e^{-3}$. Both steps included a fixed momentum $\gamma = 0.8$; a learning rate policy is implemented if the validation loss is not decreasing during 20 training epochs, $\eta = \delta \cdot \eta$ for $\delta = 0.1$.

Moreover, if the network was trained from scratch, it follows the same hyperparameters as the fine-tuning process. A k -fold Cross-Validation was used to evaluate the model performance. As such, a subset of 125 images of the real XCA dataset was randomly selected for testing and the remaining to perform the cross-validation procedure. A 5-fold cross-validation was carried out, selecting 100 images for training and 25 for validation in each fold.

3.3. Stenosis detection results

Table 4 summarizes the stenosis detection results, highlighting more outstanding metrics. The comparative analysis proved the proposed generative model efficacy as a pre-trained dataset.

In the first evaluated network, the plain CNN, a significant boost was obtained in the five-evaluation metrics by pre-training the network with the HBGM. This configuration reaches an accuracy, precision, sensitivity, F1-score, and specificity of 0.7295, 0.6854, 0.8169, 0.7449, 0.6476, respectively. In the case of the ResNets without attention module and trained from scratch, the ResNet34 surpassed the other models, obtaining a 0.8164, 0.7977, 0.8441, 0.8173, and 0.7905 in accuracy, precision, sensitivity, F1-score, and specificity, respectively. When an ImageNet pre-training was employed, the ResNet18 stood out by achieving an accuracy of 0.8574, precision of 0.8420, 0.8780 on sensitivity, 0.8567 on F1-score, and specificity of 0.8381. By pre-training the ResNets by the BGM and the HBGM, the Resnet18 + HBGM obtained the best results, outperforming the ImageNet initialization in accuracy (0.8672), precision (0.8822), F1-score (0.8587), and specificity (0.8952).

When the CBAM was trained from scratch, the ResNet18 obtained a boost in precision (0.8972) and F1-score (0.9175) concerning the ResNet18 with ImageNet initialization. Now, if the networks with attention module were previously pre-trained employing the artificial datasets, the ResNet34 + HBGM reached the best overall results in four of the five evaluation metrics: accuracy, precision, F1-score, and specificity of 0.8934, 0.9031, 0.8889, and 0.9111, respectively.

3.4. Feature maps visualization

Gradient weighted Class Activation Map (Grad-CAM) [52] uses the gradients of a given class, back-propagating the information into a particular convolutional layer (in most of the cases into the last convolutional layer). This backpropagation procedure generates a coarse localization map highlighting the most important regions in the image for predicting the class. Fig. 9 shows the Grad-CAM visualization concerning the different ResNet34 configurations for the stenosis class for representative test images, presenting intensity variations (6th row), background artifacts (2nd and 3rd rows), thin blood vessels (1st and 5th rows), and blood vessel bifurcations (4th and 5th row). The predicted probability of stenosis is shown below each Grad-CAM image. A reasonable prediction explanation produces discriminative visualizations over blood vessel regions. Thus, discriminative regions were highlighted in red, and those with lower contributions in purple.

The first row is the raw XCA images; the next four rows correspond to the ResNet34 without attention. Finally, the last three rows illustrate the outcome of the ResNet34 with CBAM. It can be seen that when the network was trained from scratch or pre-trained with ImageNet (2nd and 3rd row), the positive cases of stenosis showed similar GradCAM responses. However, when a pre-training step with synthetic datasets (4th and 5th rows), blood vessel bifurcations and stenosis had high attention areas (red tones) and medium attention (yellow tones) over blood vessel pixels. In the case of including CBAM, more refined GradCAM areas were obtained when the HBGM was employed. Showing high attention in regions over stenosis areas or their surroundings. All the configurations were sensitive to background artifacts and very thin blood vessels. Notice that if the probability in the first three images is higher than 0.5, a false positive is obtained. A probability below 0.5 represents a false negative in the last three images.

3.5. Full XCA image test

Additionally, from a full XCA image, a patch-based classification was carried on, classifying each patch as stenosis or non-

