Cover Page

1) Title of the paper:

Combining machine learning and artery characterization to identify the main bifurcations in 3D vascular trees

2) authors' affiliation and address:

RMeS lab, UFR Odontologie & LTeN, Polytech' Nantes
1rue Ch. Pauc, La Chantrerie, 44306 NANTES, France.

3) e_mail address:

Florent.Autrusseau@univ-nantes.fr

4) Conference & Publisher information:

SPIE Medical Imaging
https://spie.org/conferences-and-exhibitions/medical-imaging?SSO=1

5) bibtex entry:

```
@inproceedings{MedIm2022,
  author = {I. Essadik and A. Nouri and N. Lauzeral and R.
Touahni and R. Bourcier and F. Autrusseau},
  title = {Combining machine learning and artery
characterization to identify the main bifurcations in 3D
vascular trees},
  booktitle = {SPIE Medical Imaging},
  year = {2022}
}
```

Combining machine learning and artery characterization to identify the main bifurcations in 3D vascular trees

Ibtissam Essadik[a], Anass Nouri[a,b], Nathan Lauzeral[c], Raja Touahni[a], Romain Bourcier[c], and Florent Autrusseau[d]

[a]SETIME Laboratory, Information Processing and Artificial Intelligence Team, Faculty of Sciences, Ibn Tofail University, BP 133, 14000 Kenitra, Morocco
[b]ENSC, National School of Chemistry, Ibn Tofail University, BP 133, 14000 Kenitra, Morocco
[c]Institut du Thorax, UMR 1087, Univ. of Nantes, 8 Quai moncousu, 44007, Nantes, France
[d]RMeS lab & LTeN, Nantes, Univ. of Nantes, Polytech'Nantes, 44306, Nantes, France

ABSTRACT

The cerebral vascular system is constituted by all the arteries and veins irrigating the brain. This vascular tree starts from two pairs of arteries, the vertebral arteries and the internal carotid arteries. These latter divide into a circular shape being called the Circle of Willis (CoW). There is considerable variability in the structure of the CoW among patients. The CoW can host various vascular diseases, among which intracranial aneurysms are of particular importance because their occurrence, or more precisely their rupture, can be devastating. Intracranial aneurysms often occur at the bifurcations of the arterial tree (saccular aneurysms), as a bulge in the vessel wall. It is crucial to recognize and monitor such aneurysms. Anatomical identification of the bifurcations of the CoW can be of great help to establish a diagnosis or to plan a surgical operation. In this study, we propose an automatic solution to categorize the vascular anatomy of the CoW in 3D volumes by identifying its main constituting bifurcations. Our solution combines machine learning and a multivariate analysis (Linear Discriminant Analysis: LDA). The LDA works as a classifier and reduces the dimensionality of the dataset by transforming the selected features in a lower dimensional space. This work is a preliminary study prior to moving to human cerebrovascular images. We evaluate the proposed method using several machine learning techniques combined with a leave-one-out validation applied on a set of 30 synthetic vascular images as well as 30 mouse cerebral vasculatures.

Keywords: Bifurcation labeling, vascular tree, artery characterization, dimensionality reduction, machine learning.

1. INTRODUCTION

This work is part of a wide French research project[1] which aim is to analyze and understand the multiple factors that may lead to the manifestation of intacranial aneurysms located along the CoW within human brains. This project is articulated around several axes, such as the genetic study of aneurysm formation, the automatic detection of bifurcations and their geometrical analysis[2] in a cerebral vascular tree, the detection of aneurysms, and the automatic identification of the bifurcations of the CoW on which the risk of aneurysm formation (and rupture) is prevalent.

In this work, we present a new automatic method to identify the main bifurcations of interest of the CoW on 3D synthetic vasculatures and mice brain acquisitions. The main objective is to automatically map the bifurcations of a 3D vascular tree onto an anatomical atlas, not only based on their 3D coordinates, but, we intend to consider a full vasculature characterization. Indeed, we consider that a clustering method would inevitably be more efficient if performed on higher dimensions, *i.e.* instead of computing bifurcations clusters based solely of (x, y, z) coordinates in the 3D space, we plan to run a clustering method considering also the arteries' features such as diameters, cross-sections, tortuosities, etc. The LDA used in the pipeline of the proposed approach will help to reduce the features' dimensionality by discarding the non discriminant ones.
This preliminary study on synthetic and mice vasculatures is a first step towards the automatic identification of the main bifurcations constituting the CoW in human MRI acquisitions. We first evaluate our method on the simplest form a vascular tree may take, *i.e.* made of linear branches. The Vascusynth software* was used to provide such synthetic models. Such a layout is much simpler to process, it is composed of fewer characteristics (cross-sections or tortuosity

*https://vascusynth.cs.sfu.ca/Welcome.html

being de facto excluded). Next, in order to experiment the method on a more realistic dataset, we use mouse μ-CT scan acquisitions. Such images are more complex than the ones provided by the synthetic model and allow us to consider some extra features (the cross-sections and tortuosity) to identify the main bifurcations of the COW. Traditionally, registration techniques are used in order to compensate for the differences between various acquisitions, in order to build an anatomical atlas. In this work, we intend to avoid this registration step, as potential 3D realignment inaccuracies may bring some misclassifications.

