



Supervised learning to quantify amyloidosis in whole brains of an Alzheimer's disease mouse model acquired with optical projection tomography

DAVID NGUYEN,^{1,2,3,*} VIRGINIE UHLMANN,^{4,5} ARIELLE L. PLANCHETTE,^{1,3} PAUL J. MARCHAND,³ DIMITRI VAN DE VILLE,^{2,6} THEO LASSER,³ AND ALEKSANDRA RADENOVIC¹

¹Laboratory of Nanoscale Biology, École Polytechnique Fédérale de Lausanne, Lausanne, Vaud, Switzerland

²Medical Image Processing Lab, École Polytechnique Fédérale de Lausanne, Genève, Genève, Switzerland

³Laboratoire d'Optique Biomédicale, École Polytechnique Fédérale de Lausanne, Lausanne, Vaud, Switzerland

⁴Biomedical Imaging Group, École Polytechnique Fédérale de Lausanne, Lausanne, Vaud, Switzerland

⁵European Bioinformatics Institute, EMBL-EBI, Cambridge, United Kingdom

⁶Department of Radiology and Medical Informatics, University of Geneva, Genève, Genève, Switzerland

*david.nguyen@alumni.epfl.ch

Abstract: Alzheimer's disease (AD) is characterized by amyloidosis of brain tissues. This phenomenon is studied with genetically-modified mouse models. We propose a method to quantify amyloidosis in whole 5xFAD mouse brains, a model of AD. We use optical projection tomography (OPT) and a random forest voxel classifier to segment and measure amyloid plaques. We validate our method in a preliminary cross-sectional study, where we measure 6136 ± 1637 , 8477 ± 3438 , and 17267 ± 4241 plaques (AVG \pm SD) at 11, 17, and 31 weeks. Overall, this method can be used in the evaluation of new treatments against AD.

© 2019 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

Amyloidosis in brain tissues is associated with several neurodegenerative diseases, including Parkinson's and AD. In AD, toxic extracellular aggregates of a truncated and thus misfolded amyloid precursor protein form deposits known as amyloid plaques [1, 2]. These plaques have a spherical shape, and their size varies between approximately ten and one hundred micrometers. The neuropathological nature of the plaques is hypothesized to play a central role in the etiology of AD and is at the core of AD research [3]. The mechanisms of plaque formation and their consequences remain elusive and are commonly studied in genetically modified rodent models, such as the 5xFAD mouse model [4]. Such models are designed to reproduce the age-dependent amyloid deposition observed in humans [5].

The standard technique used to visualize and quantify amyloid plaques is histopathology wherein sections of brain tissue are sliced, mounted on glass cover slides, stained, and imaged with either a widefield or a fluorescence microscope. Histopathology is an invaluable diagnostic tool but has some limitations. Strict sample preparation protocols and a well-practiced and meticulous expertise are required. In addition, reproducibility across samples is challenging to obtain, and artefacts related to sample preparation are nearly unavoidable. Furthermore, three-dimensional renderings of whole brains remain prohibitively time-consuming. As a consequence, this technique only allows observing the development of amyloid plaques in local, arbitrarily chosen areas of the brain, making an objective comparison between specimen difficult. Other less invasive techniques exist to image amyloid plaque growth *in vivo*, such as two-photon microscopy [6],

optical coherence microscopy [7], and photoacoustic microscopy [8]. All three techniques can resolve individual plaques consistently while preserving brain integrity through the use of cranial windows. They can cover several square millimeters of tissue over hundreds of micrometers in depth, but fail at imaging over the whole brain. Differential phase contrast tomography [9] and contrast-enhanced magnetic resonance microscopy [10] can produce images of amyloidosis in whole intact rodent brains with a resolution of a few tens of micrometers. However, they require expensive instruments with specific sample preparation protocols and are therefore not routinely used. Ultramicroscopy [11] is another technique for visualizing amyloidosis in whole-excised mouse brains [12, 13]. It is a mesoscopic adaptation of light sheet microscopy, where a sheet of light illuminates the sample orthogonal to the detection path, thus producing optically sectioned fluorescence images of the organ with minimal photobleaching. Similarly, optical projection tomography (OPT) [14] can perform whole mouse brain neuroanatomical phenotyping [15] and amyloidosis imaging [16] by acquiring fluorescence projections of the organ at different angles over a complete turn. Although ultramicroscopy and OPT are complementary techniques, ultramicroscopy often offers a better resolution, while OPT is generally easier to implement and more robust to misalignment.

To quantify amyloidosis progression from image data, the standard protocol is to process specimen with an amyloid plaques-specific staining such that the plaques stand out from other tissue elements. A segmentation mask is then obtained by identifying which image elements correspond to amyloid plaques. From there, the number of plaques and volume they occupy can be estimated. When imaging the whole brain at once, a significant challenge is the staining of the intact organ. For amyloid plaques, Methoxy-X04 [17] is one of the only probes that can be used due to its ability to penetrate the blood-brain barrier. However, the excitation of Methoxy-X04 (in the near-UV/blue) also generates a strong signal from tissue autofluorescence. Therefore, quantitative image-based analysis of amyloidosis relying on voxel intensity thresholding, as typically used in histopathology data, cannot be performed. Amyloid plaques are indeed indistinguishable from tissue autofluorescence based on intensity, henceforth requiring a more complex image analysis approach. In the work of Jährling *et al.* [12], plaques could be isolated in ultramicroscopy images by applying intensity thresholding since small sub-volumes called cubes were considered instead of the whole brain volume. Quantitative analysis was then run through six of these cubes per sample. Statistical analyses are performed on the sub-volumes but they do not reflect the brain-wide plaque distribution.

In this paper, we propose a supervised learning pipeline relying on random forest voxel classifiers to segment and quantify amyloid plaques in whole 5xFAD mouse brains of different ages acquired with OPT.

Learning-based automated classification for AD has gained a strong interest in recent years. The disease is indeed currently diagnosed based on clinical examination, and classification of AD samples, preferably at an early stage of the pathology, holds a strong potential for adding validity to the diagnosis. Signal-based classification already allowed identifying markers to aid in AD diagnostic from diffusion tensor imaging scans [18], EEG signals [19], intracellular recordings [20], or multiple sources including various 3D imaging modalities [21].

Regarding tissue imaging data, BioVision [22] was proposed as a supervised training approach for image-based identification of histopathological objects. This approach, relying on a pixel-based Bayesian classifier, has been used for the quantification of amyloid plaques in histology images [23]. A similar approach was followed by Vandenberghe *et al.* [24] on the same type of data, showing that random forest classifiers exhibited better performance. The work of Li *et al.* [25] also illustrates the good performance of pixel-based decision tree algorithms, this time for the pixel-based classification of MRI image data. An essential aspect of these previous works is that the classification algorithm operates on pixels in 2D slices. Another way to approach the problem of identifying brain components from images is atlas-based segmentation [24, 26]. In

this setting, a labeled model (the atlas) of the 3D organ is deformed to fit the 3D image volume at once. From the resulting boundaries of each labeled regions, statistics can be derived at the local level.

As no atlas labels are available for mouse brain in OPT data, the approach we propose considers 3D voxels in image volumes without an underlying atlas. It is, to the best of our knowledge, the first attempt to introduce a learning-based method for amyloid plaques detection in OPT image data and, more generally, to quantify amyloid plaques relying on a pipeline incorporating information from all three dimensions in an atlas-free setting.

Our manuscript is structured as follows. Firstly, we briefly describe the sample preparation and imaging setup, which we introduced in our previous work [16]. Secondly, we present the supervised learning pipeline for segmentation and quantification of amyloidosis. Supervised learning provides an efficient way to learn, from a set of manually annotated examples, how image features such as *e.g.*, grayscale intensity and textures [27], should be combined to segment objects of interest in images [28]. In particular, pixel- and voxel-based random forest classifiers have proved their efficiency for bioimage segmentation tasks [29, 30]. We choose to rely on that specific type of algorithms for two main reasons: first, random forests are well-suited for classification tasks with few training data; and second, they can be used through the *ilastik* [31] interface. *ilastik* is an open-source image analysis software, which provides a user-friendly GUI for training and reusing trained classification workflows for prediction. In that way, we aim at making our image analysis pipeline easily reusable with or without additional training. Thirdly, we showcase the results of the segmentation and quantification in 5xFAD mouse brains of different ages. Finally, we discuss the impact and limitations of our study and give some recommendations based on our preliminary results.

