# Automatic classification of the cerebral vascular bifurcations using dimensionality reduction and machine learning

Ibtissam Essadik, Anass Nouri, Raja Touahni, Romain Bourcier, Florent Autrusseau

# Automatic classification of the cerebral vascular bifurcations using dimensionality reduction and machine learning

Ibtissam Essadik[a], Anass Nouri[a,b], Raja Touahni[a], Romain Bourcier[c], Florent Autrusseau[c,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, CHU de Nantes, Nantes Université, 44006, Nantes, France*
[d] *LTEN, Nantes Université, Polytech'Nantes, 44306, Nantes, France*

**Abstract**

This paper presents a method for the automatic labeling of vascular bifurcations along the Circle of Willis (CoW) in 3D images. Our automatic labeling process uses machine learning as well as dimensionality reduction algorithms to map selected bifurcation features to a lower dimensional space and thereafter classify them. Unlike similar studies in the literature, our main goal here is to avoid a classical registration step commonly applied before resorting to classification. In our approach, we aim to collect various geometric features of the bifurcations of interest, and thanks to dimensionality reduction, to discard the irrelevant ones before using classifiers.

In this paper, we apply the proposed method to 50 human brain vascular trees imaged via Magnetic Resonance Angiography (MRA). The constructed classifiers were evaluated using the Leave One Out Cross-Validation approach (LOOCV). The experimental results showed that the proposed method could assign correct labels to bifurcations at 96.8% with the Naive Bayes classifier. We also confirmed its functionality by presenting automatic bifurcation labels on independent images.

*Corresponding author
*Email address:* Florent.Autrusseau@univ-nantes.fr (Florent Autrusseau )

*Keywords:*
Vascular atlas, Human Brain, Artery Characterization, Dimensionality Reduction, Machine Learning Algorithms, Registration.

## 1. INTRODUCTION

Automated labeling of the bifurcations that constitute blood arteries is a critical problem with many practical applications. It could provide additional guidance to an interventional radiologist when surveying a patient's vascular system, or provide an automatic measurement of specific vessel segments [1]. It can also be used to detect trends in the development of a given disease. Among the vascular disorders, intracranial aneurysms (ICAs) are of particular interest as their occurrence, or more precisely their rupture may be devastating [2].

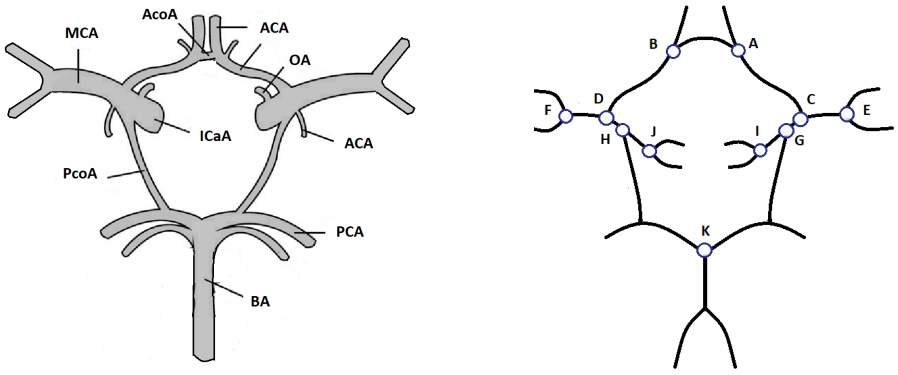
Within the framework of this study, we will specifically focus on the CoW and its constituting arteries. Indeed, the ICAs presenting the highest rupture rates, are those located onto the bifurcations of the CoW [3]. The CoW is a network of arteries which connects the left and right sides of the anterior cerebral circulation to the posterior cerebral circulation at the base of the skull (Fig. 1(a)). It originates from three primary arteries: the left and right internal carotid arteries and the vertebrobasilar artery.

The prevalence of these ICAs can vary between 5 and 8% of the population [4]. It is therefore important to understand the anatomical variants with their effects on the haemodynamic and geometrical parameters responsible for the pathogenesis of this neurological disease [5].

Analyzing the statistical variation of the arteries and bifurcations' geometry may help to quantify the risk factors related to the vascular tree geometry for aneurysm occurrence [6]. It is important to notice that the significant variability of the vasculature shapes among patients, and the important complexity of the arteries' shapes [7, 8, 9] make the automatic labeling of the CoW a very challenging problem. In order to help monitoring ICAs, the detection and recognition of the bifurcations and aneurysms are needed. The Manual anatomical labeling is

a tedious and time consuming task. Therefore, its automation becomes crucial to streamline the geometric characterization of a large number of cases [10].

In particular, we are interested in the automatic labeling of major CoW bifurcations (Fig. 1(b)), where 90% of all the cerebral aneurysms occur [10]. Bifurcations are defined as the end points of cerebrovascular vessels. Thus, the labelling of bifurcation centers identifies the intersection points of the arteries connecting them. In this study, we have selected eleven Bifurcations of Interest (BoI) presenting the highest risk of aneurysm development (see Fig. 1(a)) : The Anterior communicating Artery ($AcoA$), the Basilar Artery ($BA$), left and right Posterior Cerebral Arteries ($PCA$). They are connected to the posterior communicating arteries (PcoA). The internal carotid artery ($ICaA$), which is divided into the anterior cerebral artery ($ACA$) and the middle cerebral artery ($MCA$). In addition to the ophthalmic artery ($OA$) that is a branch of the ICaA.




(a) The area of interest

(b) Schematic representation of the BoIs.

Figure 1: Anatomy of the Circle of Willis.

In this paper, we present a new automatic method to identify the main BoI on 3D TOF-MRA human brain acquisitions. This study follows a previous work which was focused on the automatic labeling of the main bifurcations constituting the CoW in synthetic and mouse vasculatures [11]. Our suggested method is designed to satisfy the following main expectations: 1) experiment

the method on complex images; 2) propose a more robust labeling method, and 3) demonstrate how a dimensionality reduction method can replace the registration step. The remainder of the paper is organized as follows. In Section 2, a summary of the previous studies is presented. Section 3 describes the workflow of the proposed method. The evaluation on a set of 50 TOF-MRA images is presented in Section 4, experimental results are presented in section 5. And finally, section 6 provide a discussion of the strengths, limitations of the proposed method and concludes the paper.

## 2. RELATED WORKS

Structural changes in cerebral vasculatures are key indicators of many diseases affecting the brain. Primary angiopathies, vascular risk factors, vascular occlusions, and strokes all affect the functions of the brain's vascular network [12]. The frequency of cerebrovascular aneurysms is rapidly increasing [13]. Considering the risk of rupture that is life-threatening, it is essential to develop a tool to help radiologists detect aneurysms at an early stage [14].