The construction of a cerebral anatomical atlas of the human brain has already been studied in the past. Robben *et al.*[3] used graphical matching algorithms to identify and label human brain arteries. Bogunovic *et al.*[4] took advantage of the *a posteriori* maximum likelihood estimation to automatically label the CoW. In the work of Dunas *et al.*,[5] the authors constructed a probabilistic stereotactic atlas for automatic labeling of cerebral arteries. In a different approach, Ota *et al.*[6] proposed a method that constructs classifiers by the AdaBoost multi-class technique. Their method provides the anatomical names of the bronchial branches. Wang *et al.*[7] also took advantage of supervised machine learning algorithms to automatically identify and label cerebral arteries. Murat *et al.*[8] identified cerebral arteries using random forest and a Bayesian network representation of the vessels. In the paper of Zhao *et al.*,[9] a random forest was implemented based on vessel features for bifurcation detection in the Vascusynth database[10] and in a dataset of 3D clinical chest CT images. Finally, Ghanavati's approach[11] used the stochastic relaxation technique for the segmentation and automatic identification of mouse cerebrovascular structures. In contrast to the state of the art where the majority of researchers have worked on labeling the arteries constituting the CoW,[4] we propose a novel automated solution to categorize the bifurcations of the CoW (where aneurysms are most prevalent) in 3D mouse vasculatures, using machine learning classifiers.

Contrary to the vast majority of similar approaches from the literature, in this work, not only do we consider the geometric localization of the bifurcations, but we also take benefit of a full geometric characterization[2] of both the bifurcation and its mother artery. The rest of the paper is structured as follows. In section 2 the proposed approach is described. In section 3 the obtained results are presented and analyzed. Finally, we conclude and present some perspectives in section 4.

2. ATLAS CONSTRUCTION

In this work, the atlas construction follows five main phases: (1) Data segmentation (2) 3D skeleton construction and bifurcation detection and characterization, (3) Manual labeling of the bifurcations of interest, (4) Dimensionality reduction of the collected features, and (5) Automatic identification of the bifurcations using Machine Learning classifiers. Our automatic bifurcation labeling system is shown in detail in Figure 1. As we can see from Figure 1, during the training step, the LDA weights are gathered from a set of labelled vasculatures, whereas this set of weights is then applied "as is" on the test images.
To attest of the relevance of the proposed approach, we considered two types of images on which the training and test phases have been conducted : synthetic vascular tree images and μ-CT acquisitions of mouse vasculatures. We will see in the following that the proposed approach can be used to analyze different types of images as long as as they can be represented via a tree structure.

2.1 Data preprocessing

As a preprocessing step, most works[12] consider applying transformations to generate additional data for the available training set. In our case of bifurcations labeling, there is very little training data available, we thus use data augmentation[12] by warping the available training images. We have applied deformations on a 3D vascular image by following a two steps process. First, we have used affine transformations in which some simple deformations including translations, rotations and shearing were generated by applying affine displacement fields to the images of both datasets. Secondly, random elastic deformations[13,14] distortions were applied. This allowed to vastly expand the size of the training set.
As previously mentioned, two distinct datasets have been tested. The first one is based on a synthetic model and the second is composed of mouse brain vasculatures. The use of a synthetic arterial tree allows to have a full control on the process steps, indeed, using such a simplified configuration, we have full knowledge of both the bifurcations locations and the various characteristics of this modeled arterial network. For mouse acquisitions, one more processing step is needed. Effectively, we must first extract the arteries from the 3D vasculatures through a manual segmentation. For each μ-CT acquisition, a region of interest along the circle of Willis is delimited and extracted using the 3D Slicer software[†].

[†] https://www.slicer.org/

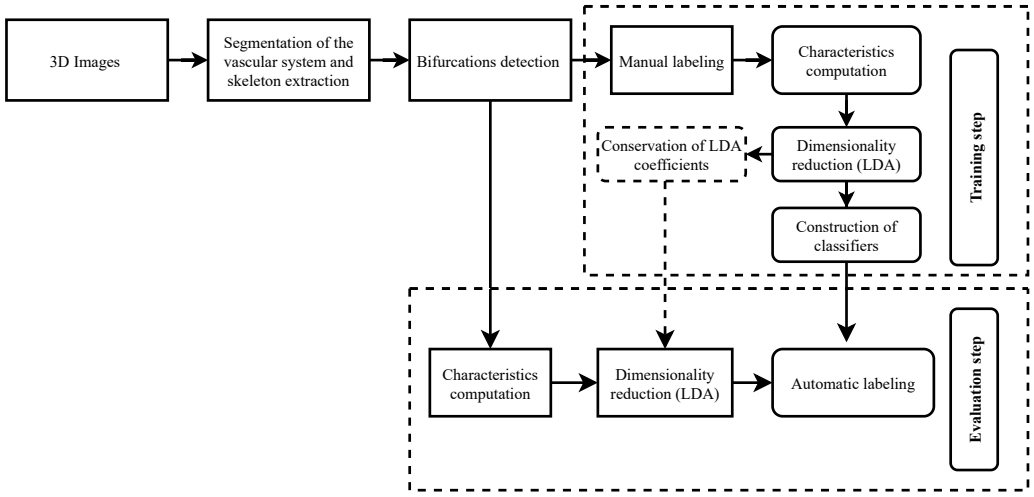


Figure 1. Pipeline of the proposed approach.