2. Materials and methods

2.1. Sample preparation

Experiment animals, 5xFAD mice [4] on a B6SJL F1 background, were generated by crossing 5xFAD transgenic females on a C57BL/6J background (34848-JAX, MMRRC) with wild type SJL/J males (000686, The Jackson Laboratory) in the local animal facility (housing conditions are described below). The disease phenotype is observed to be more robust on this hybrid background [32]. The 5xFAD mouse model is commercially-available and widely-used in AD research. The mice develop a severe amyloid pathology starting around 1.5 months with high levels of accumulation in the subiculum [4]. Animal procedures were carried out according to Swiss regulations [33] (animal protection ordinance 455.1) under the approval of the veterinary authority of the canton of Vaud (license: VD3058), and all efforts were made to minimize suffering, following the principle of the 3Rs [34]. Experiment animals were housed in ventilated cages (maximum 5 animals per cage) under a 12h light/dark cycle (lights on at 7 a.m.) and controlled atmosphere (23 °C and 50% relative humidity) with *ad libitum* access to food and water. Two 5xFAD transgenic females on a C57BL/6J background were housed with one wild type SJL/J male in the breeding cages under the same conditions as experiment animals, but the extreme aggression of SJL/J males [35] hampered the breeding capacity. Indeed, 5 out of 12 females in the 6 breeding cages (that we shared with other experimentalists) were found dead with lethal lesions from fights with the males. To refine our breeding protocol, we decided to move the males in individual cages when the females were pregnant. However, this measure was not sufficient to prevent casualties in the breeding cages, and it reduced the number of experiment animals that could be generated, more details can be found in the discussion section.

Brains were processed as described previously [16]. In summary, the mice were injected with Methoxy-X04 (Tocris) [17], a fluorescent marker that targets amyloid plaques. Then, they were deeply anesthetized using an overdose of pentobarbital and perfused with a 10% formalin solution. The brains were extracted and post-fixed in formalin overnight. Each organ was mounted in

a 1.5% agarose gel and cleared using BABB before imaging. An illustration of these steps is shown in Fig. 1(a).

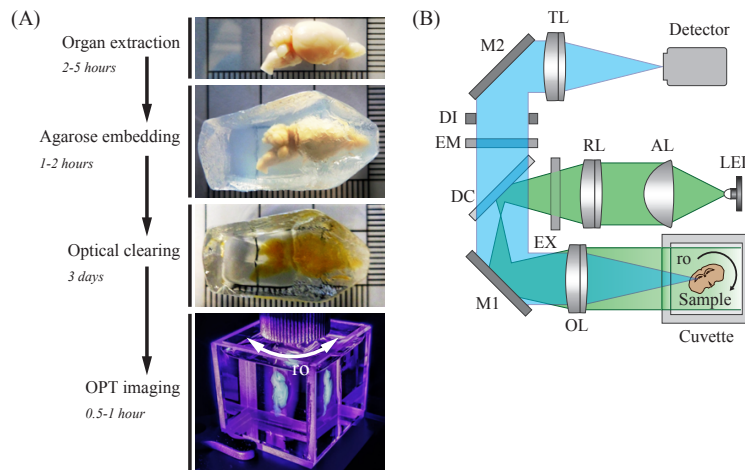


Fig. 1. Sample preparation and experimental setup. (A) Photographs of a mouse brain through the different steps of the preparation protocol. (B) Scheme of the optical projection tomography setup. AL, aspheric lens; RL, relay lens; EX, excitation filter; DC, dichroic mirror; M1/2, mirrors; OL, objective lens; EM, emission filter; DI, diaphragm; TL, tube lens; ro, rotation.

Ten experiment animals could be retrieved from the breedings and were divided into three age-groups after amyloidosis onset (Tg/0): a young one at 11 weeks old ($n = 3$), a middle-aged one at 17 weeks old ($n = 4$), and an old one at 31 weeks old ($n = 3$). A fourth control group ($n = 2$) was formed with B6SJL F1 mice (+/+) that did not have the transgenes. The control group followed the same sample preparation as the experiment groups, including Methoxy-X04 injection. All samples and their corresponding group are recapitulated in Table 1. As indicated in the third column, some samples are composed of only one hemisphere of the brain, as some specimen had to be shared with other experimentalists.

2.2. Optical projection tomography

Whole brain imaging was performed with a custom OPT setup [16] shown in Fig. 1(b). The instrument is similar to an epifluorescence microscope. It uses a 420-nm LED light source and a 300-mm achromat objective lens (OL) to produce fluorescence projections of the sample with a 0.5X magnification. The fluorescent light is filtered using a custom filter set from Chroma composed of a dichroic (DC) mirror (AT455dc), an excitation (EX) filter (AT420/40x), and an emission (EM) filter (AT465lp). From numerical simulations, the intensity in the sample plane has been estimated to be approximately of 0.25 mW/mm^2 . A diaphragm (DI) in the rear focal plane of the OL enlarges the depth of field to guarantee a sharp focus through the front half of the sample, as suggested by Sharpe *et al.* in their original work introducing OPT [14]. Projections are acquired over 360 degrees by steps of 0.3 or 0.9 degrees in approximately five minutes. The three-dimensional reconstruction of the sample is achieved by applying a filtered back-projection [36] to these projections. To do so, we perform the same procedure as described previously in [16], and refer the reader to the latter work for more details. The reconstruction procedure includes a convolution-based method to retrieve the center of rotation in the projections. Such an approach is crucial to avoid reconstruction artefacts, as precise knowledge of the position of the center of rotation in the projections is essential for an accurate filtered back-projection

Table 1. **List of samples processed.** The * indicates samples, which were partially annotated for the training of the random forest classifiers. The † indicates samples, which were used to generate Fig. 4. Tg/0, transgenic animals; +/+, wild type animals.

Id	Age [weeks]	Structure	Genotype	Image size [px ³]	Group
0 [†]	11	Whole	Tg/0	407x344x514	Young
1*	11	Hemisphere	Tg/0	283x288x517	
2	11	Hemisphere	Tg/0	268x321x569	
3	17	Hemisphere	Tg/0	300x274x590	Middle-aged
4	17	Hemisphere	Tg/0	317x243x650	
5 [†]	17	Whole	Tg/0	446x307x602	
6	17	Whole	Tg/0	442x320x577	
7 ^{*,†}	31	Whole	Tg/0	420x357x603	Old
8	31	Whole	Tg/0	411x389x597	
9	31	Whole	Tg/0	427x384x615	
10	7	Whole	+/+	352x340x512	Control
11 [†]	31	Whole	+/+	432x334x576	

reconstruction [37]. In this configuration, the OPT setup has an isotropic pixel-limited resolution of approximately 50 μm over the whole organ, due to the physical pixel size of the detector. The current resolution does not allow to measure small plaques (10-50 μm). However, the intense brightness of the plaques still makes them detectable for counting.

2.3. Image segmentation and quantification

The overall image-based segmentation and quantification process from reconstructed projections is depicted in Fig. 2. Once 3D image volumes are reconstructed with a filtered back-projection algorithm, images are normalized using Fiji [38] and its contrast enhancement tool before segmentation to accommodate for differences in dynamic range stemming from variations in, *e.g.*, image acquisition time. During normalization, 0.1% of the total amount of voxels (stack histogram) is allowed to be saturated in the image to account for outliers. This normalization step is crucial for the quality of the subsequent segmentation of amyloid plaques. As voxel intensity is part of the feature set considered by our proposed supervised learning approach, all image volumes must be brought to a comparable dynamic range, both in the training and prediction phases.

We segment plaques from background and brain image elements using a supervised learning approach. To do so, we designed a 3D voxel classification workflow in the open-source image analysis software ilastik [31]. The image analysis pipeline consists of two steps. First, training is performed from manual annotations of a few voxels in a small selection of image data. Then, the trained supervised learning algorithm predicts the type of the remaining unannotated voxels in these image data, as well as the type of each voxel for the ones left aside during training. In our case, random forests [39] are used as learning algorithm. The working principle of this approach

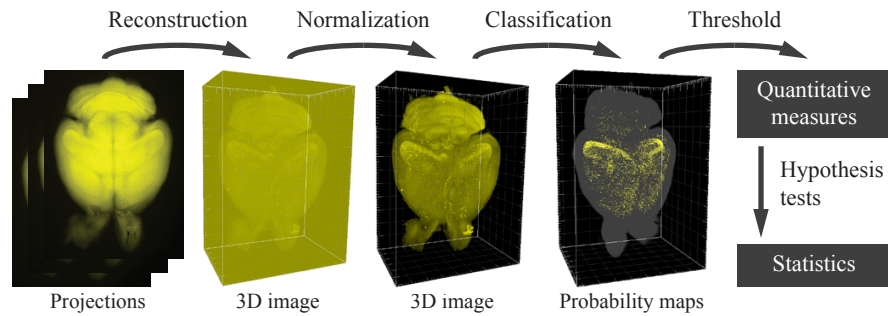


Fig. 2. **Processing pipeline.** From OPT acquisition, the projections are used to reconstruct a 3D image with a filtered back-projection algorithm. After a normalization step, a random forest classifier identifies voxels corresponding to plaques in the image volume. Quantitative measurements can be extracted from the segmentation mask, and statistical analyses can be conducted.

is as follows. A set of numerical measurements, called features, is computed for each object to be classified (in our case, the voxels). These measurements aim at describing the object in terms of, *e.g.*, color, shape, neighborhood and serve to distinguish objects belonging to different classes (in our case, plaques versus everything else). The specific nature of the features we rely on is described hereafter.

