



## Deep learning-based classification of retinal atrophy using fundus autofluorescence imaging

Alexandra Miere <sup>a,b,\*</sup>, Vittorio Capuano <sup>a</sup>, Arthur Kessler <sup>c</sup>, Olivia Zambrowski <sup>a</sup>, Camille Jung <sup>d</sup>, Donato Colantuono <sup>a</sup>, Carlotta Pallone <sup>a</sup>, Oudy Semoun <sup>a</sup>, Eric Petit <sup>b</sup>, Eric Souied <sup>a</sup>

<sup>a</sup> Department of Ophthalmology, Centre Hospitalier Intercommunal de Créteil, Créteil, France

<sup>b</sup> Laboratory of Images, Signals and Intelligent Systems (LISSI), (EA N° 3956), University Paris-Est, Créteil, France

<sup>c</sup> EPISEN – ISBS, University Paris-Est, Créteil, France

<sup>d</sup> Clinical Research Center, Centre Hospitalier Intercommunal de Créteil, Créteil, France



### ARTICLE INFO

#### Keywords:

Retinal imaging  
Deep learning  
Convolutional neural network  
Inherited retinal disease  
Geographic atrophy

### ABSTRACT

**Purpose:** To automatically classify retinal atrophy according to its etiology, using fundus autofluorescence (FAF) images, using a deep learning model.

**Methods:** In this study, FAF images of patients with advanced dry age-related macular degeneration (AMD), also called geographic atrophy (GA), and genetically confirmed inherited retinal diseases (IRDs) in late atrophic stages [Stargardt disease (STGD1) and Pseudo-Stargardt Pattern Dystrophy (PSPD)] were included. The FAF images were used to train a multi-layer deep convolutional neural network (CNN) to differentiate on FAF between atrophy in the context of AMD (GA) and atrophy secondary to IRDs. Three-hundred fourteen FAF images were included, of which 110 images were of GA eyes and 204 were eyes with genetically confirmed STGD1 or PSPD. In the first approach, the CNN was trained and validated with 251 FAF images. Established augmentation techniques were used and an Adam optimizer was used for training. For the subsequent testing, the built classifiers were then tested with 63 untrained FAF images. The visualization method was integrated gradient visualization. In the second approach, 10-fold cross-validation was used to determine the model's performance.

**Results:** In the first approach, the best performance of the model was obtained using 10 epochs, with an accuracy of 0.92 and an area under the curve for Receiver Operating Characteristic (AUC-ROC) of 0.981. Mean accuracy was  $87.30 \pm 2.96$ . In the second approach, a mean accuracy of  $0.79 \pm 0.06$  was obtained.

**Conclusion:** This study describes the use of a deep learning-based algorithm to automatically classify atrophy on FAF imaging according to its etiology. Accurate differential diagnosis between GA and late-onset IRDs masquerading as GA on FAF can be performed with good accuracy and AUC-ROC values.

### 1. Introduction

Atrophy represents a common late-stage manifestation of various retinal diseases, from advanced dry age-related macular degeneration (dry AMD, also called geographic atrophy, GA) to inherited retinal diseases (IRDs), such as Stargardt disease (STGD1), Best disease (BD) or Pseudo-Stargardt multifocal pattern dystrophy (PSPD) [1]. Despite being a leading cause of blindness in developed countries [2], the pathogenic mechanism for GA is still unelucidated. However, multiple factors have been demonstrated to participate in atrophy secondary to dry AMD, such as chronic inflammatory processes, excessive lipofuscin

accumulation in the RPE cells, dysregulation of the complement system, as well as vascular factors [2,3].

Conversely, STGD1 is a monogenic inherited disease. STGD1 is an autosomal recessive retinal dystrophy caused by the ABCA4 gene mutation. It has been extensively shown that ABCA4 dysfunction leads to the accumulation of lipofuscin and its constituent, A2E, resulting in atrophy development [1,4]. The phenotype of STGD1 overlaps, in late-onset stages, to that of PSPD, which is related to mutations of the PRPH2 gene. The term coined for this similar phenotype is Pseudo-Stargardt pattern dystrophy (PSPD), characterized by a slower progression to atrophy and higher age of onset [5]. Although classically

\* Corresponding author. Department of Ophthalmology, University Paris Est Créteil Centre Hospitalier Intercommunal de Créteil, 40, avenue de Verdun, 94010, Créteil, France.

E-mail addresses: [alexandra.miere@chicreteil.fr](mailto:alexandra.miere@chicreteil.fr), [alexandra.miere@chicreteil.fr](mailto:alexandra.miere@chicreteil.fr) (A. Miere).

patients with IRDs have much lower ages of onset than GA patients, in late forms of both STGD1 and PSPD, given the funduscopic similarities, these diseases may be confounded. Given the different prognosis, the possibility of enrolment in multicentric clinical trials providing experimental treatment for STGD1 and GA, and the need for genetic counseling if a genetic etiology is suspected, it is essential to correctly determine the etiology of the atrophy [1].