2.2 Bifurcation detection and features extraction

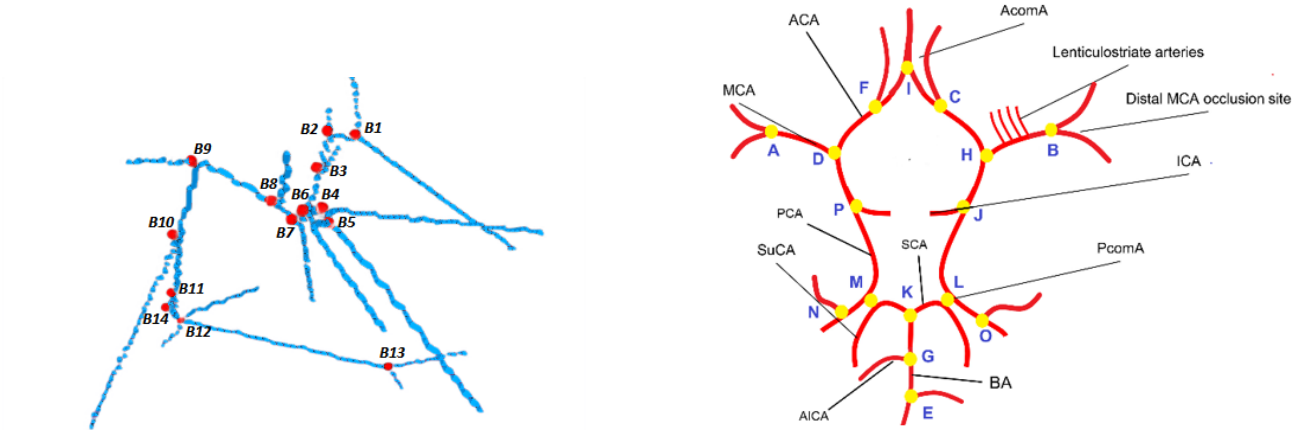
Once the region of interest on each image has been located and extracted (the CoW), we compute a 3D skeleton on which a non-directed graph is constructed. It allows, based on the work of Nouri *et al.*,[2] to browse the vasculature in 3D and hence accurately detect the 3D coordinates (x, y, z) of all the bifurcations belonging to the CoW. Afterwards, particular geometrical features that presumably have an impact on the development of intracranial aneurysms (such as the cross-section and diameters of a bifurcation branch, the angles between each pair of bifurcation branches, the mother branch tortuosity, the geodesic distance between two consecutive bifurcations along the tree) are automatically computed. However, in this work, we did not limit ourselves to the features presented in the work of Nouri *et al.*,[2] but we have enriched this set of features through several combinations which shall describe more precisely the target bifurcations in the CoW. This leads to a set of 40 features for each bifurcation of the mice CoW. For the synthetic vascular tree, 22 features were retained per bifurcation.

2.3 Manual labeling and bifurcations of interest selection

To generate an atlas using supervised machine learning algorithms, we require a training dataset of graphs built on the skeletons of the segmented vasculatures and ground truth representing the labels of the target bifurcations that have been manually labeled. In the framework of this study, we have selected the bifurcations presenting the highest risk of aneurysm formation. The involved arteries, on which those bifurcations are located, are: the basilar artery (BA), and the following ones (for both left and right localizations) anterior cerebral artery (ACA); internal carotid artery (ICA); middle cerebral artery (MCA); posterior communicating arteries ($PComA$); posterior cerebral arteries (PCA); Superior cerebellar artery ($SuCA$); anterior inferior cerebellar artery ($AICA$); superior cerebellar artery (SCA) and distal middle cerebral artery occlusion. It is important to note that not all bifurcations will necessarily be present in each vascular graph. This heterogeneity may be caused either by the patient's particular anatomy or by a flaw during the segmentation or skeletonization process. For the synthetic data, we have chosen 14 bifurcations that are labeled from 1 to 14 in Figure 2(a). Concerning the mouse acquisitions, we are primarily interested by 16 Points of Interest ($PoIs$) representing the centers of the labeled target bifurcations (points A through P in Figure 2(b)).

2.4 Dimensionality Reduction

In this Work, the LDA was used with a twofold purpose: *i)* reduce the data dimensionality and *ii)*, act as a classifier. The LDA is a supervised algorithm which aims to classify data for mutli-class problems while projecting them into a lower dimensional space, in such a way that similar data gets clustered together, while dissimilar data are spread further apart. In other terms, the LDA shall project the n-dimensional data onto a smaller dimension system while trying to achieve the best possible separation between the clusters. Consequently, the classification can be done directly by using the LDA algorithm or by applying another classification algorithm on the data mapped in a lower dimensional space. In this paper, each bifurcation of the CoW is represented by a vector of measured features: the bifurcation 3D coordinates, the arteries' diameters, their cross-sections, the angles, the tortuosity of each branch, the Euclidean distances (Vascusynth) separating the bifurcation from the previous one along the CoW, etc. All such features (along with various combinations) are used



(a) *Vascusynth* vasculature (b) Mouse vasculature

Figure 2. Schematic representation of the bifurcations of interest. (a) the 14 PoI labels (B1-B14) in a VascuSynth model of the vasculature and (b) a schematic representation of the mouse circle of Willis with the 16 PoI labels (A-P).

to discriminate the sets of bifurcations. Thus, we have fed the LDA with the dataset of labeled bifurcations along with their corresponding geometric features and furthermore, we have trained the LDA to generate a mapping between the initial feature space and the labeled bifurcations which represent the ground truth in the learning process.