In the training phase, given (a) a collection of objects, (b) their feature values, and (c) their known class labels, several decision trees are built by randomly picking features and searching for proper decision boundaries to separate between classes. Starting from the root, each intermediate node in a decision tree corresponds to a specific feature, while leaves correspond to class labels. In the prediction phase, given an object and its collection of computed features, each tree is explored to predict the class as follows: the value of the feature corresponding to the root node is examined, and dictates which branch to follow. This process is repeated at each intermediate node until a leaf is reached, which predicts the class of the object. The final class probability estimates are obtained as the percentage of trees predicting the object to belong to each considered class. The overall classification procedure with random forests is illustrated in Fig. 3.

The choice of a specific supervised learning algorithm among the vast variety of those available is generally dictated by the constraints that are inherent to the considered problem. In our case, following the principles of the 3Rs, we want our approach to be able to train based on a small number of samples. Moreover, we aim at maximizing the reusability of our pipeline, which implies limiting handcrafted steps and relying on well-established processing environments. These considerations guided our choice for random forests to segment plaques. Random forests indeed provide an excellent trade-off between generalization capability and ability to learn from few training examples [40]. This last point specifically makes them preferred over the currently popular convolutional neural networks (CNN), which require an extensive amount of labeled data for training. The small amount of samples at our disposal and the unavailability of similar datasets make deep networks inappropriate for this problem. To validate this claim, we included CNN in our performance analysis in Section 3.2. In addition, random forests are available through the *ilastik* GUI, which maximizes user-friendliness during the training phase and ease of use when predicting on new data.

In the training phase, manual annotations were provided for each of the three classes, here corresponding to background, brain, and plaques. Our underlying assumption motivating this setting is that, although plaque signal cannot easily be separated from autofluorescence signal from brain tissues using a simple intensity-based threshold, it does stand out when considering

a combination of texture, intensity, and shape measurements. The random forest classifier, therefore, acts as an adaptive way to learn a complex threshold combining these different features to isolate plaques from the rest of the image volume content. The specific voxel features we rely on were computed through the 3D volumes. They can be grouped into three categories and are as follows:

- *Intensity-based features*: raw image and Gaussian smoothing;
- *Edge-based features*: Laplacian of Gaussian (LoG), Gaussian gradient magnitude, difference of Gaussian;
- *Texture features*: structure tensor eigenvalues, Hessian of Gaussian eigenvalues.

All features were computed at various image scales; that is, on the original image volume as well as on the image volume processed by various levels of Gaussian smoothing, namely $\sigma = 0.3, 0.7, 1, 1.6, 3.5, 5, \text{ and } 10$.

When relying on learning-based approaches, the selection of training elements, which are used as ground truth, is of major importance as it influences the generalization capability and overall performance of the algorithm. We selected samples 1 and 7 (see Table 1) for training. These two image volumes, one corresponding to a young and one to an old specimen, were chosen to avoid age-induced batch effects [41]. A few manually labeled voxels were distributed through the whole volume along the X, Y and Z planes to minimize training bias and reduce the risk of overfitting. Once trained, the pipeline was used to predict plaques in all remaining dataset. To assess the false positive rate of the random forest classifier trained in this way, we also ran the pipeline on brain image volumes of negative control specimen to report the amount of erroneously predicted plaques. Results are provided in Table 4 and testify the specificity of the method.

We then reused the same background and brain voxel labels but removed the plaque class to train a 2-class random forest voxel classifier and retrieve a whole brain segmentation mask. This second step allows us to compute an estimate of total brain volume, which is used to extract the ratio of plaque volume to the whole brain.

Due to data size (see Table 1 for details), all considered image volumes were first converted from .tif stacks to .h5, processed with ilastik, and converted back from .h5 to .tif for visualization purpose.

3. Results

3.1. Imaging of amyloid pathology progression

Selected three-dimensional brain images from the three groups of mice are shown in Fig. 4(a-c). As reported previously [16], the emission signal of Methoxy-X04 is deeply mixed with tissue autofluorescence. Therefore, both the amyloid plaques and brain anatomy can be visualized in a single OPT acquisition. Amyloid plaques can be identified from the strong fluorescent signal they emit as well as from their characteristic small spherical shape. As expected, strong age-dependent amyloidosis is observed in the subiculum, indicated by the white arrows. These images already give us a qualitative feeling for the progression of the amyloid pathology, which worsens with age. However, as previously mentioned, thresholding based solely on voxel intensity as often performed on histopathology data is inefficient for quantitative analysis. The cerebellum, indicated by the blue arrows, indeed exhibits similar voxel intensity levels as plaques, possibly caused by different autofluorescence contrast mechanisms. Likewise, the cortical barrel fields (white arrowheads), which are dense anatomical features of the mouse brain, reach high levels of autofluorescence intensity. Ultimately, perfusion artefacts, such as blood-containing vessels (blue arrowheads), also corrupt the outcome of intensity threshold-based segmentation due to the strong autofluorescence of hemoglobin. Therefore, correct isolation of amyloid plaques from

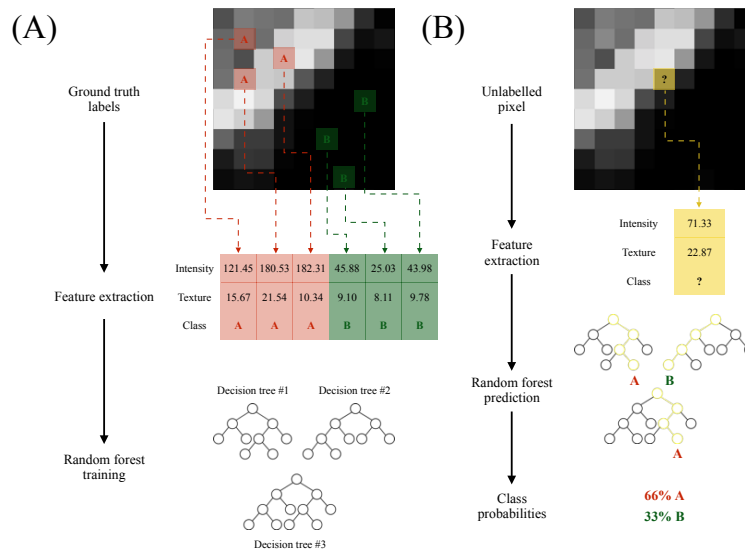


Fig. 3. Classification procedure with random forests. For the sake of simplicity, 2D pixels and a 2-class problem are depicted. (A) Training phase. A few voxels from each classes (here **A**, the object, and **B**, the background) are manually labelled as ground truth. A collection of measurements, referred to as features, are extracted to characterize these voxels. Here, only two (intensity and texture-based features) are shown. A forest of decision trees is then constructed. Each stage of a decision tree is built by randomly picking a feature and choosing a boundary value that best separates each class according to this measurement. Here, three trees are shown. (B) Classification, or prediction, phase. For each unlabeled voxel, the same set of features as in training phase are extracted. These measurements are then fed to the random forest trees, which provide class prediction. The final probability is retrieved as the percentage of trees predicting each class.

the whole brain requires to rely on a combination of several visual aspects such as their texture, shape, and intensity. A brain image from the control group is shown in Fig. 4(d). While there is a strong signal from nervous tissues, there is an absence of amyloid plaques.

3.2. Quantification of the amyloid plaques

The outcome of the voxel classification workflow described above provides an efficient way to segment amyloid plaques based on a mixture of their visual features. For each voxel in an OPT image volume of a diseased brain, the 3-class random forest classifier returns a probability value (0-1), yielding a 3D probability map image. The value of each voxel in the probability map image corresponds to the likelihood that the corresponding voxel in the original image belongs to an amyloid plaque. A segmentation mask isolating amyloid plaques is thus obtained by thresholding the 3D probability map at 0.5. This threshold can be directly interpreted as retaining only voxels with a higher-than-50-percent chance of belonging to a plaque. We found it to be sufficiently sensitive for plaque detection against the strong background while maintaining a certain level of specificity, which is discussed at the end of this section. The same procedure can be carried out to segment brain anatomy relying on the probability map output of the 2-class random forest classifier and a threshold of 0.7. The sensitivity to brain tissues is indeed observed to be higher due to a consistency in the autofluorescence signal, allowing us to increase the threshold value to gain more specificity. Segmentation results of the selected brain images from the three age-groups are shown in Fig. 4(e-h). The transparency of the brain anatomy channel