Furthermore, when hallmarks such as drusen or flecks are absent, the differential diagnosis between the two diseases is clinically challenging.

Fundus autofluorescence imaging (FAF) provides an *in vivo* metabolic mapping of the retina [6]. On FAF, areas of lipofuscin deposition are readily delineated. Moreover, FAF qualitative and quantitative analysis provided valuable outcome measures in several clinical studies [3,6].

In the last years, deep learning (DL) models have shown their effectiveness in image-centric specialties, such as Ophthalmology. The application of DL models in Ophthalmology focused immensely on color fundus photography (CFP), to detect sight-threatening diseases such as diabetic retinopathy (DR) or dry AMD [7]. Of recent, different DL models have appeared using optical coherence tomography (OCT) images or even OCT Angiography (OCTA) images, concentrating on DR or dry AMD [8,9]. Treder and collaborators, focused on FAF imaging to distinguish GA from a single class of other retinal diseases, from AMD without GA, adult-onset foveamacular vitelliform dystrophy, central serous chorioretinopathy, and epiretinal membranes [10]. Therefore, patients in the ‘other retinal disease’ class had no atrophy.

Several IRDs phenocopy GA and accurate diagnosis may be difficult based on fundus appearance alone [11]. Deep learning applied to FAF imaging may help the clinician’s task, ensure a correct referral to a specialized ocular genetics unit if a genetic cause is suspected, and ultimately assist the clinician when genetic testing is non-available. Our objective was to consider whether a pre-trained deep learning classifier can assist in distinguishing GA from atrophy secondary to STGD1 and PSPD.

## 2. Methods

### 2.1. Database

#### 2.1.1. Patients and genetic testing

Patients retrospectively included in this study had complete outer retinal and RPE atrophy secondary to either (1) genetically confirmed STGD1 at the atrophic stage, OR (2) genetically confirmed PSPD with atrophy OR (3) GA in the context of advanced AMD. This retrospective study was conducted in accordance with the tenets of the Declaration of Helsinki. Written consent was waived because of the retrospective nature of the study.

#### 2.1.2. Imaging

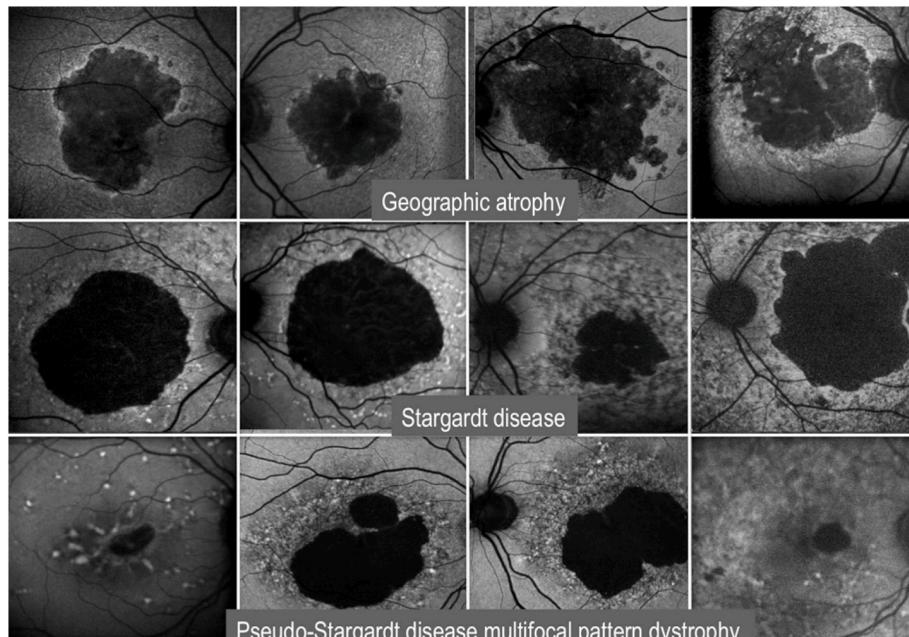
We used macula-centered fundus autofluorescence retinal images from genetically confirmed eyes with atrophy secondary to STGD1 and PSPD eyes, as well as FAF images from patients with GA from the Department of Ophthalmology of Crêteil, France. FAF images had been obtained in the Ophthalmology outpatient clinic in the Department of Ophthalmology in Crêteil between April 2007 and August 2020, using Spectralis HRA + OCT (Heidelberg Engineering, Heidelberg, Germany). High-resolution ( $1536 \times 1536$  pixels),  $30^\circ \times 30$ , as well as  $55^\circ \times 55^\circ$  degree-field-of-view images, centered on the fovea, with a minimum averaging of 30 frames, were extracted. One FAF image/eye/year of each patient was extracted. All images were deidentified and all personal data (e.g., patient name, birth date, and study date) were removed. FAF images were resized to  $224 \times 224$  pixels to meet the network specifications, with the fovea at the center. The images were labeled either ‘atrophy genetic’ or ‘GA’, according to the etiology of the outer retinal and RPE-choriocapillaris atrophy. Fig. 1 shows the FAF images of patients with GA versus eyes with atrophy secondary to IRDs.