3. EXPERIMENTAL RESULTS

Machine learning algorithms are mainly divided into four[15] categories: supervised learning, unsupervised learning, semi-supervised learning and reinforcement learning. In this work, the bifurcation labeling is considered as a multi-label classification problem. We have evaluated several classification methods in a low dimensional space in order to select the best performing one. Hence, for any new bifurcation, the identification follows a three steps process. First, anatomical features are extracted, then, the LDA mapping is applied to reduce the data into a lower dimensional representation, and finally, the identification is carried out using the best performing classification technique. Note that this approach allows to bypass any image registration stage, which is a common practice when creating anatomical atlases. Our method has been evaluated on two different datasets, VascuSynth (3D synthetic vascular tree) and μ-CT mice acquisitions. Both datasets contain 30 images of the vasculature with their segmentations and centerlines, as well as the ground truth labels indicated manually.

3.1 Synthetic vasculature

As previously explained, in this work, we will focus on both a synthetic vascular tree structure, and mice brain acquisitions. The vascusynth dataset we have been using is actually providing some extremely simplified arterial trees. Indeed, all branches are linear and exhibit a unitary diameter. Hence, as we will witness in the following, we are compelled to limit our measurements to only few characteristics.

3.1.1 VascuSynth dataset

The VascuSynth[10] library simulates 3D volumes ($100 \times 100 \times 100$ voxels) of vascular trees and generates the corresponding ground truth for segmentation and bifurcation location. Following the data augmentation step by applying elastic distortions and affine transformations, we have constructed a dataset of 30 images from a reference object containing 14 bifurcations of interest (Figure 2(a)). In this data, each individual bifurcation is characterized by several features : the 3D coordinates of its center (x_0, y_0, z_0); the coordinates of its extremities, i.e. of the next bifurcation along each branch (x_i, y_i, z_i), with $i = \{1, 2, 3\}$; the coordinates of the sum of three unitary direction vectors of the bifurcation branches (S_x, S_y, S_z); the angles between the three directional vectors of the bifurcation (θ_i) and Euclidean distances between the bifurcation center and all three extremities $\mathcal{D}_i$.

The vasculature characterisation method (proposed in previous works[2]) provides a list of features per bifurcation. Unfortunately, this method does not allow to automatically determine which of the three arteries is the mother artery. Indeed, the blood flow direction is unknown. Thus, depending on the bifurcation properties (its orientation mostly), the 3 arteries might not match from one 3D acquisition to the other, i.e. for one μ-CT scan, the mother branch might be

Table 1. Geometric properties of the bifurcations in 3D Vascusynth vasculatures.

Basic characteristics	Combinations
• (x_0, y_0, z_0) coordinates of centers • $(\max(x_i), \max(y_i), \max(z_i))$ maximum of extremities coordinates • $(\min(x_i), \min(y_i), \min(z_i))$ minimum of extremities coordinates • Directional sum vector (S_x, S_y, S_z) • Average, sum, min and max of distances $\boldsymbol{\mathcal{D}} = (\mathcal{D}_1, \mathcal{D}_3, \mathcal{D}_3)$ • Average, sum, min and max of Angles $\Theta = (\theta_1, \theta_2, \theta_3)$	• $\frac{\max(\Theta)}{\min(\Theta)}$ • $\frac{\max(\boldsymbol{\mathcal{D}})}{\min(\boldsymbol{\mathcal{D}})}$

labeled as #1 whereas it would be denoted as #2 on another scan. Hence, having no access to this information, makes the comparison of inter-bifurcation features impossible. As a consequence, we can only compare minimal-, maximal-, summed- or averaged-features from one bifurcation to the other.

Moreover, in addition to these basic features, we have also computed various combinations. Diameters, sections and tortuosities are not included for this dataset because, as previously explained, the Vascusynth images are composed of rectilinear skeleton branches. Table 1 depicts both the basic and combined features of the bifurcations for each synthetic vasculature. The proposed labeling method consists in reducing the number of characteristics calculated using the LDA algorithm. They are thus transformed into five LDA components made of various weighted combinations of these features.

3.1.2 Results

In order to quantify the importance of PoI features and the improvement brought by the LDA during the automatic labeling, we have tested three different approaches. First, our method was applied on the three dimension coordinates of the bifurcation centers without any additional features and the LDA was employed as a classifier ("Method 1"). For the second approach, we have added the complete feature vectors described in the previous section. These features were directly used as inputs for the classification methods in order to predict the labels without any dimensionality reduction ("Method 2"). Finally, in our last experiment, we have used the LDA to represent the bifurcations' features in a lower dimensional space, the classifiers were then trained in this new space ("Method 3"). The classification of PoIs was performed using different supervised machine learning algorithms and was evaluated by using a leave-one-out cross-validation on this set of 30 images (29 images used for training and 1 for testing, through 30 trials in total).

Table 2 and Figure 3 illustrate the performances of the machine learning algorithms for labeling prediction for each method on the Vascusynth database. We can notice that adding some extra geometric features improves the classification rate. Similarly, transforming and representing features in a reduced dimensional space with the LDA has a very clear impact on the bifurcation labeling of the Vascusynth vasculature. It appears that the LDA and the K-Nearest Neighbors (K-NN) algorithms present the best performances with a recognition rate reaching 98.5%.