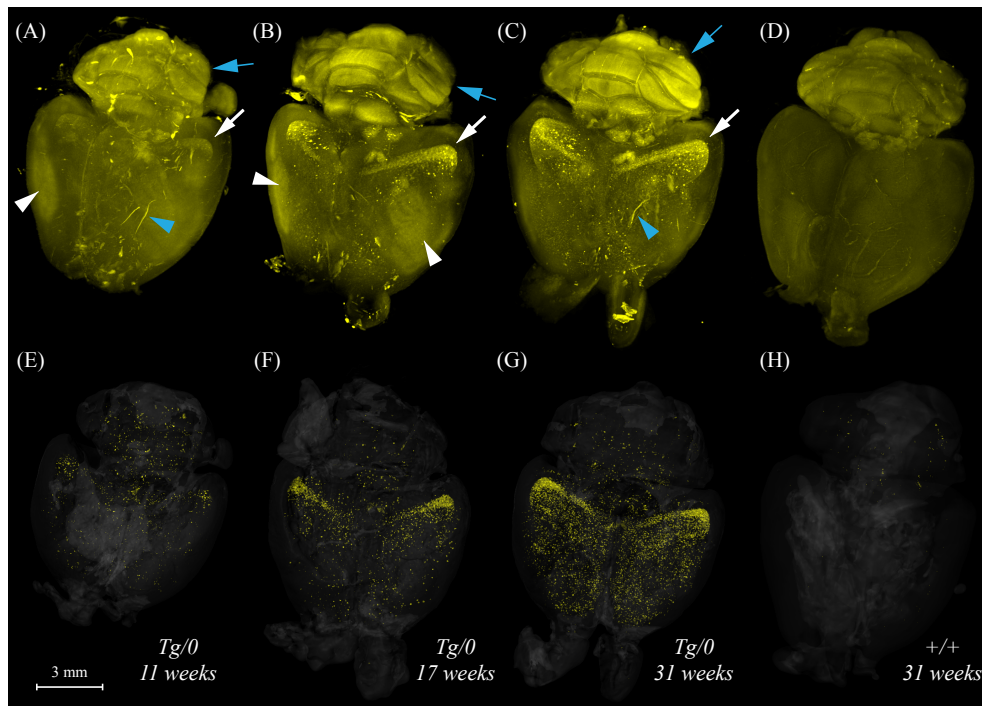


Fig. 4. Imaging of amyloid pathology progression. (A-C) Renderings of OPT images from the three age-groups: young, middle-aged, and old, respectively. Blue arrows, cerebellum; White arrows, amyloid plaques in the subiculum; Blue arrowheads, blood vessels; White arrowheads, cortical barrel fields. (D) Rendering from the control group. (E-H) Corresponding renderings after random forest classification and thresholding. The amyloid plaques, in yellow (thresholded at 0.5), are overlaid with the brain anatomy, in grey (thresholded at 0.7), whose transparency is reduced for visualization purposes.

was reduced to enhance the visibility of amyloid plaques. As all raw image volumes along with random forests predictions of brain anatomy and plaques are freely available in [Dataset 1](#) [42], readers interested in exploring the details of brain anatomy are encouraged to retrieve the data and create renderings that match their interests. As expected, we observe an increase in the number of plaques with age. For comparison purpose, we illustrate in Fig. 5 the difference between the best attainable plaque segmentation image obtained by a fine-tuned thresholding of the image (Fig. 5(b)) and the random forests result (Fig. 5(c)). Direct intensity thresholding (DIT) does not allow to isolate plaques from the autofluorescence signal of blood vessels (blue arrowheads) and the cerebellum (tan arrowheads), as well as other kinds of artefacts (purple arrowheads) whose size and shape cannot be attributed to amyloid plaques. The threshold was here fine-tuned so as to minimize background signal, excluding a large amount of fainter plaques.

Moreover, the classifier seems to have captured the different features that make amyloid plaques apparent in OPT images. For example, the cerebellum is properly excluded as it does not appear to show amyloid deposition, which is in agreement with known results [4]. Despite its strong fluorescence intensity, it does not meet the other visual criteria to be associated with the amyloid plaque class. The same observations hold for the cortical barrel fields and perfusion artefacts.

After segmentation, we perform a quantitative analysis of the amyloid plaques in the brain by computing the ratio of plaque volume to the total organ volume. This measure is referred to as plaque load and is expressed in percentage. Additionally, the total plaque count provides

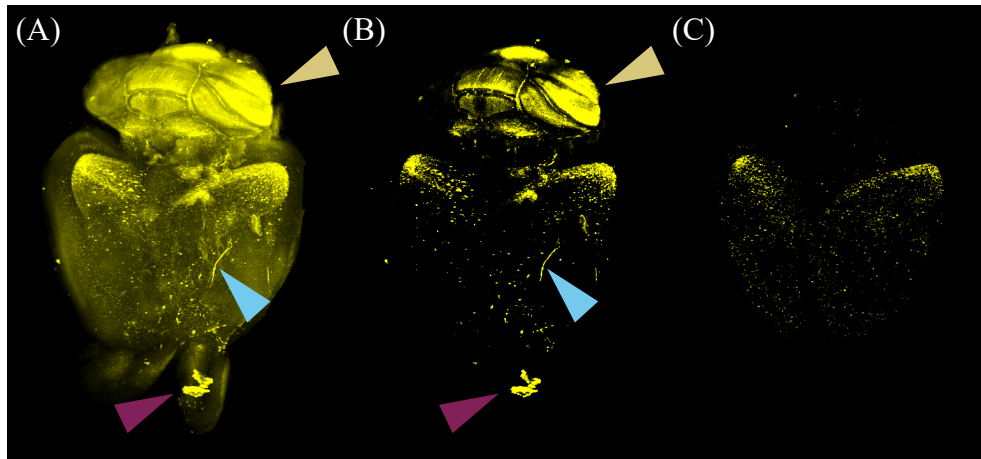


Fig. 5. Visual comparison of standard thresholding with random forests. (A) Rendering of an OPT image from an old mouse brain. (B) Corresponding rendering after thresholding voxel intensities at 70 % of the bit depth. DIT fails to isolate plaque signal from other brain structures, and results in several faint intensity plaques to be missed. Blue arrowheads, blood vessels; Tan arrowheads, cerebellum; Purple arrowhead, undetermined artefact. (C) Corresponding rendering of random forest predictions thresholded at 0.5.

another quantitative measure of amyloidosis. Regions of dense deposition, such as the subiculum, exhibits the formation of clusters of amyloid plaques in OPT images due to the resolution of the instrument and the normalization step. To refine our estimate of plaque count, we measured the average size of a single plaque and divided the total area of all larger plaque clusters by this value in each sample. Since small plaques have a size comparable to the instrument resolution, we might overestimate the plaque load.

Additionally, as mentioned previously and reported in Table 1, some samples of our study are single brain hemispheres. To allow for a fair comparison of the total number of plaques with full brain samples, we reported twice their measured plaque count. This choice is motivated by current observations of brain amyloidosis, which suggest a symmetry of the pathology between brain hemispheres [43]. To further validate this hypothesis, two 3D images of whole 31-week-old brains (samples 7 and 8 in Table 1) were digitally split and the number of plaques was counted in each hemisphere. A total amount of 8418 (left hemisphere)/7467 (right hemisphere) plaques for sample 7 and 10764 (left hemisphere)/9257 (right hemisphere) plaques for sample 8 were obtained respectively, corresponding to total plaque proportions of 0.53 (left hemisphere)/0.47 (right hemisphere) for sample 7 and 0.54 (left hemisphere)/0.46 (right hemisphere) for sample 8. We thus believe that assuming that the plaque count in a single hemisphere corresponds to half of the total plaque count in the whole brain is a reasonable hypothesis. However, a comprehensive statistical analysis of the regional deposition of amyloid plaques in the 5xFAD model should be performed prior to drawing biological conclusions. As such a study would require a much larger sample size than our resources allowed, we leave it to future work. The results of the quantitative analysis of all samples are summarized in Table 4.

Providing results from alternative methods for a reasonable comparison to our approach appears to be challenging. Existing solutions for the automated image-based assessment of amyloid plaques are indeed designed for imaging modalities of entirely different nature (*e.g.*, histopathology or MRI) and, most importantly, trained at a pixel (and not voxel) level. The following works relate to ours and are relevant comparison points, although focusing on different imaging modalities. Iordanescu *et al.* [44] proposed a machine learning approach to amyloid

plaque segmentation in MR images based on support vector machines using handcrafted features. As the features are derived from MR image intensity only, this approach is unadapted to our data, as seen from the problems highlighted in Fig 5. In histopathology images, Teboul *et al.* [45] presented an intensity thresholding-based method to detect amyloid plaques, but this approach suffers from the same limitations as the previous one. Kuan *et al.* [46] and Whitesell *et al.* [47] studied the distribution of amyloid plaques in whole mouse brain images at the mesoscopic scale relying on ad-hoc thresholding based on intensity signal from different channels, followed by a morphometric classifier. Our random forests approach combines these two steps, without the need for any hyperparameter tuning and based on few manual annotations. The work of Jähring *et al.* [12] already mentioned in Section 1 is the most sensible state-of-the-art reference for the problem we focus on. Their method relies on DIT followed by manual corrections to remove artefacts such as blood vessels. We follow the same idea but replaces the correction step by a few manual annotations for training, making the approach more automated and less dependent on parameter tuning. This, in particular, allows us to perform plaque quantification in the entire brain, as opposed to small sub-volumes.