This work is part of a wide scientific French project [15]. The main goals of this project are the following: 1) an in-depth analysis of factors such as genetical predisposition, smoking, chronic diseases which can lead to the occurence of an ICA located along the CoW, 2) the automatic detection of bifurcations and their geometrical analysis [6] and the automatic identification of the bifurcations of the CoW where the risk of rupture is higher. In this work, we present a new automated method for identifying major bifurcations of interest in the CoW within 3D human brain acquisitions. As previously mentioned, this study extends our previous work [11], where we labeled major bifurcations in both artificial images (vascusynth [16]) as well as in mouse brain acquisitions. Using a set of 30 vascuSynth synthetic images, we achieved a 98% recognition rate for the 14 main bifurcations using a Linear Discriminant Analysis (LDA). On mouse brain acquisitions ($\mu$-CT scans), labeling the 16 bifurcations of interest with the Naive Bayes classifier exhibited an accuracy of more than 95%.

The anatomical labeling may be considered as a mapping of an unlabeled case to an atlas, represented by a knowledge base of the population mean with geometric and structural variability. Many works have been devoted to the design of brain atlases and have attempted to associate various parts of the brain with a particular ability or task [17]. Fan *et al.*[18] have developed an atlas that explores the complex relationships between structure, and connectivity in order to better understand various functions of the human brain. Nowinski *et al.*[19] summarize the creation, validation and commercialization of the brain atlas.

The construction of a cerebral anatomical atlas of the human brain has already been studied in the past. Bogunovic *et al.* [7] took advantage of the a posteriori maximum likelihood estimation to automatically label the CoW. Dunas *et al.*[20] constructed a probabilistic atlas for automatic labeling of cerebral arteries. In the work of Robben *et al.*[8], the authors used graphical matching algorithms to identify and label human brain arteries.

With the development of machine learning methods, new methods addressing the automatic bifurcation labeling problem have been proposed. Ota *et al.* [21] proposed a method that provides the anatomical names of the bronchial branches using the AdaBoost multi-class technique. In the work of Murat *et al.*[22], they identified cerebral arteries using random forest and a Bayesian network. Wang *et al.*[23] also took advantage of supervised machine learning algorithms to automatically label cerebral arteries. To identify the bifurcations, they relied on the interdependence of the vessel axis as well as several features such as centerline length or curvature. Their method is also based on a model of the vessel centerline, therefore, they did not take into account the characteristics of the diameter and cross section of the arteries. In the work of Zhao *et al.*[24], Random Forest algorithm was implemented based on vessel features for bifurcation detection on both artificial and 3D clinical chest CT images. All these approaches rely on a registration step before the creation of the atlas, in order to align the images and increase their correspondence. Registration is the process of finding a geometrical transformation that spatially aligns two distinct images. It is widely used in medical imaging, in particular, neuroradiologists

may resort to such methods to quantify aneurysm growth [25].

In this work, we present an automatic solution to categorize CoW bifurcations (where aneurysms are prevalent) from 3D MRA brain acquisitions. In the proposed method, we apply machine learning classifiers onto various geometric features of the cerebral arteries. We hereby exploit more features characterizing each bifurcation of interest than was previously done from competing methods of the literature [7, 23]. For instance, not only do we consider the bifurcations coordinates, but also the arteries' thickness (diameters and cross-sections) are taken into consideration by our process. We believe that such properties may be highly discriminant when it comes to classify a given bifurcation.

Furthermore, in this work, one of our basis assumptions was that using n-dimensional clustering combined with a dimensionality reduction step might successfully replace the registration stage. To do so, we have used the Linear Discriminant Analysis (LDA). We also provide an additional experimental evaluation of the approach using rigid registration before the labeling step. Using evaluation measures, we demonstrate the advantages of the proposed method over state of the art methods.

## 3. MATERIALS AND METHODS

The whole pipeline architecture of the automatic labeling is shown in Figure 2. The MRA data were preprocessed before generating feature vectors for each bifurcation. After preprocessing, the features were combined to form the concatenated feature vector. A dimensionality reduction step was applied before classification. We used a Linear Discriminant Analysis to project the concatenated feature vector onto a reduced number of dimensions where the points representing the data belonging to the same group (bifurcation label) tend to be close to each other. Eight classifiers were used : Support Vector Machine (SVM), Logistic Regression (LR), LDA, Decision Tree (DT), Random Forest (RF), K-Nearest Neighbor (KNN), Naive Bayes (NB), XGBoost and Quadratic Discriminant Analysis (QDA). Further details will be provided for all these clas-

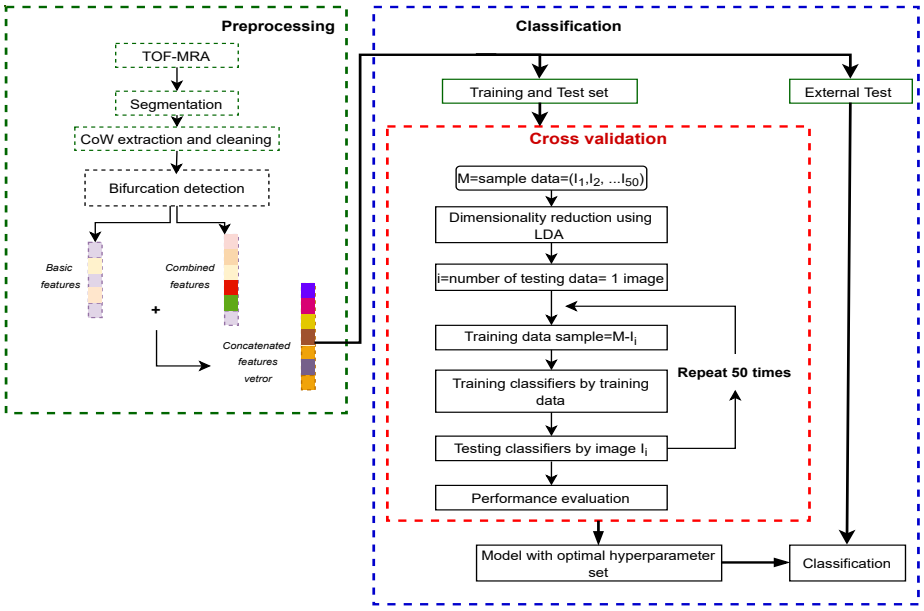



Figure 2: The pipeline architecture of the data preprocessing and automatic labeling method.

sifiers in section 3.2.2.

### *3.1. Preprocessing*

#### *3.1.1. Subjects*

In this study, 50 TOF MRA source images were used to develop the automatic labeling solution. The data includes images with and without aneurysms. The images were acquired from various University Hospitals, all over France. 15 images were reconstructed within a matrix of size $500 \times 500 \times z$, 15 images have a resolution of $512 \times 512 \times z$, and 20 other images are of size $475 \times 475 \times z$, where $z$ refers to the number of slices in the 3D volume.