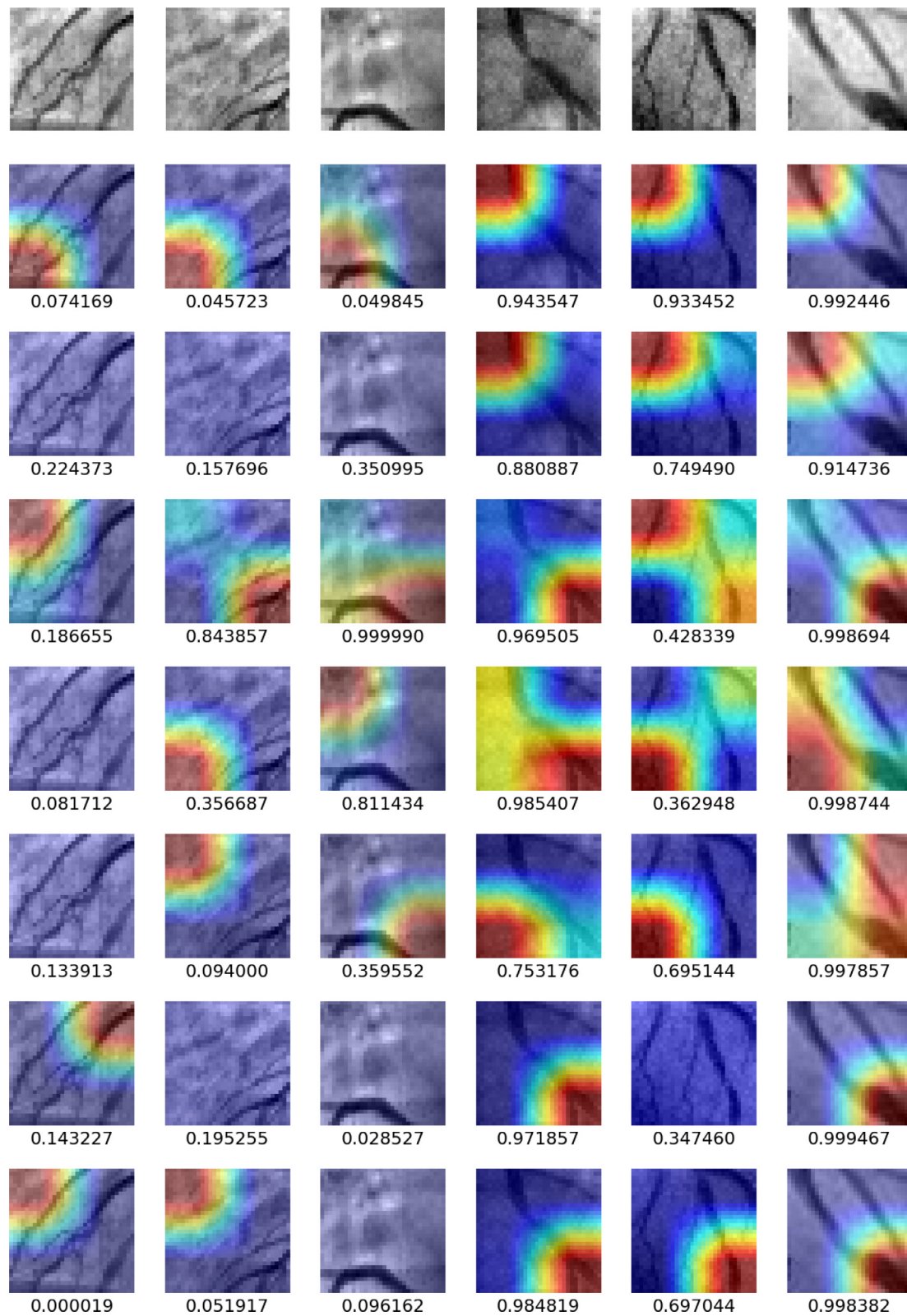


Fig. 9. Visual prediction explanation for three cases of non-stenosis and stenosis, from left to right. Different ResNet34 network configuration are shown: trained from scratch, ImageNet, BGM, HBGM. Then the ResNet34 + CBAM are shown: trained from scratch, BGM, HBGM. The first row are the raw XCA images. Below each image, the probability of being stenosis is reported. High attention areas are highlighted in red, while low attention regions are highlighted in purple tones.

Table 4

Stenosis detection results. Each network was evaluated under three scenarios: random weights initialization (N/A), pre-trained with the baseline generative model (BGM), and pre-trained with the proposed generative model (HBGM). Furthermore, the ResNet networks were fine-tuned, using the ImageNet weights as an additional comparison.