In order to evaluate the performance of our method and validate our design choices, we obtained manual annotations for the brain and plaque elements, serving as gold standard. As our resources did not allow for obtaining a labeling of amyloid plaques in a complete 3D OPT image volume, we selected 18 slices (2D) from three different datasets to be manually annotated by an expert (brain tissues and amyloid plaques). We enforced that our selection contained slices from different image volumes so as to ensure the reliability of the evaluation and assess the generalization capability of our approach.

We carried out a performance analysis of the random forest predictions for brain and plaque segmentation. We provide a comparison with most commonly used methods to approach the plaque identification problem in other modalities, namely DIT and LoG, and with CNN, which is the state-of-the-art approach for segmentation tasks in general. As LoG is a blob-detection approach, it is not suited for brain anatomy and is considered for plaques only. LoG filtering was performed with a 3D Gaussian detector with a standard deviation tuned to the average plaque size. Patch-based CNN were designed following the popular U-Net architecture [48]. The same images and labels used to train random forests were split into patches that were augmented with rotations, shifts, shear transforms and flip transformations. Following the same procedure as for random forests, a network was first trained to identify amyloid plaques, and another one was trained to identify brain anatomy. To ensure convergence in spite of the extreme sparsity of labels, we used a pixel-weighted soft-max cross-entropy loss [49] to optimize class-balancing, and trained for 10 epochs. Proper training of the networks was ensured by monitoring the evolution of the validation loss over successive epochs. Plots and Pickle files containing each network training history are provided as supplementary material in [Dataset 1](#) [42] for the interested reader.

We provide a receiver operating characteristic (ROC) analysis of each method in Fig. 6. ROC curves are built by reporting the true positive rate against the false positive rate for all possible threshold values within an image dynamic range. Each point in the ROC curve thus corresponds to what can be achieved with one specific threshold value, and the corresponding areas under the ROC curves (AUC) is an estimate of the global accuracy of each method. AUC values are provided in Table 2. In addition to ROC analysis, we reported the accuracy, sensitivity, specificity,

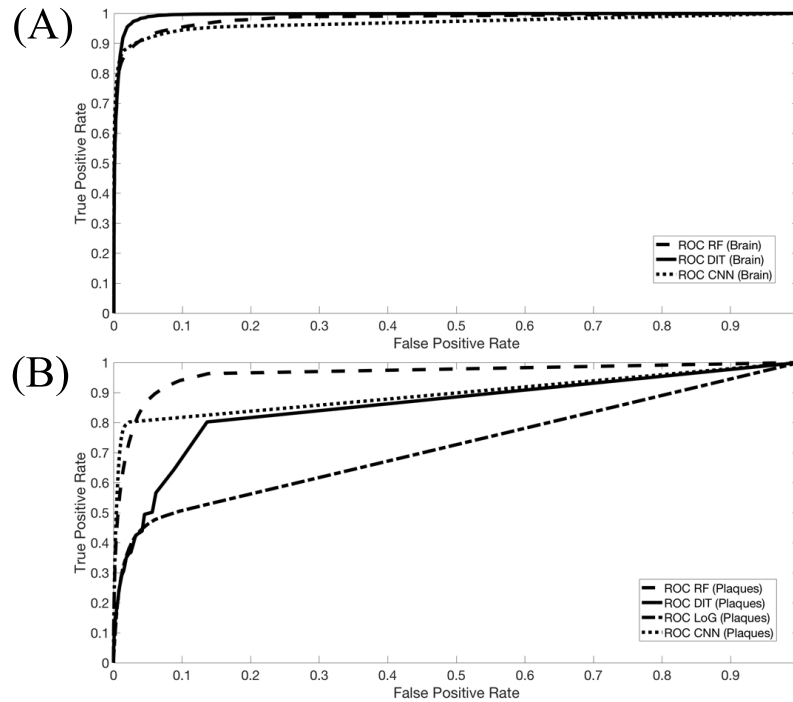


Fig. 6. **Performance evaluation of the random forest classifier.** Comparison of ROC analysis for (A) brain voxels identification and (B) plaque voxels identification using the random forest predictions (RF, dashed line), and competing approaches such as direct intensity thresholding (DIT, solid line), Laplacian of Gaussian filtering (LoG, dash-dotted line), and convolutional neural networks predictions (CNN, dotted line).

and Dice index for each considered approach. These quantities are defined as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN},$$

$$\text{Sensitivity} = \frac{TP}{TP + FN},$$

$$\text{Specificity} = \frac{TN}{TN + FP},$$

$$\text{Dice} = \frac{2TP}{2TP + FP + FN},$$

where TP stands for true positives, TN true negatives, FP false positives, and FN false negatives. In contrast to the ROC curve, these metrics were obtained for specific thresholds chosen as follows. For DIT, the optimal value for brain segmentation was obtained using Otsu's method on each considered dataset. For plaque segmentation, an optimal threshold value was identified by visual assessment for each dataset. The optimal value for LoG was computed as $\mu + \alpha * \sigma$ for each dataset, where μ and σ are the mean and standard deviation of the LoG filtered dataset, respectively, and α is an integer manually set by visual assessment. For probabilities predicted either by random forests or CNN, since the considered values directly reflected prediction certainty, thresholds were fixed according to the difficulty of the task, namely 70% (0.7) for brain anatomy and 50% (0.5) for amyloid plaques. Results are reported in Table 3.

Random forest classifier performance is on par with DIT when it comes to isolating brain

Table 2. **Performance evaluation of the random forest classifier.** Comparison of the area under the curve (AUC) corresponding to the ROC curves presented in Fig. 6 for brain and plaque identification using the random forest predictions (RF), and competing approaches such as direct intensity thresholding (DIT), Laplacian of Gaussian filtering (LoG), and convolutional neural networks predictions (CNN).

	DIT	LoG	CNN	RF
Brain	0.99	N/A	0.96	0.99
Plaques	0.92	0.75	0.89	0.96

Table 3. **Performance evaluation of the random forest classifier.** Comparison of the accuracy, sensitivity, specificity and dice metrics for brain and plaque identification using the random forest predictions (RF), and competing approaches such as direct intensity thresholding (DIT), Laplacian of Gaussian filtering (LoG), and convolutional neural networks predictions (CNN).

		Accuracy	Sensitivity	Specificity	Dice
Brain	DIT	0.94	0.86	0.99	0.91
	CNN	0.91	0.73	0.99	0.84
	RF	0.94	0.87	0.98	0.91
Plaques	DIT	0.80	0.49	0.84	0.35
	LoG	0.46	1.00	0.40	0.27
	CNN	0.21	0.07	0.59	0.12
	RF	0.96	0.88	0.97	0.83

anatomy from the background, which is expected from the strong contrast making this step a relatively easy segmentation task. CNN also performs reasonably well, and would probably reach equivalent performance with more training annotations. However, when considering the segmentation of amyloid plaques, random forest predictions significantly outperforms all other methods. DIT offers high specificity but has a high false negative rate, which translates into a low sensitivity. The accuracy of DIT is boosted by its good capability in identifying true negatives, but its actual performance in detecting plaques is quite poor as seen from the low Dice index. LoG exhibits poor discrimination power for identifying plaques. Although it misses none of them (yielding a false negative count of zero and a maximal sensitivity), it obtains low specificity, low accuracy, and even worse Dice index, indicating that lots of false positives are captured. CNN performs poorly most likely due to the sparsity of labels in the training set, and would probably yield better results with more annotations, which would, however, come at a cost in terms of manual labour. Moreover, setting up CNN requires a significant amount of hyperparameter tuning, script writing, and data pre-processing. In contrast, the ilastik interface allows training random forests models with zero coding expertise, no hyperparameter tuning, and few annotations that can easily be created as brush strokes through the interactive GUI.

Arguably, a valid estimate of brain volume could be obtained by segmenting brain anatomy

using DIT and combined to the amyloid plaque segmentation outcome provided by random forests to compute the plaque load. This would, however, require identifying an appropriate intensity threshold value to segment brain anatomy for each individual dataset (*e.g.*, using Otsu's method). As amyloid plaque segmentation is achieved by training a 3-class (background-brain-plaque) random forests classifier, the 2-class (brain-background) random forest classifier used for brain anatomy segmentation is simply obtained by removing the plaque class, without adding additional training labels. Random forests predictions for brain anatomy, therefore, come at no extra cost and provide equivalently good results. It however offers a way to segment brain anatomy that is better suited to batch processing: the threshold used to segment is then a common value to all datasets, that directly relates to an interpretable quantity, namely the minimum acceptable probability for a voxel to be considered as part of the brain. For all considered methods, the lower performance observed for plaque versus brain anatomy detection can be explained by the small nature of plaque elements: as they are composed of few voxels, minor disagreements between expert annotation and automated segmentation results are severely penalized.

All in all, random forests through ilastik pragmatically appear to be the solution offering the best compromise between simplicity and performance for the task of segmenting amyloid plaques in OPT images.