Using zero-padding along $x$ and $y$ coordinates, we resized all images to the same resolution of $512 \times 512 \times z$ voxels (preserving the slice number). High-resolution images have a voxel size of $0.49 \times 0.49 \times 0.57 \mu m$ and were all saved with a 16-bit per voxel depth.

*3.1.2. Bifurcations of interest and annotations*

To allow extraction of quantitative features from the vessel structures, the acquired brain vasculature must be segmented in 3D. The MRA data was segmented using 3D Slicer[1]. Indeed, we must first extract the arteries from the 3D vasculatures by manual segmentation. For each image, a region of interest along the CoW is delineated. Note that this manual segmentation is only performed once, for the learning phase of the anatomical atlas construction process. Obviously, once the model is set up, an automatic segmentation will be applied to any new image provided to the model.

Once the vascular tree (and more particularly its CoW) has been properly segmented, a 3D skeleton is computed and its corresponding 3D undirected graph is built [26]. The labeling algorithm presented in this work consists of estimating the label of each bifurcation using machine learning classifiers. Thus, generating an atlas requires a training set of vascular graphs and ground truth annotation of BoIs on these graphs. Note that not all BoIs will be present in every vascular graph. This variability can be caused either by the patient's particular vasculature anatomy or by the segmentation process and the subsequent skeletonization.

In the present study, only the bifurcations with the highest risk of aneurysm formation along the CoW were selected. We first prune the centerline model to a region around the CoW. For a complete representation of the Circle of Willis, this yields to 13 bifurcations. We have witnessed in our dataset that both posterior cerebral arteries were regularly missing in the MRA-TOF acquisitions. Therefore, the two bifurcations constituted by the left and right PCA and PCoA arteries (in Figure 1) are not considered in our study when the communicating artery (PCA or AcoA) was actually missing in the majority of the MRA subjects in our dataset. Due to the heterogeneity of the different cases, we are thus mainly interested by eleven bifurcation points, which are labeled as $A$ through $K$ in Figure 1.

[1] `https://www.slicer.org/`

*3.1.3. Features extraction*

It is widely accepted that the geometrical features of the bifurcation plays an important role in the aneurysms formation [27]. Similarly to our previous work, we hereby propose to classify the bifurcation points based on their associated geometric features. That is, for each BoI, we compute several geometrical features from its three branches, and moreover, we also consider various combined features. Overall, we end up gathering 61 computed features identifying the bifurcations: 35 collected geometrical features, and 26 combinations. The 35 basis features were collected using the works in [6].

The basic feature vector contains: the 3D coordinates of the center of each bifurcation, the extremity coordinates of the branches, the sum of the three unit direction vectors of all branches $(S_x, S_y, S_z)$, the tortuosity $(\mathcal{T}_i)$, the maximum and minimum diameters$(\varnothing_i^{min}, \varnothing_i^{max})$, the three angles vector between each pair of branches $\Theta_i$ and the cross-section area of the arteries $(\mathcal{A}_i)$, where $i = \{1, 2, 3\}$ refers to the bifurcation branch. Note that for the bifurcations' classification, we rely on the comparison of the minimum, maximum, sum or average of the three computed features composing each bifurcation. The combinations of characteristics are very critical for the classification accuracy. For each bifurcation, we have computed the ratio between the minimum and maximum diameters, angles and cross-section ( $\frac{\max(\varnothing_i^{max})}{\min(\varnothing_i^{min})}$ , $\frac{\max(\Theta)}{\min(\Theta)}, \frac{max(\mathcal{A}_i)}{min(\mathcal{A}_i)}$). Also, we suppose that the product of the largest branch's section by it's own tortuosity, or between the largest branch's tortuosity and it's corresponding section can help identifying each bifurcation.

In this work, we add new features related to the bifurcation orientation. For each branch constituting the bifucation, we make the projection of its directional vector with respect to the $xy$, $xz$ and $yz$ plans.

Thus we compute the angles $\theta^i_{xy}$,$\theta^i_{xz}$ and $\theta^i_{yz}$. Consequently, we get three angles computed for each artery. Unfortunately, the characterization process [6] cannot relate a given angle with it's corresponding arteries, we thus use the average, sum, min and max of the three directional angles along each axis, i.e.

$\Theta_O = (\theta^i_{xy}, \theta^i_{xz}, \theta^i_{yz})$. All those 61 features are presented in Table 1 and are considered as input data to train machine learning algorithms for bifurcation recognition and labeling in MRA images.

| Feature | Description |
|---|---|
| Basics | $(x_0, y_0, z_0)$: coordinates (coords) of centers<br>$(\max(x_i), \max(y_i), \max(z_i))$: maximum of extremities coords<br>$(\min(x_i), \min(y_i), \min(z_i))$: minimum of extremities coords<br>$(S_x, S_y, S_z)$: Directional sum vector<br>$(\Theta = (\theta_1, \theta_2, \theta_3))$: Average, sum, min and max of the three Angles<br>$\left(\Theta_O = \left(\theta^i_x, \theta^i_y, \theta^i_z\right)\right)$: Average, sum, min and max of the three orientation angle<br>$(\boldsymbol{\mathcal{T}} = (\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3))$: Average, sum, min and max of tortuosity<br>$\left(\oslash_{\boldsymbol{max}} = (\varnothing^1_{max}, \varnothing^2_{max}, \varnothing^3_{max})\right)$: Average, min & max of the maximum diameter<br>$\left(\oslash_{\boldsymbol{min}} = (\varnothing^1_{min}, \varnothing^2_{min}, \varnothing^3_{min})\right)$ : Average, min & max of the minimum diameter<br>$(\boldsymbol{\mathcal{A}} = (\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3))$: Average, sum, minimum and maximum of Area-Cross-section. |
| Combinations | $\left( \frac{\max(\varnothing^{max}_i)}{\min(\varnothing^{min}_i)} \right.$ , $\frac{\max(\Theta)}{\min(\Theta)}, \left.\frac{max(\mathcal{A}_i)}{min(\mathcal{A}_i)}\right)$<br>$\max(\boldsymbol{\mathcal{A}}) \times \max(\boldsymbol{\mathcal{T}})$ , $\max(\mathcal{T}_i) \times \mathcal{A}_i$, $\max(\mathcal{A}_i) \times \mathcal{T}_i$<br>$\frac{\max(\mathcal{A}_i)}{\max(\varnothing^i_{max})}$, $\frac{\max(\mathcal{A}_i)}{\min(\varnothing^i_{min})}$ , $\frac{\max(\oslash_{max})}{\min(\oslash_{min})}$, $\frac{\max(\oslash_{min})}{\min(\oslash_{min})}$, $\frac{\max(\boldsymbol{\mathcal{A}})}{\max(\oslash_{max})}$ |

Table 1: Feature set of collected data.