Architecture	Pre-training	ACC	Pr	Se	F1	Sp
Plain	N/A	0.5557 ± 0.0646	0.3524 ± 0.2975	0.4678 ± 0.4191	0.3807 ± 0.3129	0.6381 ± 0.3946
CNN	BGM	0.5492 ± 0.0226	0.5154 ± 0.0186	0.8102 ± 0.3035	0.6001 ± 0.1573	0.3048 ± 0.2450
[17]	HBGM	0.7295 ± 0.0172	0.6854 ± 0.0195	0.8169 ± 0.0329	0.7449 ± 0.0155	0.6476 ± 0.0381
ResNet18	N/A	0.8164 ± 0.0390	0.8339 ± 0.0782	0.7966 ± 0.0959	0.8070 ± 0.0393	0.8349 ± 0.1190
	ImageNet	0.8574 ± 0.0369	0.8420 ± 0.0618	0.8780 ± 0.0571	0.8567 ± 0.0346	0.8381 ± 0.0842
	BGM	0.7656 ± 0.0177	0.7418 ± 0.0084	0.7898 ± 0.0395	0.7647 ± 0.0227	0.7429 ± 0.0063
	HBGM	0.8672 ± 0.0210	0.8822 ± 0.0165	0.8373 ± 0.0395	0.8587 ± 0.0250	0.8952 ± 0.0162
ResNet18 + CBAM	N/A	0.8098 ± 0.0286	0.8719 ± 0.0388	0.7153 ± 0.0776	0.7825 ± 0.0392	0.8984 ± 0.0409
	BGM	0.7885 ± 0.0159	0.8143 ± 0.0139	0.7288 ± 0.0284	0.7690 ± 0.0199	0.8444 ± 0.0119
	HBGM	0.8721 ± 0.0205	0.8538 ± 0.0252	0.8881 ± 0.0173	0.8705 ± 0.0201	0.8571 ± 0.0266
ResNet34	N/A	0.8164 ± 0.0436	0.7977 ± 0.0693	0.8441 ± 0.0540	0.8173 ± 0.0380	0.7905 ± 0.0922
	ImageNet	0.7672 ± 0.0520	0.7885 ± 0.0574	0.7153 ± 0.1116	0.7447 ± 0.0689	0.8159 ± 0.0775
	BGM	0.7738 ± 0.0241	0.7575 ± 0.0358	0.7864 ± 0.0314	0.7709 ± 0.0220	0.7619 ± 0.0471
	HBGM	0.8295 ± 0.0175	0.8496 ± 0.0086	0.7864 ± 0.0349	0.8165 ± 0.0216	0.8698 ± 0.0063
ResNet34 + CBAM	N/A	0.8361 ± 0.0274	0.8972 ± 0.0345	0.7492 ± 0.0706	0.8140 ± 0.0386	0.9105 ± 0.0354
	BGM	0.7131 ± 0.0073	0.7059 ± 0.0125	0.6983 ± 0.0198	0.7018 ± 0.0087	0.7270 ± 0.0211
	HBGM	0.8934 ± 0.0090	0.9031 ± 0.0226	0.8746 ± 0.0275	0.8880 ± 0.0100	0.9111 ± 0.0238
ResNet50	N/A	0.7131 ± 0.0742	0.7427 ± 0.1162	0.6746 ± 0.1910	0.6834 ± 0.1099	0.7492 ± 0.1774
	ImageNet	0.7607 ± 0.0629	0.7648 ± 0.1012	0.7559 ± 0.0938	0.7536 ± 0.0597	0.7651 ± 0.1219
	BGM	0.7770 ± 0.0251	0.8547 ± 0.0396	0.6542 ± 0.0732	0.7372 ± 0.0431	0.8921 ± 0.0407
	HBGM	0.7639 ± 0.0228	0.7692 ± 0.0244	0.7322 ± 0.0421	0.7496 ± 0.0271	0.7937 ± 0.0284
ResNet50 + CBAM	N/A	0.7934 ± 0.0487	0.7708 ± 0.0891	0.8542 ± 0.0913	0.8013 ± 0.0336	0.7365 ± 0.1544
	BGM	0.8361 ± 0.0137	0.8222 ± 0.0175	0.8441 ± 0.0225	0.8327 ± 0.0142	0.8286 ± 0.0211
	HBGM	0.8541 ± 0.0141	0.8366 ± 0.0090	0.8678 ± 0.0292	0.8517 ± 0.0161	0.8413 ± 0.0100