#### 2.1.3. Data split

A two-class classification system [‘atrophy genetic’ (containing atrophic stages of STGD1 and PSPD) and ‘GA’] was implemented. For classification, we used two approaches:

##### 2.1.3.1. Approach 1.

A ‘hold out’ validation technique, by dividing the dataset into three mutually exclusive subsets, using the function



**Fig. 1.** Fundus autofluorescence in eyes with geographic atrophy secondary to dry Age-related macular degeneration (upper row) versus atrophy secondary to inherited retinal disease (middle row: STGD1, right lower row: PSPD).

train\_test\_split of scikit-learn: the training set (70% of the images), the validation set (10% of the images), and the test set (20% of the images). The split is summarized in Table 1.

Here it is important to note that all images of a patient belonged to only one dataset (training/validation/test). There was a strict separation of the training data from the validation and test data to prevent intra-eye and inter-eye correlations. Images of patients that were used for training the deep learning classifier were not used to test it.

**2.1.3.2. Approach 2.** K-fold cross-validation technique, in which the dataset is randomly split into k mutually exclusive subsets equal in size. Train test split of scikit-learn was used to split the dataset into a training set of 70% and a test set of 30%. Ten-fold cross-validation was then implemented by separating the test set into 10 subsets of equal size. Then the classifier is trained and tested  $k = 10$  times. The final accuracy is the average of the 10 iterations. This method limits overfitting. Moreover, by using different splits and multiple runs of the algorithm, the robustness and repeatability of the classification is confirmed [12, 13].

## 2.2. Development of a deep learning classifier

Input images underwent a pre-processing step, as illustrated in Fig. 3.

We used a pre-trained ResNet101 to perform the classification task. The pre-trained weights were from the ImageNet (<http://www.image-net.org/>) dataset. Global average pooling (GAP) was used to minimize overfitting, by reducing the total number of parameters in the model. The classification part of the CNN was replaced by a head, consisting of only one neuron [which will be referred to as Dense (1)], having a sigmoid function as its activation function. The output of this function is a probability vector between 0 and 1. If the probability is  $> 0.5$ , the image is predicted as GA. If the probability is  $< 0.5$ , the image is predicted as 'atrophy genetic'. The loss function used was binary cross-entropy.

Data augmentation on the training set, consisting of a horizontal flip, fill mode, width shift, height shift, rotation, and zoom, was performed automatically using Keras, using *ImageDataGenerator API*.

The model was optimized using the "Adam Optimization Algorithm", with a learning rate of 0.001 and an exponential decay rate for the first (beta1) and second (beta2) moment estimates of 0.9 and 0.999, respectively [14]. In *approach 1*, the model was then evaluated with the test set of 63 images (41 FAF images for 'atrophy genetic' and 22 FAF images for 'GA'). Because of the class imbalance between 'atrophy genetic' and 'GA', class weights were implemented in Keras using the *class\_weight* from the *sklearn* module. The class weights implemented for each class were: 0.77 for 'atrophy genetic' and 1.43 for 'GA'. To obtain convergence and to avoid overfitting, different numbers of epochs were tested (100, 20, and 10). In *approach 2*, we used 10-fold cross-validation, to allow the use of all subsets exactly once as a test set.

The method for *approach 1* is summarized in Fig. 2.

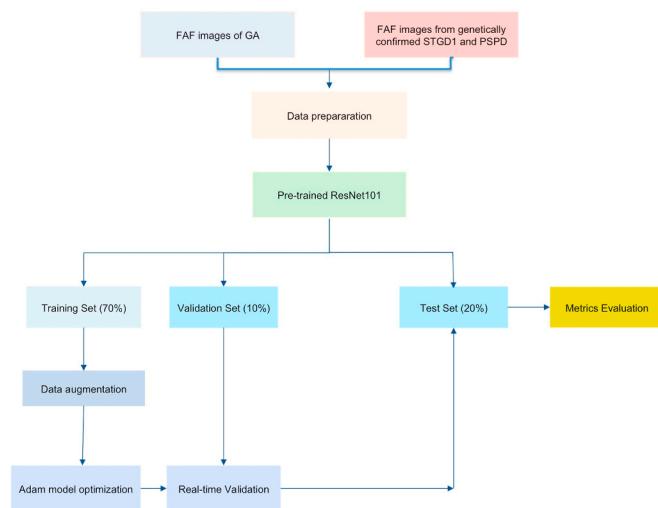
The performance was evaluated through a comparison of the CNNs output to the ground truth. Receiver Operating Characteristics (ROC) curves were generated by plotting the true-positive rate and the false-positive rate and adjusting the threshold on the output neuron from 0.5.