### *3.2. Classification*

After the preprocessing step, the combined and basic features were merged to form the final concatenated feature vector that represents, in our context, the descriptor of a target bifurcation. Once the descriptor vector, composed of 61 feature, is assembled, it is being fed to the dimensionality reduction step prior to be used by the classification algorithms. The performances of the final model were assessed by a separate test set to ensure generalization. The complete architecture of the Atlas construction pipeline is shown in Figure 2.

#### *3.2.1. Dimensionality reduction*

A dimensionality reduction step can be applied before the classification process due to a possible overlapping or irrelevant features that may lead to overfitting during the training [28]. Independent Component Analysis and non-negative matrix factorization represent two unsupervised techniques for dimensionality reduction. One of the most well-known unsupervised multivariate methods is Principal Component Analysis [29]. For unsupervised classification, a labelling of the various classes is not necessary, whereas for supervised clustering, the classes labels must be known while carrying out dimensionality reduction.

In this work, we aim to tackle the bifurcation clustering issue using supervised classification methods, and more particularly, we have used the LDA as a dimensionality reduction algorithm.

The main goal of the LDA is to map input data into lower dimensions with the highest variance using a linear function [30]. In other words, the LDA determines a linear combination feature vectors in the aim to maximize the mean difference between the classes or to maximize the interclass variance. Hence, similar data are clustered altogether whereas dissimilar ones are further separated. In this Work, the LDA was used with a twofold purpose: *i)* reduce the data dimensionality and *ii)*, act as a classifier in order to predict a bifurcation label. Therefore, bifurcation labeling can be performed directly using the LDA algorithm or by applying another classification algorithm on the mapped data

within the lower dimension space.

### *3.2.2. Machine learning classification*

Over the past decades, Artificial Intelligence, and more particularly Machine Learning (ML), have been increasingly used for various medical imaging tasks [31]. In this paper, the bifurcation labeling process is considered as a multi-label classification problem. Using the variables generated with the LDA, we have tested the following nine classifiers : LDA, SVM, LR, RF, QDA, DT, NB, KNN and XGBoost. The first one was developed by Fisher in 1936 [32]. Its principle is to project the data set in a lower dimension space where the points belonging to the same group tend to gather together. The QDA is quite similar to the LDA, it allows quadratic decision boundaries between classes [33]. SVM was first suggested by Vapnik in 1995 [34], it aims to find the optimal hyperplane separating the data set with differing labels into multiple hyperspaces [35]. Similarly to SVM, LR also produces a hyperplane using a linear transformation function and a sigmoid activation function to split the data and provide the probability that the unseen data is classified into a given group. DT represents the hierarchical exemplification of knowledge relationships that contains nodes and connections [36]. Each node represents the characteristics of a category to be classified and each subset defines a value that can be taken by the node. Following the same principle, RF is a bagging algorithm that predicts the labels of unseen data based on the votes of all the embedded decision trees [37]. Each decision tree is generated by randomly choosing the data from the total available data. The KNN classifier groups the data into coherent clusters or subsets and ranks the newly entered data based on their similarity to the previously formed data [38]. And finally, XGBoost is a popular supervised machine learning algorithm used for classification and prediction problems. It is a modified version of the gradient boosting algorithm, which classifies data by combining model resulting from classification decision trees [39]. In the context of our work, the best algorithm was selected by cross-validation, as described within the next section, based on the evaluation metrics.

## 4. EVALUATION METHODOLOGY

Overall, 50 MRA-TOF volumes were selected and a LOOCV step was performed to build the predictive model with optimal hyperparameters. Among the 50 volumes, 49 were used for training, and one was used for validation. Each and every volume was iteratively used as the "validation set".

The classification performances were systematically analyzed for the different combinations of classifiers using LOOCV in the section 5. The reliability was evaluated using the area under the ROC curve (AUC). In parallel, sensitivity, specificity and F1-Score were measured for additional performance evaluation, all these metrics are presented below.

$$\text{Precision} = \frac{T_p}{T_p + f_p}$$
$$\text{Recall} = \frac{T_p}{T_p + f_n}$$
$$\text{F1} - \text{Score} = \frac{2 \times Recall \times Precision}{Recall + Precision}$$
$$\text{Accuracy} = \frac{T_p + T_n}{T_p + f_n + f_p + T_n}$$

where the True positive ($T_p$) is the number of positive samples predicted as positive; True Negative ($T_n$) is the number of negative samples predicted as negative. False Positive ($F_p$): the number of negative samples predicted as positive; and False Negative ($F_n$): the number of negative samples predicted as positive.

## 5. EXPERIMENTAL RESULTS

Traditionally, registration techniques are used to attenuate the differences between various acquisitions in order to construct an anatomical atlas. In this section, we compare the effects of registration and dimensionality reduction (using LDA) on the classification results. Both experiments are performed with a LOOCV on 50 MRA 3D images.

### *5.1. Labeling using registration*

In the first experiment, all volumes composing our dataset are rigidly aligned using the General Registration (BRAINS) library [40] (the target reference had a dimension of $512 \times 512 \times 210$ voxels). From each registered image, we first compute a 3D undirected graph on the skeleton of the segmented MRA-TOF. Our experimental protocol was as follows:

1. **Method A**: From the registered images, we only collect the bifurcations' 3D coordinates. We then apply our labeling approach on those coordinates.
2. **Method B**: We compute geometrical features described in the previous section from non registered images and concatenate them with the 3D coordinates of bifurcations's centers used in the "Registration - Method A". Note that here, the geometrical features were collected prior launching the registration process, in order to avoid any distortion of the arteries' geometry. Furthermore, the concatenated data are used as classification inputs without any dimensionality reduction since we expect that such a step might allow to avoid resorting to registration.

Table 2 shows the accuracy and F1 score for each method separately. We can notice that the integration of several additional geometric features improves the classification rate for the majority of the ML classifiers. Using geometric features, XGBoost and QDA showed the highest accuracies of 86.7% and 83% respectively. However, KNN showed the lowest performance with an accuracy of 60.7%. Our findings suggest that the performance of each algorithm fluctuates with different inputs for the two labeling methods. The results also illustrate that potential inaccuracies in 3D realignment can lead to classification errors.

### *5.2. Labeling without registration*

In this scenario, we have tested three different approaches. First, our method was applied on the three-dimensional coordinates of the bifurcations centers without any additional features where the LDA was used as a classifier ("Method

Table 2: The Method A indicates the labeling approach based only on the coordinates after registration, and Method B the labeling approach based on the coordinates after registration combined with the geometrical features.

| Classifier | Method A | | Method B | |
|---|---|---|---|---|
| | Accuracy | F1-Scores | Accuracy | F1-Scores |
| SVM | 62.3 | 54.5 | 74.4 | 73.1 |
| DT | 74.2 | 71.5 | 78.9 | 78 |
| LR | 51.9 | 42.6 | 58.3 | 56.0 |
| KNN | 74.5 | 70.2 | 60.7 | 59.1 |
| NB | 73.1 | 68.6 | 68.3 | 66.6 |
| RF | 77.5 | 73.7 | 81.5 | 81.0 |
| XGBoost | 78.9 | 76.0 | 86.7 | 86.3 |
| QDA | 75.8 | 71.7 | 83.9 | 83.2 |

1"). For the second approach ("Method 2"), the collected features were added to the bifurcations' center coordinates and directly used as inputs for the classification without any dimensionality reduction. Finally, in the third experiment, the LDA was used to represent the bifurcations' features in a lower dimensional space, the classifiers were then trained in the new components of this reduced space ("Method 3"). These three tests were performed to demonstrate the importance of BoI features and the improvement brought by the LDA in predicting the bifurcation labels during an automatic evaluation.