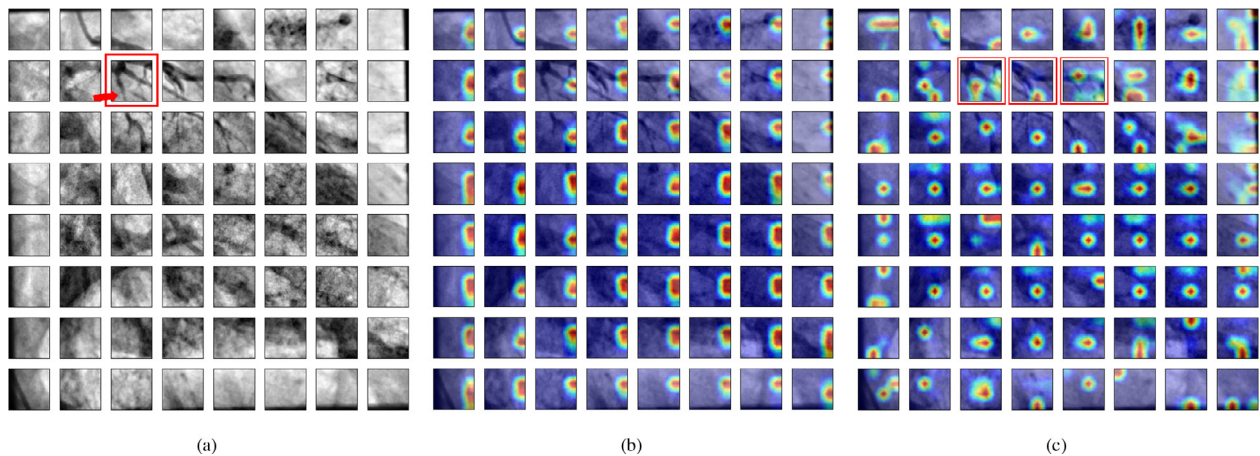


Fig. 10. (a) XCA image with a stenosis case, highlighted in a red box. (b) ResNet34 + ImageNet detection. (c) ResNet34 + CBAM + HBGM detection. In (b) and (c) the classified stenosis cases are highlighted in red boxes.

stenosis employing the ResNet34 pre-trained on the ImageNet and the ResNet34 + CBAM pre-trained on the HBGM. The results are shown in Fig. 10, where the highlighted patches in red illustrate the patches classified as stenosis cases. It can be seen that the XCA image tested on the ResNet34 pre-trained on the ImageNet is not available to detect the single case of stenosis. On the other hand, the one pre-trained on the HBGM successfully detects the stenosis case but presents two false-positive cases in patches surrounding the true positive case. Nevertheless, the GradCAM map particularly shows high attention in the stenosis region in the true positive case.

4. Discussion

Transfer learning from a CNN pre-trained on the ImageNet dataset has become a starting point from many medical imaging tasks, such as anomaly detection and image segmentation. As an alternative to transfer the weight from ImageNet, this work introduces a Hierarchical Bezier-based Generative Model to create an artificial dataset to pre-train a CNN to improve stenosis detection.

However, as the network goes deeper, pre-training on the ImageNet shows a performance degradation. As such, deeper models require a greater amount of data to be well-tuned to avoid overfitting and poor performance on the test data.

The achieved results employing the proposed generative model surpassed the ImageNet pre-training. It is noteworthy that the ImageNet dataset has 1,281,167 images that optimize the network for 600,000 iterations [21], which is 128 times bigger than the HBGM dataset and runs 12,000 times more epochs. Notice that the HBGM performs a pre-training step only for 50 epochs and fine-tuning for 200 epochs. By including CBAM, finer GradCAM maps are generated. Thus, higher attention over blood vessel pixels in bifurcation and stenosis areas can be seen.

One drawback of the proposed generative model is that the FID between synthetic and real data is computed after creating the dataset. This implies that the parameters that govern the model are not optimized, were handcrafted. The generative model is governed by a set of free parameters that control the background of the image, the number, and the width of the Bezier curve that simulate real coronary artery vessels. Furthermore, parameters also in-

clude settings, allowing control of the stenosis grade of a vessel. At this point, the FID between the generated images and the real ones cannot be controlled as a discriminator loss function in the GANs, such as the ones used for Chest X-ray [53], MR imaging [54], and Corneal Tomography [55]. For such a reason, it seems natural to explore GANs using additional prior information and hyperparameters such as the ones employed for the proposed generative model.

The limited amount of annotated XCA images restricts the classification process to a patch-based classification strategy. For such a reason, pre-processing or post-processing techniques that allow a classification refinement, such as an initial bounding box prediction [22,26,28] or redundant frames data extraction of the full XCA test [23,27], cannot be implemented in this work.

Additional real XCA training data is required to implement the current approach into the clinical praxis. Moreover, training data, including the stenosis severity, will be helpful to validate a stenosis quantification that is also commonly required for specialists. Nevertheless, the current classification method is robust enough to be the first tool to retrieve candidate stenosis regions given a full XCA image.