Table 2. Accuracy [%] and F1-Scores [%] on the VascuSynth generated images

Classifiers	Method 1		Method 2		Method 3	
	Accuracy	F1-Score	Accuracy	F1-Score	Accuracy	F1-Score
SVM	87.4	85.5	96.0	95.5	**97.8**	**97.6**
Decision tree	80.8	78.3	87.1	84.6	**94.0**	**93.2**
Logistic regression	73.9	67.7	92.7	91.6	**94.4**	**93.4**
K-NN	85.6	83.5	95.2	94.5	**98.5**	**98.4**
Naive Bayes	75.7	70.9	93.7	93.1	**97.5**	**97.3**
Random forest	84.0	81.5	96.0	95.6	**96.9**	**96.7**
XGBoost	82.6	80.4	94.7	93.8	**96.7**	**96.1**
LDA	87.5	85.7	95.6	95.5	**98.5**	**98.5**

For each PoI, we report the accuracy, sensitivity and F-score of the classification. Figure 4 shows these metrics' performances for each PoI separately using the LDA classifier. In order to examine the misclassified PoIs , we show the confusion matrix of the automatic labeling for the VascuSynth dataset in Figure 5. The matrix's diagonal shows the probabilities of correctly classifying each PoI with its ground truth label. The recognition error for each bifurcation can also be seen in the confusion matrix, where we can observe the symmetrically located "confused" bifurcation points,

(a) Accuracy (b) F1-Score

Figure 3. Accuracy and F1-Scores on the VascuSynth generated images

e.g. B_7, B_{12} and B_{14}. Based on the confusion matrix, we can observe that the two PoIs B_7 and B_{14} suffer the most from misclassification, B_7 is occasionally detected as B_6 and B_{14} is sometimes incorrectly recognized as B_{12}. These misclassifications can be explained by the strong similarities between both bifurcations. Effectively, their constituting arteries frequently are located within the same neighborhood, and have very similar lengths. Strong similarities and a close proximity can be found between these pairs of bifurcations.

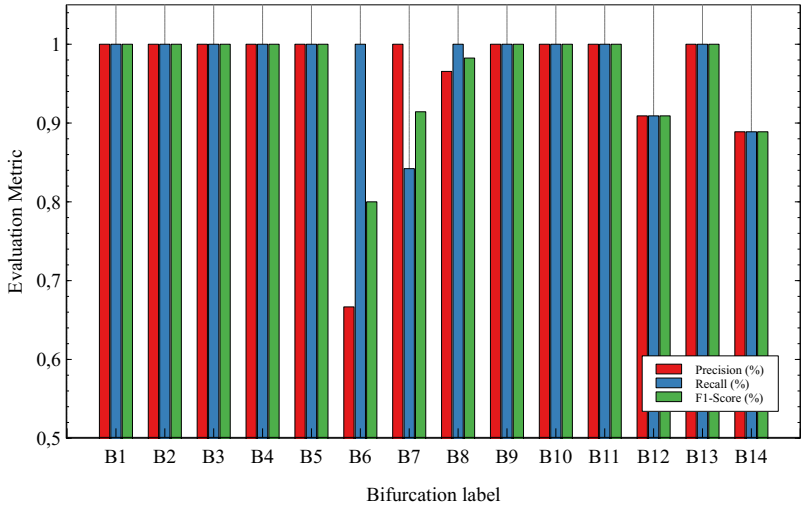


Figure 4. Detail of the classification performance of the 14 PoIs in Vascusynth Images with the LDA.

3.2 Mice cerebrovascular trees

As was observed in the previous section, the Vascusynth dataset, due to its simplified tree representation forced us to discard some important geometrical parameters during the analysis. Effectively, the arteries diameters and tortuosities could not be considered by the LDA; that is why in the following we have also considered mouse brain vasculatures in our analysis. Another very important reason to choose the mouse vasculatures is the very neat segmentation they can offer. Indeed, the mice brains went through a Barium Sulfate injection, allowing the μ-CT scan to acquire only the vascular tree (hence automatically discarding the grey/white matters). This allows us to avoid going through a tedious segmentation process as this is not the purpose of this study.

3.2.1 Mice Dataset

The Mice dataset contains 11 μ-CT acquisitions and thanks to image augmentation, we have generated 30 different distortions by applying both affine and elastic deformations. Similarly to the vascusynth dataset approach, here we

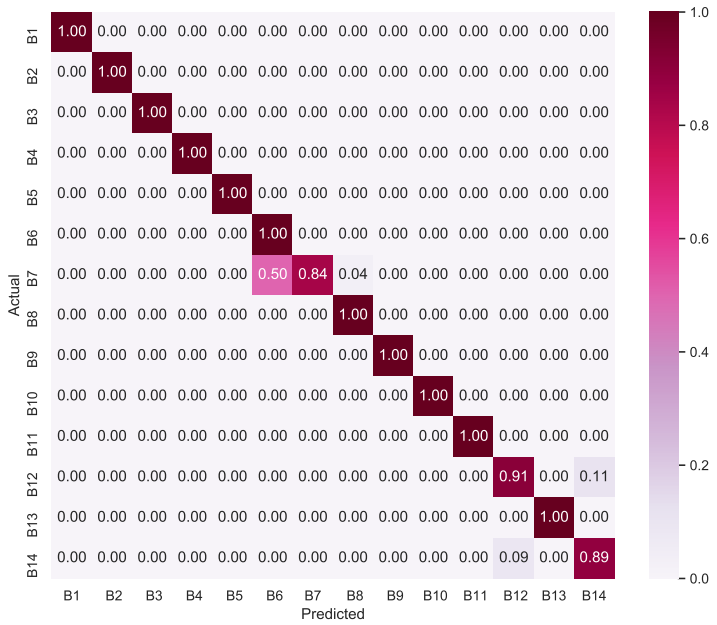


Figure 5. Confusion matrix of the automatic bifurcation labeling on the VascuSynth dataset using the LDA.