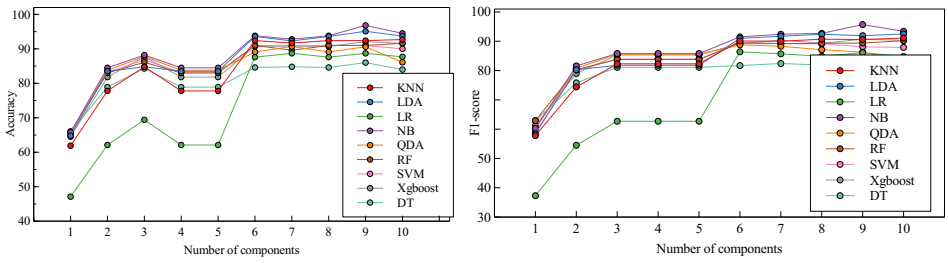
Our goal being to achieve an efficient clustering of the BoI based on various geometrical features, we do not know *a priori* which particular feature might be discriminative or not. We have hence opted for an approach where we deliberately exploit the large amounts of features, as presented earlier (61 geometrical features), and subsequently, we aim to reduce this initial set of features to a smaller one, which hopefully shall be more representative, to feed our machine learning classifiers. In this section, we demonstrate the effect of dimensionality

reduction methods (such as the LDA) on the performance of learning algorithms to predict the BoI labels.

### *5.2.1. Dimensionality reduction*

In a scenario where $N$ classes are to be clustered, the LDA transforms the set of data into $(N-1)$ discriminant functions. In this study, we aim to classify eleven bifurcations of interest, the LDA thus transforms the set of computed features into ten components. In order to select the best number of LDA components that will improve the evaluation rate of our classification, we have plotted the accuracy and F1-Score as a function of the LDA components. Figures 3(a,b) show that all classifiers achieve a maximum accuracy and F1-Score when the dataset dimensions are reduced to 9 components.

It is important to note that KNN and XGBoost reach the maximum accuracy of 92% when the dimensions of the dataset are reduced to 9, while the DT and LR algorithms reach 86% and 88.5% accuracy respectively. Moreover, the NB model after using LDA performs better when the dataset dimension is reduced to 9, reaching 96.8% of accuracy. The QDA and LDA classifiers respectively achieve 95% and 90.6% of accuracy.




(a) Accuracy (b) F1 score

Figure 3: Effect of the number of LDA components on the Accuracy and F1-Scores metrics.

### *5.2.2. Predictive models performance*

We examine the impact of the features and dimensionality reduction using the LDA on the performance of 9 popular ML algorithms. Specifically, we use

9 supervised learning algorithms to predict the BoI of labels. The accuracy and F1-Score of these algorithms are shown in Table 3 and Figure 6. We can notice that adding several additional geometric features improves the classification rate. Similarly, transforming and representing features in a LDA-reduced dimensional space with the 9 components has a very clear impact on the labeling of the human vasculature's bifurcations. Moreover, DT and LR have the worst performance in terms of accuracy and F1-Score. Although NB after dimensionality reduction exhibits the best performances (maximum accuracy (96.8%) and maximum F1-Score (95.7%)), the second highest accuracy is achieved by the LDA classifier with 95.1%. NB, LDA, and XGBoost show higher classification performances, with AUC values equal to 0.97, 0.96, and 0.95, respectively, while the DT algorithm presents the lower performance, with an AUC of 0.92. In addition, QDA, LR and RF, reach similar classification performances with AUC values of 0.93.(see Fig. 7).

We report the accuracy, AUC, and F1-Score for each particular BoI in Figure 4 using the NB classifier. We can notice that most bifurcations are correctly labeled with an identification rate of 100%. The bifurcations' misclassifications can be observed within the confusion matrix (Figure 5). From this confusion matrix, we can notice that the bifurcation points G, H and J suffer the most from classifications errors. G is occasionally detected as I and H is sometimes incorrectly recognized as being J. These misclassifications are mainly due to the variability of the vascular tree, especially missing or the obvious features of the two vessels ICaA and MCA. More interestingly, due to the absence of the ACA vessels, the D and C bifurcations are barely affected by the above reason.

Table 4 presents a comparison of the precision and recall for each bifurcation of interest for our proposed method as well as for two state-of-the-art algorithms [7, 23]. For each analyzed bifurcation, the best performances, in terms of Accuracy, Precision and Recall have been set to bold fonts in Table4, thus highlighting the good performances of our approach compared to state-of-the-art methods.

Moreover, it is important to point out that a comparison with these com-

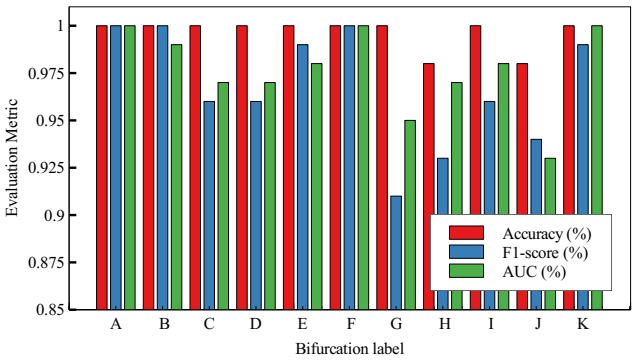



Figure 4: Details of the classification performances of the 11 BoIs in MRA-TOF images with the NB classifier.