Since our method has been designed to be easily reusable, all results provided in this paper can be directly reproduced by importing the ilastik pipelines, which holds all the necessary information regarding algorithmic parameters, and running it on the raw data (image analysis pipeline, raw data and training labels are available in [Dataset 1](#) [42], including manual ground truth). Annotations provided for training can be visualized and the algorithm re-trained from additional user-provided brush strokes on images through the ilastik's GUI.

Ultimately, we performed a statistical analysis of our results, illustrated with boxplots in Fig. 7. For the plaque load, there is no statistically significant difference between groups as determined by one-way ANOVA ($F(2, 7) = 4.733, p = .05$). However, we still performed a post-hoc Tuckey comparison since $p = \alpha = .05$, and we found a statistically significant difference between the young and old group ($p < .05$) (with a 5% chance of it being a false positive). Moreover, a statistically significant difference in total plaque number is observed by one-way ANOVA ($F(2, 7) = 9.609, p = .01$). Tuckey's test showed that the old group differs significantly ($p < .05$) from the middle-aged and the young group. Shapiro-Wilk normality tests have been performed on each group for the plaque load and total number of plaque. None of these tests rejected the null hypothesis that the data are normally distributed ($p < .05$), which justifies the statistical tests performed in our analysis.

As a quality control, we performed the quantitative analysis on negative controls (Table 1). Both the plaque load and plaque count show a reasonably low amount of false positives (Table 4), demonstrating the specificity of our analysis.

4. Discussion

The segmentation of amyloid plaques from whole images of 5xFAD mice brains acquired with OPT shows promising results, as seen in Fig. 4(e-h). However, limited statistically significant differences are drawn from the group analysis. We believe that this restricted statistical significance, observed in spite of the visually encouraging segmentation outcome obtained with our method, can be attributed to three different factors: the plaque load, number of plaques and number of samples. We now describe in more details how these three aspects affected results in our experimental setting, and propose solutions to minimize their adverse effect in further experimental designs.

Firstly, plaque load is a standard measure of amyloidosis in sections with high plaque density from histopathology data. In our experimental design, sample calculation is based on the findings of Bolmont *et al.* [7], who observed a difference of approximately 1.6 percent in the plaque

Table 4. **Quantitative measurements of amyloid plaque quantity in all samples.** The * indicates samples which were partially annotated for the training of the random forest classifiers. The † indicates samples, which were used to generate Fig. 4. The mean and standard deviation (SD) is included for each measure and group, which are compared in this study. The control group serves as quality control.

Id	Plaque load [%]		Total number of plaques		Group
0 [†]	0.05	Mean: 0.11 SD: 0.07	4595	Mean: 6136 SD: 1637	Young
1*	0.09		5958		
2	0.19		7854		
3	0.42	Mean: 0.25 SD: 0.14	12642	Mean: 8477 SD: 3438	Middle-aged
4	0.28		9784		
5 [†]	0.12		4898		
6	0.16		6582		
7* [†]	0.27	Mean: 0.42 SD: 0.15	13602	Mean: 17267 SD: 4241	Old
8	0.56		21913		
9	0.43		16287		
10	0.001	Mean: 0.002 SD: 0.001	114	Mean: 143 SD: 41	Control
11 [†]	0.003		172		

load between 2 and 4 months of age, using a different mouse model. However, when calculated for the whole brain of 5xFAD mice, plaque load hardly reached 0.5 percent, with differences of 0.1 – 0.2 percent between groups. Therefore, our design is not appropriate to observe such small differences: there is not enough plaque load difference between the age groups we consider when observing the whole brain at once, as opposed to studying a small, specific local section of it. Moreover, comparison of sections in OPT images is hard to achieve due to the difficulty of reliably finding matching sections in the image volumes. Larger groups should thus be considered to observe a statistically significant difference in amyloid plaque load.

Secondly, only limited information about the total number of plaques can be found in the literature. Jährling *et al.* [12] observe a difference of several thousands of plaques between young (10 weeks old) and old (28-34 weeks old) animals. Although these measurements come from a different mouse model of amyloidosis and the total plaque count does not represent a whole brain, they were the most sensible known results for our work and thus served as a basis for our experimental design. Our results suggest that their plaque number estimate was a reasonable starting hypothesis, but that the variability in total plaque count was underestimated. The inter-specimen plaque number variance observed in small cubes is indeed drastically lower than in the entire brain volume. The young (11 weeks old) and middle-aged (17 weeks old) groups seem to be too close in terms of age to be able to observe a statistically significant difference. Our quantification method might not be sensitive enough to capture changes in amyloidosis under 6 weeks of age difference, as there is not enough disease progression in this period to detect a statistically significant difference in levels of amyloidosis. For that reason, we recommend having at least 8 weeks (2 months) of age difference between groups.

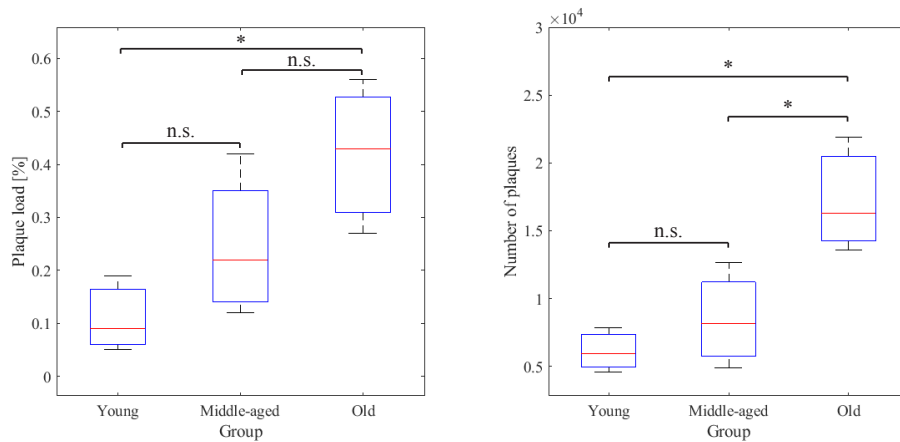


Fig. 7. Representation of the amyloidosis statistical analysis Boxplots of the quantitative measures of amyloid progression for each group of mice. *, $p < .05$ (Tuckey's test); n.s., not statistically significant.

Thirdly, when using 5xFAD mice on a B6SJL F1 background, the extreme aggressivity of the SJL/J males in the breeding cages must be considered in the experimental design as casualties are more likely to happen, thus reducing the number of generated specimens. Having witnessed this, we suggest using the congenic 5xFAD strain, on a C57BL/6J background instead. Animals should then be generated by back-crossing females 5xFAD (34848-JAX, MMRRC) with C57BL/6J males (000664, The Jackson Laboratory). We nevertheless believe that the results of our quantitative analysis can be used as a preliminary study. Based on our findings, in order to observe a statistically significant difference of 0.15 percent with a standard deviation of 0.15 percent in the plaque load relying on a one-tailed t-test, 16 animals are required in each age-group ($\alpha = .05$ and power = .85). Concerning the experiment duration, in our experience, 8-12 animals can be sacrificed per day, and the remaining sample preparation takes 4 days (count 1 day to embed the samples in agarose and 10 – 20 minutes per following day to change clearing solutions) plus 1 day of imaging.

In summary, it is noteworthy to take into consideration that quantifying the plaque load in whole mouse brains is inherently subject to more variability compared to other techniques, which focus on smaller regions of interest. Therefore, a more substantial amount of samples is necessary and care must be taken in the experimental design to take into account the variability factors mentioned above. Our technique, which considers the organ in its entirety, however offers a wealth of information for AD research. When looking at individual images, our approach allows to isolate amyloid plaques from the rest of the brain adequately. Although plaque load might be overestimated due to the instrument resolution, causing smaller bright plaques to appear with a size of approximately 50 μm , it is, to the best of our knowledge, the first time that the plaque load and the total number of plaques are measured in the whole brain of 5xFAD mice.

Our results can be related to those of other studies using different imaging modalities. Oakley *et al.* [4] quantify amyloid beta relying on ELISA and qualitatively assess amyloid deposition on histopathology images, which both exhibit an increase in plaque quantity with age. The quantitative results obtained with ELISA follow a similar trend as ours. From histopathology images, plaques are observed to be largely located in the subiculum. Additional structural labels could be incorporated in our approach so as to perform atlas-based segmentation and quantify the spatial distributions of the plaques to validate these observations. A full study of the distribution of plaque in the brain and its age-related evolution is a future research

direction for experiments involving larger animal cohorts. Hernández *et al.* [50] measure amyloid plaque density from two-photon microscope images relying on direct intensity thresholding and ad-hoc post-processing. Quantification results are provided for a single time point (11-months old animals) and extracted from image volumes corresponding to subparts of the cortex and hippocampus. Their results, being absolute numbers estimated on subvolumes of unknown location, are difficult to compare to ours. To perform a meaningful comparison, the area images with two-photon microscopy should be matched to the corresponding OPT image subvolume. This could be achieved in a single experiment, as animals used for two-photon *in vivo* imaging could be sacrificed, and imaged using OPT. The approach we propose therefore holds a strong potential to complement existing methods that investigate amyloid plaque deposition from a quantitative perspective.