extract the basic features, *i.e.* the 3D coordinates of each bifurcation center (x_0, y_0, z_0), the coordinates of the branches' end points, i.e. of the next bifurcation along each branch (x_i, y_i, z_i), where $i = \{1, 2, 3\}$, and the sum of the three unit direction vectors of all branches (S_x, S_y, S_z). Some features, (which could not be collected on VascuSynth) such as the tortuosity $(\mathcal{T}_i)$, the diameters $(\varnothing_i)$, or the arteries' cross-section area $(\mathcal{A}_i)$, may also play an important role for the recognition and classification of the bifurcations. It is important to note that the diameter feature is a bit different, indeed, the arteries might not be perfectly circular, they (often) present an elliptical shape, and we thus compute for each and every artery both a maximum $(\varnothing_{max})$ and minimum diameter $(\varnothing_{min})$ [2] .

As previously explained (within the VascuSynth dataset description), out of the collected characterization dataset, we unfortunately have no means to link the branch of a given bifurcation to the branch of another bifurcation. In short, branch #1 on bifurcation N might not be labeled as #1 on bifurcation K. We must thus only compare some invariant properties, which is why we turned to the comparison of minimum, maximum, sums or averages of the three given features composing a bifurcation.

However, these basic features by themselves, may not be sufficient to categorize each individual bifurcation. We suggest again using some combinations of features in order to enrich the characterization. The choice of these combinations of features is very important. For instance, a bifurcation presenting a significant discrepancy between its branches' diameters (having one thick and one thin branch) might help to exploit a useful measure, we might for instance consider computing the ratio $\frac{\varnothing_{min}}{\varnothing_{max}}$ (ratio between the minimum and maximum diameters). Similarly for the angles (*resp.* cross-section), we suppose that having a large and a small value can help identifying each bifurcation, hence the computation of $\frac{\max(\Theta)}{\min(\Theta)}$ *resp.* $\frac{\max(\boldsymbol{\mathcal{A}})}{\min(\boldsymbol{\mathcal{A}})}$. All basic and combined features of the mice bifurcations (*i.e*, 40 features) are presented in Table 3 and they are also included to train the machine learning algorithms for bifurcation recognition and labeling. In this Table, $max(\boldsymbol{\mathcal{A}})$ refers to $max(\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3)$, i.e. the maximum cross-section area among the three branches, whereas the combination $(\max(\mathcal{A}_i) \times \mathcal{T}_i)$ refers to the largest branch's area multiplied by it's own tortuosity.

For the statistics of the bifurcations detected in the characterization phase, in total, we could obtain 480 PoI across the 30 test images (having 16 PoI/image). However, only 407 PoI were actually enumerated and labeled. This is due to the important variability between vasculatures, indeed not all the bifurcation centers are necessarily present in each vascular graph. Bifurcation H was present in more than 96 % of the objects, whereas bifurcations B, C, D and K were present in 93.3 % of the objects (28/30). As for the ones labeled E, F, G, I, J, L, M, N, P, they were present in at least

Table 3. Geometric properties of one bifurcation in 3D mouse vasculatures.

Basic characteristics	Combinations
• (x_0, y_0, z_0) coordinates of centers	• $\frac{\max(\Theta)}{\min(\Theta)}$
• $(\max(x_i), \max(y_i), \max(z_i))$ maximum of extremities coordinates	• $\frac{\max(\boldsymbol{\mathcal{A}})}{\min(\boldsymbol{\mathcal{A}})}$
• $(\min(x_i), \min(y_i), \min(z_i))$ minimum of extremities coordinates	• $\max(\boldsymbol{\mathcal{A}}) \times \max(\boldsymbol{\mathcal{T}})$
• Directional sum vector (S_x, S_y, S_z)	• $\max(\mathcal{A}_i) \times \mathcal{T}_i$
• Average, sum, min and max of the three Angles $(\Theta = (\theta_1, \theta_2, \theta_3))$	• $\max(\mathcal{T}_i) \times \mathcal{A}_i$
• Average, sum, min and max tortuosity $(\boldsymbol{\mathcal{T}} = (\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3))$	• $\frac{\max(\mathcal{A}_i)}{\varnothing_{maxi}}$
• Average, min & max of the maximum diameter $(\varnothing_{\boldsymbol{max}} = (\varnothing_{max1}, \varnothing_{max2}, \varnothing_{max3}))$	• $\frac{\max(\mathcal{A}_i)}{\varnothing_{mini}}$
• Average, min & max of the minimum diameter $(\varnothing_{\boldsymbol{min}} = (\varnothing_{min1}, \varnothing_{min2}, \varnothing_{min3}))$	• $\frac{\max(\varnothing_{max})}{\min(\varnothing_{min})}$
• Average, sum, minimum and maximum of Area (Cross-section) $(\boldsymbol{\mathcal{A}} = (\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3))$	• $\frac{\max(\varnothing_{min})}{\min(\varnothing_{min})}$
	• $\frac{\max(\boldsymbol{\mathcal{A}})}{\max(\varnothing_{max})}$

80 % of the images. While only 76.7 % (23/30) had a bifurcation O and 66.7 % (20/30) had the one labeled A. As pointed out in earlier works, there can be a significant structural variability[3,4] in the vasculature of human subjects, we tend to think that such a variability also exists for mice vasculatures.