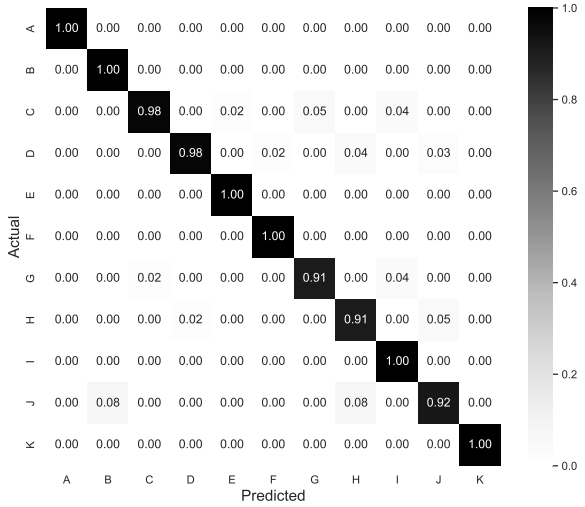



Figure 5: Confusion matrix of the automatic bifurcation labeling on the MRA images using the NB classifier.

peting methods is to be conducted with extreme care. Indeed, the images being used in [7, 23] were all acquired using the same MRA scanner, and hence all images within these two studies present the exact same resolution. Our dataset allows a better generalization, as all 50 tested images exhibited different image

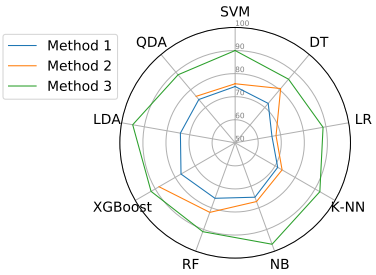
Table 3: The Method 1 indicates the labeling approach based only on the 3D coordinates. Method 2 refers to the labeling based on 3D coordinates combined to geometrical features and Method 3 is the approach based on the dimensionality reduction (LDA) applied on coordinates and geometrical features. For each method, we provide the Accuracy (A)[%] ,F1-Scores (F1) [%] and fitting time (FT).

| Classifiers | Method 1 | | | Method 2 | | | Method 3 | | |
|---|---|---|---|---|---|---|---|---|---|
| | A | F1 | FT | A | F1 | FT | A | F1 | FT |
| SVM | 74.4 | 67.9 | 52.72 | 72.1 | 69.3 | 8.02 | 90.0 | 88.1 | 1.08 |
| DT | 72.3 | 68.0 | 0.56 | 80.7 | 77.7 | 1.84 | 86.0 | 83.8 | 0.45 |
| LR | 66.2 | 60.4 | 0.56 | 72.3 | 68.0 | 3.07 | 88.7 | 86.1 | 0.59 |
| KNN | 71.3 | 66.7 | 0.32 | 73.5 | 70.3 | 1.37 | 92.4 | 90.6 | 0.45 |
| NB | 69.1 | 63.6 | 0.34 | 75.1 | 70.4 | 1.04 | 96.8 | 95.7 | 0.42 |
| RF | 75.6 | 71.3 | 10.49 | 82.1 | 79.2 | 32.73 | 91.0 | 89.4 | 11.87 |
| XGBoost | 77.1 | 72.3 | 57.36 | 88.1 | 85.6 | 86.58 | 92.1 | 90.7 | 80.23 |
| LDA | 74.1 | 68.4 | 0.65 | - | - | - | 95.1 | 91.9 | 0.60 |
| QDA | 74.5 | 69.7 | 0.77 | 66.3 | 64.5 | 1.20 | 88.5 | 86.3 | 0.58 |

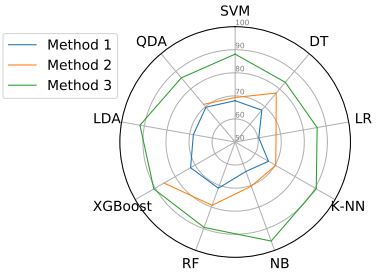
resolutions acquired from various scanners.

### *5.3. Testing on independant images*

Contrary to the works in [23], we have tested our approach on two independent images not belonging to the 50 images dataset on which we performed the LOOCV. This allows to analyse the behavior of our approach when considering a novel image. These two images have a complete CoW, i.e. for each image all 11 BoIs were correctly detected. However, only 18 BoIs were characterized due to the geometrical configuration of these two images (short branches, irrelevant branches belonging to bifurcations, etc.) Afterwards, the computed feature vectors were reduced using the LDA weights saved during the training phase. The new components obtained in this phase are used as inputs to the classifi-




(a) Accuracy




(b) F1-Score

Figure 6: Accuracy and F1-Scores on MRA brain acquisitions

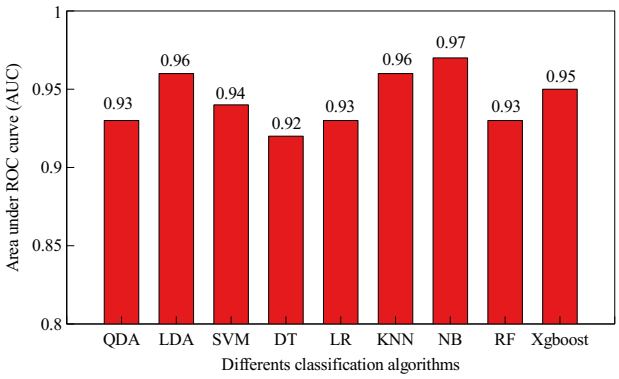



Figure 7: Comparison of different classification algorithms based on the AUC.

cation algorithms. As a result, there was only one misclassified bifurcation, the L-bifurcation, which was confused with the H-bifurcation. Thus, we achieved a labeling rate of 94% for these two images.

Table 4: The comparison of the evaluation of the 11 bifurcations of interest with accuracy (A), precision (P) and recall (R) is reported for the proposed method as well as for the best performing methods in the state of the art [7, 23].

| | **Wang [23]** | | | **Bogunovic [7]** | | | **Our method** | | |
|---|---|---|---|---|---|---|---|---|---|
| **BoI** | A | P | R | A | P | R | A | P | R |
| A | 99.3 | 95.9 | 94.0 | 96 | 95 | **100** | **100** | **100** | **100** |
| B | 99.3 | 100 | 87.8 | **100** | **100** | **100** | **100** | **100** | **100** |
| C | 99.2 | 93.9 | 92.0 | **100** | **100** | **100** | 99.7 | 95.2 | 98.0 |
| D | 99.8 | 96.2 | **100** | 98 | **100** | 98 | **100** | 99.7 | 98.1 |
| E | 98.2 | 90.7 | 78.0 | 80 | 80 | **100** | **100** | **97.8** | **100** |
| F | 97.6 | 77.8 | 84.0 | 84 | 84 | **100** | **100** | **97.8** | **100** |
| G | 99.4 | **100** | 90.0 | 98 | **100** | **97** | **99.9** | 93.9 | 95.8 |
| H | **99.5** | 96.0 | 96.0 | 98 | **97** | **100** | 98.4 | 90.3 | 91.4 |
| I | *N/A* | *N/A* | *N/A* | *N/A* | *N/A* | *N/A* | **100** | **93.1** | **100** |
| J | *N/A* | *N/A* | *N/A* | *N/A* | *N/A* | *N/A* | **97.6** | **89.8** | **91.9** |
| K | 97.4 | 79.6 | 78.0 | 84 | 84 | **100** | **100** | **100** | **100** |
| Average | 98.84 | 92.23 | 88.86 | 93.11 | 93.33 | 99.44 | 99.60 | 96.14 | 97.74 |

## 6. DISCUSSION AND CONCLUSION

In this work, we have proposed a method for anatomical labeling of the major 11 bifurcations comprising the anterior part of the cerebral vasculature on human MRA . Our method learns features of the arteries that describe the architecture of the vascular network using a subset of manually labeled samples. Features such as anatomical variation in branch length, orientation angles and tortuosity can be captured using this method. The problem is formulated as a supervised ML solution. The total performances of nine different ML algorithms were evaluated using a Linear Discriminant Analysis performing dimensionality reduction. A leave-one-out cross validation, performed on 50 brain MRA acquisitions, has shown higher accuracy than the best performing state-of-the-art methods ([7, 23]).