5. Conclusions

In this paper, a Hierarchical Bezier Generative Model has been proposed to address the problem of a small and poorly diversified database for stenosis detection in XCA images. A large-scale labeled dataset consisting of 10k images was created using the proposed approach. Extensive experiments showed that using this dataset to pre-train the CNNs and a posterior fine-tuning with real XCA image achieved the best overall performance on four (of five) evaluation metrics and competitive results on the remainder. Mainly, employing the ResNet34 + CBAM, an accuracy, precision, sensitivity, and F1-score of 0.8934, 0.9031, 0.8746, 0.8880, 0.9111, respectively, were achieved. Also, when incorporating CBAMs into the network, higher performance was reached than training from scratch. Moreover, it demonstrated the value of transferring the weights pre-trained using a more alike (artificial) dataset instead of the ImageNet dataset for stenosis detection tasks with only limited data available. One future extension of the proposed HGBM is to learn the parameters that govern the model in a deep learning way, such as classical GAN. The HGBM only requires a handful of free parameters compared to the millions of parameters that requires a traditional GAN.

Declaration of Competing Interest

The authors declare that they have no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Acknowledgment

This work was supported in part by the University of Guanajuato Grant NUA 147347, Convocatoria Institucional de Investigación Científica – CIIC 2022 (DAIP), and in part by the Mexican Council of Science and Technology CONACyT under Grant 626154/755609, and by the Mexican National Council of Science and Technology under project Cátedras-CONACyT No. 3150-3097.