3.2.2 Results

Similarly to the experiment on the VascuSynth dataset, we have performed PoIs recognition by applying the following three tests: first, for "Method 1", we used only the 3D coordinates of the bifurcation centers and applied machine learning techniques for labels classification. Second, in "Method 2", beside using the 3D coordinates, we have added the features computed for each bifurcation (Table 3) without applying any dimensionality reduction. Afterwards, we applied the machine learning classifiers. For the third and last test ("Method 3"), we used a dataset containing both the 3D coordinates of the bifurcations and their constituting features, the only difference lies in the use of the LDA for a dimensionality reduction prior launching the classification (seven LDA components were retained). The Figure 6 shows a 3D graphical representation of each bifurcation group using only the first three LDA components.

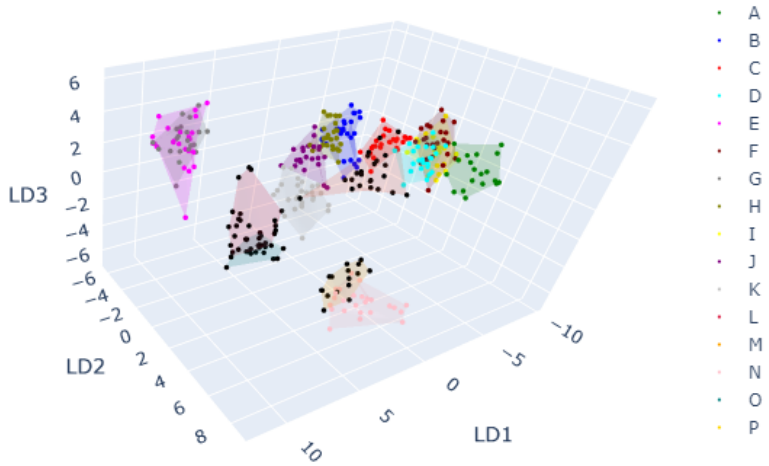


Figure 6. 3D representation of the 16 bifurcations in the mouse images with labels in the LDA plane.

For the comparison of the three tested approaches, the Table 4 and Figure 7 illustrate the average accuracy and F1-Score of each tested classification algorithm. We can observe that the accuracy of the Naive Bayes algorithm performs best with a recognition rate above than 95 % and the LDA offers the second best performances. For every PoI, we report in figure 9 the accuracy, sensitivity and F1-Score values of the classification using the Naive Bayes classifier. The confusion matrix constructed for the automatic labeling (Figure 8) shows the probabilities of correctly classifying the PoIs with their true labels. Moreover, it also shows the classification errors. As we can observe, PoIs G and O are the two bifurcations with the more classification errors, G is detected as being E and point O as L.

There are several causes for these misclassifications: 1) the absence of one or more arteries forming the PoI. 2) the small arteries of one PoI are close or adjacent to the small arteries of another PoI. 3) the arteries are frequently of the same order of magnitude in position, length and diameter. Therefore, the algorithm mistakenly extends a bifurcation to its neighboring cluster.

Table 4. Accuracy [%] and F1-Scores [%] on the Mice generated images

Classifiers	Method 1		Method 2		Method 3	
	Accuracy	F1-Score	Accuracy	F1-Score	Accuracy	F1-Score
SVM	84.3	80.8	87.6	85.6	**93.4**	**92.0**
Decision tree	82.9	78.7	84.3	81.4	**87.7**	**85.0**
Logistic regression	61.6	54.1	77.4	73.4	**88.3**	**85.4**
K-NN	85.4	82.1	82.2	78.8	**94.6**	**93.3**
Naive Bayes	83.9	80.4	81.9	79.2	**95.2**	**94.1**
Random forest	86.4	83.7	92.3	90.9	**93.6**	**92.5**
XGBoost	85.3	82.1	87.9	85.3	**91.7**	**90.2**
LDA	84.1	80.7	90.4	88.3	**95.0**	**93.7**

(a) Accuracy (b) F1-Score

Figure 7. Accuracy and F1-Scores on the Mice generated images

4. CONCLUSION AND PERSPECTIVES

In this work, we have presented an automated solution to label the main bifurcations in 3D vascular images. Such a process is rather difficult to handle on human MRI data. Effectively, the noise inherent to MRI acquisitions may lead to important misclassification errors. We have thus opted for a simplified setup, before turning to human brains anatomical atlas generation. In this study, we have tested our solution on two different datasets : one synthetic dataset (VascuSynth) that we have set to be composed of 14 bifurcations, this is the simplest setup one could imagine in such a context. Our second setup was a biological dataset containing 3D mice brain images. The mice were injected with a contrast liquid, thus making the vasculature much easier to segment, and most importantly noise free. The 16 main bifurcations forming the circle of Willis were labeled on these mice vascular trees. We have shown that using the bifurcations' geometrical properties along with their respective 3D coordinates, can be exploited for an improved classification. Thanks to a dimensionality reduction step (using a Linear Discriminant Analysis), we can avoid using a registration method prior to process a 3D clustering. Indeed, our results showed that such an approach can significantly improve the prediction of the bifurcations labels without relying on a registration step. In this study, a leave-one-out validation was performed on a set of 30 synthetic VascuSynth images, and we were able to reach a recognition rate of 98% for the main 14 bifurcations using the LDA.