Without registration, we have shown how the geometrical characteristics of bifurcations, as well as their associated 3D coordinates, can be used to improve bifurcation labeling. We can avoid applying a registration stage before performing a 3D clustering thanks to a dimensionality reduction phase. Indeed, our findings demonstrate that, without relying on a registration step, such an approach can greatly enhance the prediction of bifurcation labels. However, our bifurcation labeling results, with an accuracy of more than 96.8% using the NB algorithm, compare favorably with those presented by Bogunović et al. [7] who achieved a correct labeling rate of the vascular tree of 95%. The labeling of the target vascular system is achieved by a maximum a posteriori probability estimation where the labeling probability of individual bifurcations is regularized by the prior structural knowledge of the graph they span. Furthermore, their method requires a global rigid realignment step before going through the labeling process. As for the method presented in [23], the proposed approach is based on independent vascular tree features, the bifurcations coordinates were not considered, and hence registration methods did not apply. In contrary, our method consists in identifying and classifying the bifurcations based on more features containing the bifurcations' coordinates along with their extremities. Our proposed method for bifurcation classification could easily be used in conjunction with an aneurysm detection method such as the one proposed in [41]. This would allow to automatically identify the aneurysm-bearing bifurcation. We have shown in this work that resorting to registration methods may not necessarily be the best option for vascular bifurcations labeling. Indeed, we found that the overall performances of all ML algorithms were significantly degraded when applied onto the features collected from registered images compared to the features issued from the LDA without any registration step.

Although our presented results may shed some light on the effect of dimensionality reduction methods in labeling the bifurcations instead of going through a registration step, our research also has some limitations. First, the image data set used in this work was small, and thus, required using LOOCV. In future studies, we plan to estimate classification performances based on MRA images with

relatively larger sizes. Second, in this work, dimensionality reduction was used to cluster various geometric features of vascular bifurcations. However, although this helps to provide more efficient combinations of the said features for better discrimination, it remains quite challenging to determine which features are best for the task at hand. Perhaps Feature Selection Techniques could be useful and provide different results. Thus, we plan to perform our detailed labeling using features selection methods in the future. Third, Some of the newer Artificial Intelligence approaches, such as Deep Learning, CNN, RNN and CapsNet were ignored because the dataset was fairly simple and easy to process with standard ML algorithms. Therefore, we also plan to investigate some Deep Learning approaches for further improvement.

## References


[1] D. Robben, E. Türetken, S. Sunaert, V. Thijs, G. Wilms, P. Fua, F. Maes, P. Suetens, Simultaneous segmentation and anatomical labeling of the cerebral vasculature, Medical Image Analysis 32 (2016) 201–215. doi:10.1016/j.media.2016.03.006.

[2] M. Sánchez van Kammen, C. Moomaw, I. Schaaf, R. Brown, D. Woo, J. Broderick, J. Mackey, G. Rinkel, J. Huston, Y. Ruigrok, Heritability of circle of willis variations in families with intracranial aneurysms, PLOS ONE 13 (2018) e0191974. doi:10.1371/journal.pone.0191974.

[3] T. Sichtermann, A. Faron, R. Sijben, N. Teichert, J. Freiherr, M. Wiesmann, Deep learning-based detection of intracranial aneurysms in 3d tof-mra, American Journal of Neuroradiology 40. doi:10.3174/ajnr.A5911.

[4] K. S. Lee, J. Zhang, V. Nguyen, J. Han, J. Johnson, R. Kirollos, M. Teo, The evolution of intracranial aneurysm treatment techniques and future directions, Neurosurgical Review 45 (2022) 1–25. doi:10.1007/s10143-021-01543-z.

[5] R. Pascalau, V. Padurean, D. Bartos, A. Bartos, B. Szabo, The geometry of the circle of willis anatomical variants as a potential cerebrovascular risk factor, Turkish Neurosurgery 29. doi:10.5137/1019-5149.JTN.21835-17.3.

[6] A. Nouri, 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.

[7] H. Bogunovic, et al., Anatomical labeling of the circle of willis using maximum a posteriori probability estimation, IEEE transactions on medical imaging 32.

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

[9] S. A. Iqbal, Comprehensive study of the anatomical variations of the circle of willis in adult human brains, J Clin Diagn Res. 11 (2013) 2423–7.

[10] D. Robben, S. Sunaert, V. Thijs, G. Wilms, F. Maes, P. Suetens, Anatomical labeling of the Circle of Willis using maximum a posteriori graph matching, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 8149 LNCS (PART 1) (2013) 566–573. doi:10.1007/978-3-642-40811-3_71.

[11] I. Essadik, A. Nouri, N. Lauzeral, R. Touahni, R. Bourcier, F. Autrusseau, Combining machine learning and artery characterization to identify the main bifurcations in 3d vascular trees, in: SPIE Medical Imaging conference, 2022, p. 16. doi:10.1117/12.2605459.

[12] M. I. Todorov, J. C. Paetzold, O. Schoppe, G. Tetteh, S. Shit, V. Efremov, K. Todorov-völgyi, M. Düring, M. Dichgans, M. Piraud, B. Menze, A. Ertürk, Machine learning analysis of whole mouse brain vasculature, Nature Methods 17 (2020) 442–449. doi:10.1038/s41592-020-0792-1.

[13] B. Thompson, R. Brown, S. Amin-Hanjani, J. Broderick, K. Cockroft, E. Connolly, G. Duckwiler, C. Harris, V. Howard, s. C. Johnston, P. Meyers, A. Molyneux, C. Ogilvy, A. Ringer, J. Torner, Guidelines for the management of patients with unruptured intracranial aneurysms: A guideline for healthcare professionals from the american heart association/american stroke association, Stroke 46. doi:10.1161/STR.0000000000000070.

[14] S. Chabert, T. Mardones, R. Riveros, M. Godoy, A. Veloz, R. Salas, P. Cox, Applying machine learning and image feature extraction techniques to the problem of cerebral aneurysm rupture, Research Ideas and Outcomes 3 (2017) e11731. doi:10.3897/rio.3.e11731.

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

[16] P. Jassi, G. Hamarneh, Vascusynth: Vascular tree synthesis software, Insight Journal January-June (2011) 1–12. doi:10380/3260.