References

- [1] L.S. Athanasiou, D.I. Fotiadis, L.K. Michalis, *Atherosclerotic Plaque Characterization Methods Based on Coronary Imaging*, Academic Press, New York, NY, USA, 2017.
- [2] National Heart, Lung, and Blood Institute, *Atherosclerosis*, 2021, (<https://www.nhlbi.nih.gov>).
- [3] World Health Organization, *Cardiovascular Diseases (CVDs)*, 2021, (<https://www.who.int/health-topics/cardiovascular-diseases>).
- [4] K.R. Nandalur, B.A. Dwamena, A.F. Choudhri, M.R. Nandalur, R.C. Carlos, Diagnostic performance of stress cardiac magnetic resonance imaging in the detection of coronary artery disease: a meta-analysis, *J. Am. Coll. Cardiol.* 50 (14) (2007) 1343–1353, doi:10.1016/j.jacc.2007.06.030.
- [5] E.N. Manson, V.A. Ampoh, E. Fiagbedzi, J.H. Amuasi, J.J. Flether, C. Schandorf, Image noise in radiography and tomography: causes, effects and reduction techniques, *Curr. Trends Clin. Med. Imaging* 2 (5) (2019) 555620, doi:10.19080/CTCMI.2019.03.555620.
- [6] C.-F. Chang, K.-H. Chang, C.-H. Lai, T.-H. Lin, T.-J. Liu, W.-L. Lee, C.-S. Su, Clinical outcomes of coronary artery bifurcation disease patients underwent culotte two-stent technique: a single center experience, *BMC Cardiovasc. Disord.* 19 (1) (2019) 1–8, doi:10.1186/s12872-019-1192-2.
- [7] T. Wan, H. Feng, C. Tong, D. Li, Z. Qin, Automated identification and grading of coronary artery stenoses with X-ray angiography, *Comput. Methods Programs Biomed.* 167 (2018) 13–22, doi:10.1016/j.cmpb.2018.10.013.
- [8] A.H.N. Kishore, V.E. Jayanthi, Automatic stenosis grading system for diagnosing coronary artery disease using coronary angiogram, *Int. J. Biomed. Eng. Technol.* 31 (3) (2019) 260–277, doi:10.1504/IJBET.2019.102974.
- [9] S. Sameh, M.A. Azim, A. AbdelRaouf, Narrowed coronary artery detection and classification using angiographic scans, in: 2017 12th International Conference on Computer Engineering and Systems (ICCSES), IEEE, Cairo, Egypt, 2017, pp. 73–79, doi:10.1109/ICCSES.2017.8275280.
- [10] A. Esteva, K. Chou, S. Yeung, N. Naik, A. Madani, A. Mottaghi, Y. Liu, E. Topol, J. Dean, R. Socher, Deep learning-enabled medical computer vision, *NPJ Digit. Med.* 4 (1) (2021) 1–9, doi:10.1038/s41746-020-00376-2.
- [11] L. Hou, D. Samaras, T.M. Kurc, Y. Gao, J.E. Davis, J.H. Saltz, Patch-based convolutional neural network for whole slide tissue image classification, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 2424–2433, doi:10.1109/CVPR.2016.266.
- [12] C. Shorten, T.M. Khoshgoftaar, A survey on image data augmentation for deep learning, *J. Big Data* 6 (1) (2019) 1–48, doi:10.1186/s40537-019-0197-0.
- [13] J. Yosinski, J. Clune, Y. Bengio, H. Lipson, How transferable are features in deep neural networks?, in: Z. Ghahramani, M. Welling, C. Cortes, N.D. Lawrence, K.Q. Weinberger (Eds.) *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8–13 2014, Montreal, Quebec, Canada, 2014*, pp. 3320–3328.
- [14] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A.C. Courville, Y. Bengio, Generative adversarial nets, in: Z. Ghahramani, M. Welling, C. Cortes, N.D. Lawrence, K.Q. Weinberger (Eds.) *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8–13 2014, Montreal, Quebec, Canada, 2014*, pp. 2672–2680.
- [15] S. Woo, J. Park, J.-Y. Lee, I.S. Kweon, CBAM: convolutional block attention module, in: V. Ferrari, M. Hebert, C. Sminchisescu, Y. Weiss (Eds.), *Computer Vision – ECCV 2018 – 15th European Conference, Munich, Germany, September 8–14, 2018, Proceedings, Part VII, Lecture Notes in Computer Science*, volume 11211, Springer, 2018, pp. 3–19, doi:10.1007/978-3-030-01234-2_1.
- [16] J. Hu, L. Shen, S. Albanie, G. Sun, E. Wu, Squeeze-and-excitation networks, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 42 (8) (2020) 2011–2023, doi:10.1109/TPAMI.2019.2913372.
- [17] K. Antczak, L. Liberadzki, Stenosis detection with deep convolutional neural networks, in: *MATEC Web of Conferences*, volume 210, EDP Sciences, 2018, p. 04001, doi:10.1051/mateconf/201821004001.
- [18] E. Ovalle-Magallanes, J.G. Avina-Cervantes, I. Cruz-Aceves, J. Ruiz-Pinales, Transfer learning for stenosis detection in X-ray coronary angiography, *Mathematics* 8 (9) (2020) 1510, doi:10.3390/math8091510.
- [19] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: Y. Bengio, Y. LeCun (Eds.), *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7–9, 2015, Conference Track Proceedings*, 2015, pp. 1–14. [arXiv:1409.1556](https://arxiv.org/abs/1409.1556).
- [20] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S.E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7–12, 2015*, IEEE Computer Society, 2015, pp. 1–9, doi:10.1109/CVPR.2015.7298594.
- [21] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27–30, 2016*, IEEE Computer Society, 2016, pp. 770–778, doi:10.1109/CVPR.2016.90.
- [22] B. Au, U. Shaham, S. Dhruva, G. Bours, E. Cristea, A. Coppi, F. Warner, S.-X. Li, H. Krumholz, Automated characterization of stenosis in invasive coronary angiography images with convolutional neural networks. [arXiv preprint arXiv:1807.10597](https://arxiv.org/abs/1807.10597).
- [23] C. Cong, Y. Kato, H.D. Vasconcellos, J. Lima, B.A. Venkatesh, Automated Stenosis Detection and Classification in X-ray Angiography Using Deep Neural Network, in: I. Yoo, J. Bi, X. Hu (Eds.), *2019 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2019, San Diego, CA, USA, November 18–21, 2019*, IEEE, 2019, pp. 1301–1308, doi:10.1109/BIBM47256.2019.8983033.
- [24] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, *Adv. Neural Inf. Process. Syst.* 25 (2012) 1097–1105.
- [25] W. Wu, J. Zhang, H. Xie, Y. Zhao, S. Zhang, L. Gu, Automatic detection of coronary artery stenosis by convolutional neural network with temporal constraint, *Comput. Biol. Med.* 118 (2020) 103657, doi:10.1016/j.combiomed.2020.103657.