For the mice acquisitions, the labeling of the 16 bifurcations of interest with the Naive Bayes classifier exhibited an accuracy above 95%. Thanks to these encouraging results with both synthetic and mice datasets, in our future works, we

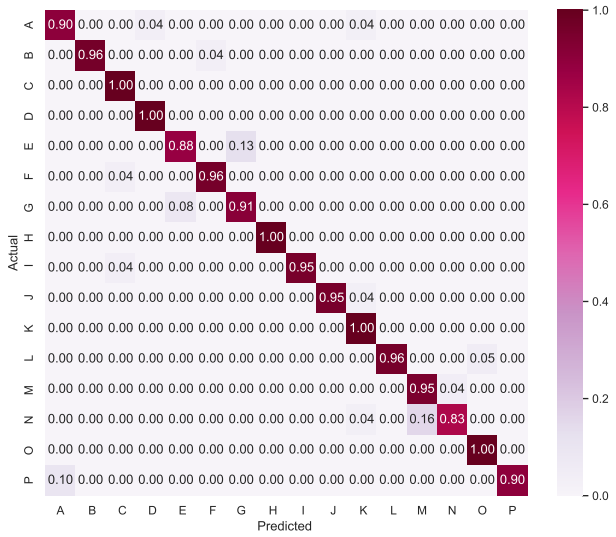


Figure 8. The confusion matrix of the automatic labeling of mouse images with Naive Bayes classifier.

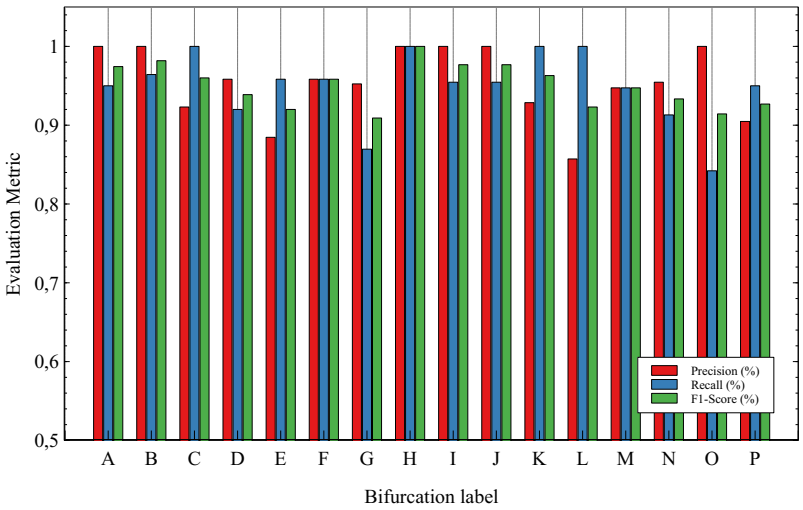


Figure 9. Detail of classification performance of 16 PoIs in mouse images with Naive Bayes algorithm.

plan to conduct similar experiments on human MRI acquisitions. Future works might encompass:(1) using larger datasets, (2) finding new features best describing the arteries bifurcations, (3) applying deep learning architectures and algorithms for bifurcation classification without requiring features computation, (4) testing the proposed method onto 3D human MRI acquisitions and (5) exploitation of non-linear dimensionality reduction techniques, such as Quadratic Discriminant Analysis, diffusion maps or isomap.

REFERENCES

[1] Bourcier, R. et al., "Understanding the pathophysiology of intracranial aneurysm: The ican project," Neurosurgery **80**, 621–626 (2017).

[2] Nouri, A. et al., "Characterization of 3D bifurcations in micro-scan and MRA-TOF images of cerebral vasculature for prediction of intra-cranial aneurysms," Computerized Medical Imaging and Graphics **84** (2020).

[3] Robben, D. et al., "Simultaneous segmentation and anatomical labeling of the cerebral vasculature," Medical Image Analysis **32**, 201–215 (2016).

[4] Bogunovic, H. et al., "Anatomical labeling of the circle of willis using maximum a posteriori probability estimation," IEEE transactions on medical imaging **32** (2013).
[5] Dunås, T. et al., "A Stereotactic Probabilistic Atlas for the Major Cerebral Arteries," Neuroinformatics **15**(1), 101–110 (2017).
[6] Ota, S. et al., "Automated anatomical labeling of bronchial branches using multiple classifiers and its application to bronchoscopy guidance based on fusion of virtual and real bronchoscopy - art. no. 69160g," Proc SPIE **6916** (2008).
[7] Wang, X. et al., "Automatic labeling of vascular structures with topological constraints via HMM," Lecture Notes in Computer Science **10434 LNCS**, 208–215 (2017).
[8] Bilgel, M. et al., "Automated anatomical labeling of the cerebral arteries using belief propagation," Proc. of SPIE **866918** (2013).
[9] Zhao, M. and Hamarneh, G., "Bifurcation detection in 3d vascular images using novel features and random forest," (2014).
[10] Jassi, P. and Hamarneh, G., "Vascusynth: Vascular tree synthesis software," Insight Journal **January-June**, 1–12 (2011).
[11] Ghanavati, S. et al., "Automatic anatomical labeling of the complete cerebral vasculature in mouse models," NeuroImage **95** (2014).
[12] Simard, P., Steinkraus, D., and Platt, J., "Best practices for convolutional neural networks applied to visual document analysis.," 958–962 (2003).
[13] Ronneberger, O. et al., "U-net: Convolutional networks for biomedical image segmentation," LNCS **9351**, 234–241 (2015).
[14] Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., and Ronneberger, O., "3d u-net: Learning dense volumetric segmentation from sparse annotation," CoRR **abs/1606.06650** (2016).
[15] Sarker, I. H., "Machine learning: Algorithms, real-world applications and research directions," SN Computer Science **2**(3), 1–21 (2021).