[17] N. Strominger, R. Demarest, L. Laemle, Gross Anatomy of the Brain, Humana Press, Totowa, NJ, 2005, Ch. 2, pp. 1–10. doi:10.1007/978-1-59259-730-7_1.
URL https://doi.org/10.1007/978-1-59259-730-7_1

[18] L. Fan, H. Li, J. Zhuo, Y. Zhang, J. Wang, L. Chen, Z. Yang, C. Chu, S. Xie, A. R. Laird, P. T. Fox, S. B. Eickhoff, C. Yu, T. Jiang, The human brainnetome atlas: A new brain atlas based on connectional architecture., Cereb Cortex 26 (8) (2016) 3508–3526. doi:10.1093/cercor/bhw157.

[19] W. L. Nowinski, Human brain atlasing: past, present and future, The Neuroradiology Journal 30 (6) (2017) 504–519, pMID: 29096577. arXiv:https://doi.org/10.1177/1971400917739274, doi:10.1177/1971400917739274.
URL https://doi.org/10.1177/1971400917739274

[20] T. Dunås, et al., A Stereotactic Probabilistic Atlas for the Major Cerebral Arteries, Neuroinformatics 15 (1) (2017) 101–110.

[21] S. Ota, 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. doi:10.1117/12.771834.

[22] M. Bilgel, et al., Automated anatomical labeling of the cerebral arteries using belief propagation, Proc. of SPIE 866918.

[23] X. Wang, Y. Liu, Z. Wu, X. Mou, M. Zhou, M. A. Ballester, C. Zhang, Automatic labeling of vascular structures with topological constraints via HMM, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Vol. 10434 LNCS, 2017, pp. 208–215. doi:10.1007/978-3-319-66185-8_24.

[24] M. Zhao, G. Hamarneh, Bifurcation detection in 3d vascular images using novel features and random forest, in: 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI), 2014, pp. 421–424. doi:10.1109/ISBI.2014.6867898.

[25] J. Geng, Y. Wang, Z. Ji, W. Wang, Y. Yin, G. Yang, X. Fan, T. Li, P. Hu, C. He, H. Zhang, Advantages of 3d registration technology (3drt) in clinical application of unruptured intracranial aneurysm follow-up: A novel method to judge aneurysm growth, Journal of Neuroradiologydoi:10.1016/j.neurad.2022.08.004.

[26] A. Wang, X. Yan, Z. Wei, Imagepy: an open-source, python-based and platform-independent software package for bioimage analysis, Bioinformatics 34 (18) (2018) 3238–3240. arXiv:https://academic.oup.com/bioinformatics/article-pdf/34/18/3238/25731720/bty313.pdf, doi:10.1093/bioinformatics/bty313.
URL https://doi.org/10.1093/bioinformatics/bty313

[27] X. Zhang, W. Hao, S. Han, C.-F. Ren, L. Yang, Y. Han, B. Gao, Middle cerebral arterial bifurcation aneurysms are associated with bifurcation angle and high tortuosity, Journal of Neuroradiology 49. doi:10.1016/j.neurad.2021.12.001.

[28] J. Zang, Y. Huang, L. Kong, B. Lei, P. Ke, H. Li, J. Zhou, D. Xiong, G. Li, J. Chen, X. Li, Z. Xiang, Y. Ning, F. Wu, K. Wu, Effects of brain atlases and machine learning methods on the discrimination of schizophrenia patients: A multimodal mri study, Frontiers in Neuroscience 15. doi:10.3389/fnins.2021.697168.
URL https://www.frontiersin.org/article/10.3389/fnins.2021.697168

[29] B. Igne, R. W. Bondi, C. Airiau, Chapter 8 - multivariate data analysis for enhancing process understanding, monitoring, and control—active pharmaceutical ingredient manufacturing case studies, in: A. P. Ferreira, J. C. Menezes, M. Tobyn (Eds.), Multivariate Analysis in the Pharmaceutical Industry, Academic Press, 2018, pp. 185–210. doi:https://doi.org/10.1016/B978-0-12-811065-2.00009-6.
URL https://www.sciencedirect.com/science/article/pii/B9780128110652000096

[30] A. Tharwat, T. Gaber, A. Ibrahim, A. E. Hassanien, Linear discriminant analysis: A detailed tutorial, Ai Communications 30 (2017) 169–190,. doi:10.3233/AIC-170729.

[31] O. Shafaat, J. D. Bernstock, A. Shafaat, V. S. Yedavalli, G. Elsayed, S. Gupta, E. Sotoudeh, H. I. Sair, D. M. Yousem, H. Sotoudeh, Leveraging artificial intelligence in ischemic stroke imaging, Journal of Neuroradiology 49 (4) (2022) 343–351. doi:https://doi.org/10.1016/j.neurad.2021.05.001.
URL https://www.sciencedirect.com/science/article/pii/S0150986121001061

[32] R. Fisher, The use of multiple measurements in taxonomic problems, Annals Eugen. 7 (1936) 179–188. doi:10.1111/j.1469-1809.1936.tb02137.x.

[33] A. Starzacher, B. Rinner, Evaluating knn, lda and qda classification for embedded online feature fusion, in: 2008 International Conference on Intelligent Sensors, Sensor Networks and Information Processing, 2008, pp. 85–90. doi:10.1109/ISSNIP.2008.4761967.

[34] C. Cortes, V. Vapnik, Support-vector networks, Machine Learning 20 (1995) 273–297. doi:10.1007/BF00994018.

[35] C. Wilson, Effectiveness of feature and classifier algorithms in character recognition systems, Proceedings of SPIE - The International Society for Optical Engineeringdoi:10.1117/12.143627.

[36] B. Jijo, A. Mohsin Abdulazeez, Classification based on decision tree algorithm for machine learning, Journal of Applied Science and Technology Trends 2 (2021) 20–28.

[37] J. Ali, R. Khan, N. Ahmad, I. Maqsood, Random forests and decision trees, International Journal of Computer Science Issues(IJCSI) 9.

[38] K. Taunk, S. De, S. Verma, A. Swetapadma, A brief review of nearest neighbor algorithm for learning and classification, in: 2019 International Conference on Intelligent Computing and Control Systems (ICCS), 2019, pp. 1255–1260. doi:10.1109/ICCS45141.2019.9065747.

[39] V. Moni, S. Cumarasurier, Human papillomavirus targeted immunotherapy outcome prediction using machine learning, International Journal for Research in Applied Science and Engineering Technology (IJRASET)doi:10.22214/ijraset.2021.37197.

[40] H. Johnson, G. Harris, K. Williams, Brainsfit: Mutual information registrations of whole-brain 3d images, using the insight toolkit, The Insight

Journaldoi:10.54294/hmb052.

URL https://www.slicer.org/

[41] F. Claux, M. Baudouin, C. Bogey, A. Rouchaud, Dense, deep learning-based intracranial aneurysm detection on tof mri using two-stage regularized u-net, Journal of Neuroradiologydoi:10.1016/j.neurad.2022.03.005.