- [26] K. Pang, D. Ai, H. Fang, J. Fan, H. Song, J. Yang, Stenosis-Detnet: sequence consistency-based stenosis detection for X-ray coronary angiography, *Comput. Med. Imaging Graph.* 89 (2021) 101900, doi:[10.1016/j.compmedimag.2021.101900](https://doi.org/10.1016/j.compmedimag.2021.101900).
- [27] J.H. Moon, W.C. Cha, M.J. Chung, K.-S. Lee, B.H. Cho, J.H. Choi, et al., Automatic stenosis recognition from coronary angiography using convolutional neural networks, *Comput. Methods Programs Biomed.* 198 (2021) 105819, doi:[10.1016/j.cmpb.2020.105819](https://doi.org/10.1016/j.cmpb.2020.105819).
- [28] V.V. Danilov, K.Y. Klyshnikov, O.M. Gerget, A.G. Kutikhin, V.I. Ganyukov, A.F. Frangi, E.A. Ovcharenko, Real-time coronary artery stenosis detection based on modern neural networks, *Sci. Rep.* 11 (1) (2021) 1–13, doi:[10.1038/s41598-021-87174-2](https://doi.org/10.1038/s41598-021-87174-2).
- [29] P.Y. Simard, D. Steinkraus, J.C. Platt, et al., Best practices for convolutional neural networks applied to visual document analysis, in: *Seventh International Conference on Document Analysis and Recognition (ICDAR)*, volume 3, 2003, pp. 958–963, doi:[10.1109/ICDAR.2003.1227801](https://doi.org/10.1109/ICDAR.2003.1227801), Edinburgh, UK.
- [30] J. Keelan, E.M.L. Chung, J.P. Hague, Simulated annealing approach to vascular structure with application to the coronary arteries, *R. Soc. Open Sci.* 3 (2) (2016) 150431, doi:[10.1098/rsos.150431](https://doi.org/10.1098/rsos.150431).
- [31] C. Jaquet, L. Najman, H. Talbot, L. Grady, M. Schaap, B. Spain, H.J. Kim, I. Vignon-Clementel, C.A. Taylor, Generation of patient-specific cardiac vascular networks: a hybrid image-based and synthetic geometric model, *IEEE Trans. Biomed. Eng.* 66 (4) (2018) 946–955, doi:[10.1109/TBME.2018.2865667](https://doi.org/10.1109/TBME.2018.2865667).
- [32] G. Tetteh, V. Efremov, N.D. Forkert, M. Schneider, J. Kirschke, B. Weber, C. Zimmer, M. Piraud, B.H. Menze, Deepvesselnet: vessel segmentation, centerline prediction, and bifurcation detection in 3-d angiographic volumes, *Front. Neurosci.* 14 (2020) 1285, doi:[10.3389/fnins.2020.592352](https://doi.org/10.3389/fnins.2020.592352).
- [33] I. Iakovou, N. Foin, A. Andreou, N. Viceconte, C. Di Mario, New strategies in the treatment of coronary bifurcations, *Herz* 36 (3) (2011) 198–213, doi:[10.1007/s00059-011-3459-y](https://doi.org/10.1007/s00059-011-3459-y).
- [34] A.P. Antoniadis, P. Mortier, G. Kassab, G. Dubini, N. Foin, Y. Murasato, A.A. Giannopoulos, S. Tu, K. Iwasaki, Y. Hikichi, et al., Biomechanical modeling to improve coronary artery bifurcation stenting: expert review document on techniques and clinical implementation, *Cardiovasc. Intervent.* 8 (10) (2015) 1281–1296, doi:[10.1016/j.jcin.2015.06.015](https://doi.org/10.1016/j.jcin.2015.06.015).
- [35] C. Chiastra, F. Iannaccone, M.J. Grundeken, F.J.H. Gijzen, P. Segers, M. De Beule, P.W. Serruys, J.J. Wykrzykowska, A.F.W. van der Steen, J.J. Wentzel, Coronary fractional flow reserve measurements of a stenosed side branch: a computational study investigating the influence of the bifurcation angle, *Biomed. Eng. Online* 15 (1) (2016) 1–16, doi:[10.1186/s12938-016-0211-0](https://doi.org/10.1186/s12938-016-0211-0).
- [36] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, S. Hochreiter, GANs trained by a two time-scale update rule converge to a local nash equilibrium, *arXiv preprint arXiv:1706.08500*.
- [37] S. Ayyachamy, V. Alex, M. Khened, G. Krishnamurthi, Medical image retrieval using Resnet-18, in: *Medical Imaging 2019: Imaging Informatics for Healthcare, Research, and Applications*, International Society for Optics and Photonics, San Diego, CA, USA, 16–21 February 2019, volume 10954, International Society for Optics and Photonics, 2019, p. 1095410, doi:[10.1117/12.2515588](https://doi.org/10.1117/12.2515588).
- [38] X. Yu, S.-H. Wang, Abnormality diagnosis in mammograms by transfer learning based on resnet18, *Fundam. Inform.* 168 (2–4) (2019) 219–230, doi:[10.3233/FI-2019-1829](https://doi.org/10.3233/FI-2019-1829).