5. Conclusion

In this paper, we propose an image-based analysis pipeline for the quantification of amyloidosis. The pipeline is applied to whole brain images of 5xFAD mice, a mouse model of AD. Image volumes are acquired with an optical projection tomography instrument, and amyloid plaques are segmented relying on a random forest voxel classifier. The plaque load and the total number of plaques are then consistently measured in the whole organ. The pipeline has been tested on 3D OPT images of brains of mice at different ages to illustrate the age-dependent disease progression. This preliminary study shows a statistically significant increase in the number of plaques in old animals (31 weeks old) compared to young and middle-aged ones (11 weeks old and 17 weeks old, respectively). Although the other group differences are not statistically significant, we believe the study should be repeated with more animals to draw more complete conclusions with regards to the disease progression. While there isn't a strict statistically significant difference in the plaque load ($p = \alpha$), a Tuckey post-hoc analysis revealed that the old group differs significantly from the young one. Isolation of plaques from the strong background autofluorescence signal is observed to be successful, and tests on negative controls show a negligible false positive rate. We provide the image volumes and manual annotations used for training, as well as the pre-trained ilastik workflows for download and further use by the community in [Dataset 1](#) [42].

In the future, we would like to take advantage of the tissue autofluorescence to segment regions of interest in the brain, such as the subiculum, which is the region where most of the deposition occurs in 5xFAD mice and estimate a local plaque load. Additionally, the pipeline could be reused for data acquired with other mesoscopic imaging modalities, such as ultramicroscopy, and in different rodent models of amyloidosis. Ultimately, we would like to use light sheet microscopy to establish a gold standard for amyloid plaques in the whole brain and compare it with the performance of our OPT classification method.

Data availability

OPT reconstructed data sets and ilastik pipelines are available in [Dataset 1](#) [42], along with a README document on how to use them.

Funding

EU Framework Programme for Research and Innovation (Grant agreement 686271); Swiss National Science Foundation (Grant 200020-162343); EMBL-EBI.

Acknowledgments

The authors would like to thank Dr. Arno Bouwens for the helpful discussions and suggestions for amyloid plaques segmentation, and Octave Martin for the manual segmentation of the OPT datasets.

D.N., AL.P., PJ.M., D.VDV., T.L., and A.R. were supported by the EU Framework Programme for Research and Innovation (Grant agreement 686271). V.U. was supported partly by the Swiss National Science Foundation (Grant 200020-162343), and partly by EMBL-EBI core fundings.

Disclosures

The authors declare that there are no conflicts of interest related to this article.

References

1. S. Han, M. Kollmer, D. Markx, S. Claus, P. Walther, and M. Fändrich, "Amyloid plaque structure and cell surface interactions of β -amyloid fibrils revealed by electron tomography," *Sci. reports* **7**, 43577 (2017).
2. R. H. Takahashi, T. Nagao, and G. K. Gouras, "Plaque formation and the intraneuronal accumulation of β -amyloid in alzheimer's disease," *Pathol. international* **67**, 185–193 (2017).
3. J. Hardy and D. J. Selkoe, "The amyloid hypothesis of alzheimer's disease: progress and problems on the road to therapeutics," *science* **297**, 353–356 (2002).
4. H. Oakley, S. L. Cole, S. Logan, E. Maus, P. Shao, J. Craft, A. Guillozet-Bongaarts, M. Ohno, J. Disterhoft, L. Van Eldik, R. Berry, and R. Vassar, "Intraneuronal β -amyloid aggregates, neurodegeneration, and neuron loss in transgenic mice with five familial alzheimer's disease mutations: potential factors in amyloid plaque formation," *J. Neurosci.* **26**, 10129–10140 (2006).
5. J. Götz, J. Streffer, D. David, A. Schild, F. Hoernkli, L. Pennanen, P. Kurosinski, and F. Chen, "Transgenic animal models of alzheimer's disease and related disorders: histopathology, behavior and therapy," *Mol. psychiatry* **9**, 664 (2004).
6. P. Yan, A. W. Bero, J. R. Cirrito, Q. Xiao, X. Hu, Y. Wang, E. Gonzales, D. M. Holtzman, and J.-M. Lee, "Characterizing the appearance and growth of amyloid plaques in app/ps1 mice," *J. Neurosci.* **29**, 10706–10714 (2009).
7. T. Bolmont, A. Bouwens, C. Pache, M. Dimitrov, C. Berclaz, M. Villiger, B. M. Wegenast-Braun, T. Lasser, and P. C. Fraering, "Label-free imaging of cerebral β -amyloidosis with extended-focus optical coherence microscopy," *J. Neurosci.* **32**, 14548–14556 (2012).
8. S. Hu, P. Yan, K. Maslov, J.-M. Lee, and L. V. Wang, "Intravital imaging of amyloid plaques in a transgenic mouse model using optical-resolution photoacoustic microscopy," *Opt. letters* **34**, 3899–3901 (2009).
9. B. Pinzer, M. Cacquevel, P. Modregger, S. McDonald, J. Bensadoun, T. Thuring, P. Aebischer, and M. Stampanoni, "Imaging brain amyloid deposition using grating-based differential phase contrast tomography," *Neuroimage* **61**, 1336–1346 (2012).
10. C. Duffeant, M. Vandesquille, K. Herbert, C. M. Garin, S. Alves, V. Blanchard, E. E. Comoy, F. Petit, and M. Dhenain, "Contrast-enhanced mr microscopy of amyloid plaques in five mouse models of amyloidosis and in human alzheimer's disease brains," *Sci. reports* **7**, 4955 (2017).
11. K. Becker, N. Jährling, S. Saghaei, and H.-U. Dodt, "Ultramicroscopy: light-sheet-based microscopy for imaging centimeter-sized objects with micrometer resolution," *Cold Spring Harb. protocols* **2013**, pdb-top076539 (2013).
12. N. Jährling, K. Becker, B. M. Wegenast-Braun, S. A. Grathwohl, M. Jucker, and H.-U. Dodt, "Cerebral β -amyloidosis in mice investigated by ultramicroscopy," *PloS one* **10**, e0125418 (2015).
13. T. Liebmann, N. Renier, K. Bettayeb, P. Greengard, M. Tessier-Lavigne, and M. Flajolet, "Three-dimensional study of alzheimer's disease hallmarks using the idisco clearing method," *Cell reports* **16**, 1138–1152 (2016).
14. J. Sharpe, U. Ahlgren, P. Perry, B. Hill, A. Ross, J. Hecksher-Sørensen, R. Baldock, and D. Davidson, "Optical projection tomography as a tool for 3d microscopy and gene expression studies," *Science* **296**, 541–545 (2002).
15. J. A. Gleave, M. D. Wong, J. Dazai, M. Altaf, R. Mark Henkelman, J. P. Lerch, and B. J. Nieman, "Neuroanatomical phenotyping of the mouse brain with three-dimensional autofluorescence imaging," *Physiol. genomics* **44**, 778–785 (2012).
16. D. Nguyen, P. J. Marchand, A. L. Planchette, J. Nilsson, M. Sison, J. Extermann, A. Lopez, M. Sylwestrzak, J. Sordet-Dessimoz, A. Schmidt-Christensen, D. Holmberg, D. Van De Ville, and T. Lasser, "Optical projection tomography for rapid whole mouse brain imaging," *Biomed. optics express* **8**, 5637–5650 (2017).
17. W. E. Klunk, B. J. Bacskai, C. A. Mathis, S. T. Kajdasz, M. E. McLellan, M. P. Frosch, M. L. Debnath, D. P. Holt, Y. Wang, and B. T. Hyman, "Imaging $a\beta$ plaques in living transgenic mice with multiphoton microscopy and methoxy-x04, a systemically administered congo red derivative," *J. Neuropathol. & Exp. Neurol.* **61**, 797–805 (2002).
18. A. Ebadi, J. L. Dalboni da Rocha, D. B. Nagaraju, F. Tovar-Moll, I. Bramati, G. Coutinho, R. Sitaram, and P. Rashidi, "Ensemble classification of alzheimer's disease and mild cognitive impairment based on complex graph measures from diffusion tensor images," *Front. neuroscience* **11**, 56 (2017).
19. E. Gallego-Jutglà, M. Elgendi, F. Vialatte, J. Solé-Casals, A. Cichocki, C. Latchoumane, J. Jeong, and J. Dauwels, "Diagnosis of alzheimer's disease from eeg by means of synchrony measures in optimized frequency bands," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE, (IEEE, 2012)*, pp. 4266–4270.

20. S. Beker, V. Kellner, G. Chechik, and E. A. Stern, "Learning to classify neural activity from a mouse model of alzheimer's disease amyloidosis versus controls," *Alzheimer's & Dementia: Diagn. Assess. & Dis. Monit.* **2**, 39–48 (2016).
21. K. R. Gray, P. Aljabar, R. A. Heckemann, A. Hammers, D. Rueckert, and A. D. N. Initiative, "Random forest-based similarity measures for multi-modal classification of alzheimer's disease," *NeuroImage* **65**, 167–175 (2013).
22. C. Chubb, Y. Inagaki, P. Sheu, B. Cummings, A. Wasserman, E. Head, and C. Cotman, "Biovision: an application for the automated image analysis of histological sections," *Neurobiol. aging* **27**, 1462–1476 (2006).
23. F. Letronne, G. Laumet, A.-M. Ayril, J. Chapuis, F. Demiautte, M. Laga, M. E. Vandenberghe, N. Malmanche, F. Leroux, F. Eysert, Y. Sottejeau, L. Chami, A. Flaig, C. Bauer, P. Dourlen, M. Lesaffre, C. Delay, L. Huot, J. Dumont, E. Werkmeister, F. Lafont, T. Mendes, F. Hansmannel, B. Dermaut, B. Deprez, A.-S. Hérard, M. Dhenain, N. Souedet, F. Pasquier, D. Tulasne, C. Berr, J.-J. Hauw, Y. Lemoine, P. Amouyel, D. Mann, R. Déprez, F. Checler, D. Hot, T. Delzescaux, K. Gevaert, and J.-C. Lambert, "Adam30 downregulates app-linked defects through cathepsin d activation in alzheimer's disease," *EBioMedicine* **9**, 278–292 (2016).
24. M. E. Vandenberghe, A.-S. Hérard, N. Souedet, E. Sadouni, M. D. Santin, D. Briet, D. Carré, J. Schulz, P. Hantraye, P.-E. Chabrier, T. Rooney, T. Debeir, V. Blanchard, L. Pradier, M. Dhenain, and T. Delzescaux, "High-throughput 3d whole-brain quantitative histopathology in rodents," *Sci. reports* **6**, 20958 (2016).
25. Y. Li, X. Zhu, P. Wang, J. Wang, S. Liu, F. Li, and M. Qiu, "Detection of $\alpha\beta$ plaque deposition in mr images based on pixel feature selection and class information in image level," *Biomed. engineering online* **15**, 108 (2016).
26. M. Grand'Maison, S. P. Zehntner, M.-K. Ho, F. Hébert, A. Wood, F. Carbonell, A. P. Zijdenbos, E. Hamel, and B. J. Bedell, "Early cortical thickness changes predict β -amyloid deposition in a mouse model of alzheimer's disease," *Neurobiol. disease* **54**, 59–67 (2013).
27. A. Deppeeringe, A. Foncubierta-Rodriguez, D. Van De Ville, and H. Müller, "Three-dimensional solid texture analysis in biomedical imaging: review and opportunities," *Med. image analysis* **18**, 176–196 (2014).
28. C. Sommer and D. W. Gerlich, "Machine learning in cell biology—teaching computers to recognize phenotypes," *J Cell Sci pp. jcs–123604* (2013).
29. A. Kreshuk, C. N. Straehle, C. Sommer, U. Koethe, M. Cantoni, G. Knott, and F. A. Hamprecht, "Automated detection and segmentation of synaptic contacts in nearly isotropic serial electron microscopy images," *PLoS one* **6**, e24899 (2011).
30. P. Strnad, S. Gunther, J. Reichmann, U. Krzic, B. Balazs, G. De Medeiros, N. Norlin, T. Hiiragi, L. Hufnagel, and J. Ellenberg, "Inverted light-sheet microscope for imaging mouse pre-implantation development," *Nat. methods* **13**, 139 (2016).
31. C. Sommer, C. N. Straehle, U. Koethe, and F. A. Hamprecht, "Ilastik: Interactive learning and segmentation toolkit." in *ISBI*, vol. 2 (2011), p. 8.
32. The Jackson Laboratory, "Mouse Strain Datasheet - 34840-JAX," <https://www.jax.org/strain/006554>.
33. Federal Food Safety and Veterinary Office of The Swiss Confederation, "Animal Experimentation," <https://www.blv.admin.ch/blv/en/home/tiere/tierversuche.html>.
34. W. M. S. Russell, R. L. Burch, and C. W. Hume, *The principles of humane experimental technique*, vol. 238 (Methuen London, 1959).
35. The Jackson Laboratory, "Mouse Strain Datasheet - 000686," <https://www.jax.org/strain/000686>.
36. A. C. Kak and M. Slaney, *Principles of computerized tomographic imaging* (IEEE press New York, 1988).
37. D. Dong, S. Zhu, C. Qin, V. Kumar, J. V. Stein, S. Oehler, C. Savakis, J. Tian, and J. Ripoll, "Automated recovery of the center of rotation in optical projection tomography in the presence of scattering," *IEEE journal biomedical health informatics* **17**, 198–204 (2013).
38. J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A. Cardona, "Fiji: an open-source platform for biological-image analysis," *Nat. methods* **9**, 676 (2012).
39. L. Breiman, "Random forests," *Mach. learning* **45**, 5–32 (2001).
40. M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim, "Do we need hundreds of classifiers to solve real world classification problems?" *The J. Mach. Learn. Res.* **15**, 3133–3181 (2014).
41. L. Shamir, "Assessing the efficacy of low-level image content descriptors for computer-based fluorescence microscopy image analysis," *J. microscopy* **243**, 284–292 (2011).
42. D. Nguyen, V. Uhlmann, A. L. Planchette, P. J. Marchand, D. Van De Ville, T. Lasser, and A. Radenovic, "Supervised learning to quantify amyloidosis in whole brains of an alzheimer's disease mouse model acquired with optical projection tomography," *The Open Science Framework* (2019). Retrieved in May 2019 from <https://osf.io/tvnyq/>.
43. C. A. Raji, J. T. Becker, N. D. Tsopelas, J. C. Price, C. A. Mathis, J. A. Saxton, B. J. Lopresti, J. A. Hoge, S. K. Ziolko, S. T. DeKosky, and W. E. Klunk, "Characterizing regional correlation, laterality and symmetry of amyloid deposition in mild cognitive impairment and alzheimer's disease with pittsburgh compound b," *J. neuroscience methods* **172**, 277–282 (2008).
44. G. Iordanescu, P. N. Venkatasubramanian, and A. M. Wyrwicz, "Automatic segmentation of amyloid plaques in mr images using unsupervised support vector machines," *Magn. resonance medicine* **67**, 1794–1802 (2012).
45. O. Teboul, A. Feki, A. Dubois, B. Bozon, A. Faure, P. Hantraye, M. Dhenain, B. Delatour, and T. Delzescaux, "A standardized method to automatically segment amyloid plaques in congo red stained sections from alzheimer transgenic mice," in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology*

- Society*, (IEEE, 2007), pp. 5593–5596.
46. L. Kuan, Y. Li, C. Lau, D. Feng, A. Bernard, S. M. Sunkin, H. Zeng, C. Dang, M. Hawrylycz, and L. Ng, “Neuroinformatics of the allen mouse brain connectivity atlas,” *Methods* **73**, 4–17 (2015).
 47. J. D. Whitesell, A. R. Buckley, J. E. Knox, L. Kuan, N. Graddis, A. Pelos, A. Mukora, W. Wakeman, P. Bohn, A. Ho, K. E. Hirokawa, and J. A. Harris, “Whole brain imaging reveals distinct spatial patterns of amyloid beta deposition in three mouse models of alzheimer’s disease,” *J. Comp. Neurol.* (2018).
 48. O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical image computing and computer-assisted intervention*, (Springer, 2015), pp. 234–241.
 49. T. Falk, D. Mai, R. Bensch, Ö. Çiçek, A. Abdulkadir, Y. Marrakchi, A. Böhm, J. Deubner, Z. Jäckel, K. Seiwald, A. Dovzhenko, O. Tietz, C. Dal Bosco, S. Walsh, D. Saltukoglu, T. L. Tay, M. Prinz, K. Palme, M. Simons, I. Diester, T. Brox, and O. Ronneberger, “U-net: deep learning for cell counting, detection, and morphometry,” *Nat. methods* **16**, 67 (2019).
 50. J. C. C. Hernández, O. Bracko, C. J. Kersbergen, V. Muse, M. Haft-Javaherian, M. Berg, L. Park, L. K. Vinarsik, I. Ivasyk, D. A. Rivera, Y. Kang, M. Coertes-Canteli, M. Peyrounette, V. Doyeux, A. Smith, J. Zhou, G. Otte, J. D. Beverly, E. Davenport, Y. Davit, C. P. Lin, S. Strickland, C. Iadecola, S. Lorthois, N. Nishimura, and C. B. Schaffer, “Neutrophil adhesion in brain capillaries reduces cortical blood flow and impairs memory function in alzheimer’s disease mouse models,” *Nat. neuroscience* p. 1 (2019).