- [39] D. Sarwinda, R.H. Paradisa, A. Bustamam, P. Anggia, Deep learning in image classification using residual network (resnet) variants for detection of colorectal cancer, *Proc. Comput. Sci.* 179 (2021) 423–431, doi:[10.1016/j.procs.2021.01.025](https://doi.org/10.1016/j.procs.2021.01.025).
- [40] K.A. Mohan, R.M. Panas, J.A. Cuadra, SABER: a systems approach to blur estimation and reduction in X-Ray imaging, *IEEE Trans. Image Process.* 29 (2020) 7751–7764, doi:[10.1109/TIP.2020.3006339](https://doi.org/10.1109/TIP.2020.3006339).
- [41] S. Lee, M.G. Kang, Poisson-Gaussian noise reduction for X-Ray images based on local linear minimum mean square error shrinkage in nonsubsampled contourlet transform domain, *IEEE Access* 9 (2021) 100637–100651, doi:[10.1109/ACCESS.2021.3097078](https://doi.org/10.1109/ACCESS.2021.3097078).
- [42] A. Borji, Pros and cons of GAN evaluation measures, *Comput. Vision Image Underst.* 179 (2019) 41–65, doi:[10.1016/j.cviu.2018.10.009](https://doi.org/10.1016/j.cviu.2018.10.009).
- [43] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein Generative Adversarial Networks, in: *International Conference on Machine Learning*, PMLR, Sydney, Australia, 2017, pp. 214–223.
- [44] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, Improved techniques for training GANs, *Adv. Neural Inf. Process. Syst.* 29 (2016) 2234–2242.
- [45] H. Zhang, I.J. Goodfellow, D.N. Metaxas, A. Odena, Self-attention generative adversarial networks, in: K. Chaudhuri, R. Salakhutdinov (Eds.), *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9–15 June 2019, Long Beach, California, USA, Proceedings of Machine Learning Research*, volume 97, PMLR, 2019, pp. 7354–7363.
- [46] P. Dimitrakopoulos, G. Sfikas, C. Nikou, Wind: Wasserstein inception distance for evaluating generative adversarial network performance, in: *2020 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2020, Barcelona, Spain, May 4–8, 2020*, IEEE, 2020, pp. 3182–3186, doi:[10.1109/ICASSP40776.2020.9053325](https://doi.org/10.1109/ICASSP40776.2020.9053325).
- [47] B. Liu, C. Tan, S. Li, J. He, H. Wang, A data augmentation method based on generative adversarial networks for grape leaf disease identification, *IEEE Access* 8 (2020) 102188–102198, doi:[10.1109/ACCESS.2020.2998839](https://doi.org/10.1109/ACCESS.2020.2998839).
- [48] S. Ioffe, C. Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift, in: F.R. Bach, D.M. Blei (Eds.), *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6–11 July 2015, JMLR Workshop and Conference Proceedings*, volume 37, JMLR.org, 2015, pp. 448–456.
- [49] V. Nair, G.E. Hinton, Rectified linear units improve restricted Boltzmann machines, in: J. Fürnkranz, T. Joachims (Eds.), *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, June 21–24, 2010, Haifa, Israel, Omnipress, 2010, pp. 807–814.
- [50] K. Antczak, L. Liberadzki, Deep Stenosis Detection Dataset, 2020, (<https://github.com/KarolAntczak/DeepStenosisDetection>).
- [51] N. Qian, On the momentum term in gradient descent learning algorithms, *Neural Netw.* 12 (1) (1999) 145–151, doi:[10.1016/S0893-6080\(98\)00116-6](https://doi.org/10.1016/S0893-6080(98)00116-6).
- [52] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: visual explanations from deep networks via gradient-based localization, in: *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22–29, 2017*, IEEE Computer Society, 2017, pp. 618–626, doi:[10.1109/ICCV.2017.74](https://doi.org/10.1109/ICCV.2017.74).
- [53] S. Kora Venu, S. Ravula, Evaluation of deep convolutional generative adversarial networks for data augmentation of chest X-ray images, *Fut. Internet* 13 (1) (2021) 8, doi:[10.3390/fi13010008](https://doi.org/10.3390/fi13010008).
- [54] K. Lei, M. Mardani, J.M. Pauly, S.S. Vasanawala, Wasserstein GANs for MR imaging: from paired to unpaired training, *IEEE Trans. Med. Imaging* 40 (1) (2021) 105–115, doi:[10.1109/TMI.2020.3022968](https://doi.org/10.1109/TMI.2020.3022968).
- [55] H. Abdelmotaal, A.A. Abdou, A.F. Omar, D.M. El-Sebaity, K. Abdelazeem, Pix2pix conditional generative adversarial networks for scheinplung camera color-coded corneal tomography image generation, *Transl. Vis. Sci. Technol.* 10 (7) (2021), doi:[10.1167/tvst.10.7.21](https://doi.org/10.1167/tvst.10.7.21).