**Table 1**

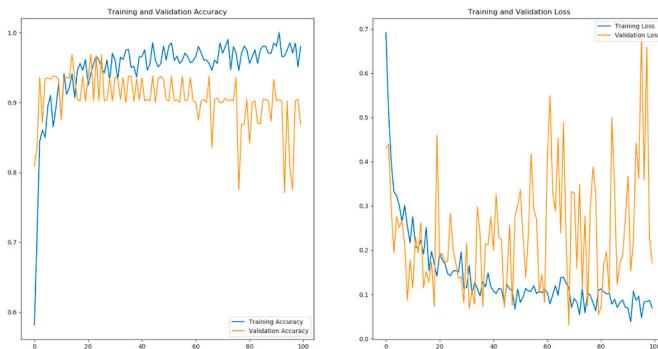
The split of the dataset for each class consisted of a training set (70%), a validation set (10%), and a test set (20%).

(n) number of fundus autofluorescence (FAF) images in each dataset.

	Training set	Validation set	Test set	Total
Geographic atrophy (GA) (n)	77 (70%)	11 (10%)	22 (20%)	110
Atrophy genetic (n)	143 (70.1%)	20 (9.8%)	41 (20.1%)	204
Total (n)	220 (70.06%)	31 (9.87%)	63 (20.06%)	314



**Fig. 2.** Approach 1. Illustration of the development of a deep learning model. Various fundus autofluorescence images of patients with geographic atrophy (GA) in the context of AMD or atrophy secondary to Stargardt disease (STGD1) or pseudo-Stargardt pattern dystrophy (PSPD) were extracted from the Créteil database and were as input in a pretrained ResNet101. The pre-trained weights were from the ImageNet (<http://www.image-net.org/>) dataset. The images were partitioned into three sets: the training set (70% of the images), the validation set (10% of the images), and the test set (20% of the images). Data augmentation was performed on the training set to increase the original dataset and to reduce overfitting of the final model. The model was optimized using Adam Optimization Algorithm. The model was then evaluated with a test set of 63 images. The output of the model was the metric evaluation of the performance of the model (accuracy, sensitivity, specificity, precision, recall, F1-score).



**Fig. 3.** One hundred epochs. Learning curves: training and validation accuracy (right), training, and validation loss (left). Note that an accuracy value of 0.93 is attained after approximately 20 epochs, followed by a plateau. The model began overfitting after 20 epochs.

## 3. Results

We included a total of 314 FAF images, as shown in Table 1: 110 images of eyes with GA (noted here as class 1) and 204 eyes with genetically confirmed STGD1 or PSPD (noted here as class 0).

### 3.1. Approach 1

At first, the model was trained during 100 epochs, to find inflection points within the learning curves and to adjust further the number of epochs.

On the test set, the model training with 100 epochs yields an accuracy is 0.89. Here, FAF images from the 'atrophy genetic', corresponding to class 0, were all correctly predicted, with a precision of 0.85 and a

recall of 1. Conversely, a correct prediction for FAF images in ‘GA’ (class 1) occurred in 15 out of 22 images, leading to a precision of 1, but a recall of 0.68.

The training with 100 epochs was repeated three more times, as can be seen in [Table 2](#).

The number of epochs was then reduced to 10 epochs and repeated four times. The model accuracy was however better with 10 epochs compared to 100 epochs, showing better generalization ability ([Fig. 4](#)).

The test accuracy for the test set was 0.92 in the first training using 10 epochs, with a correct prediction for ‘atrophy genetic’ (class 0) in 39 out of 41 instances, and a precision of 0.93/a recall of 0.95. Conversely, a correct prediction for FAF images in ‘GA’ (class 1) occurred in 19 out of 22 cases, leading to a precision of 0.90, but a recall of 0.86. The model was trained with 100, 20, and 10 epochs 4 times each, as can be seen in [Table 2](#) below.

Integrated gradient visualization was used to assess the pixels the model relies on when making a prediction [15]. [Fig. 5](#) below shows examples of correct, while [Fig. 6](#) shows incorrect attributions. The ROC curves in Approach 1 and Approach 2 are shown in [Fig. 7](#).

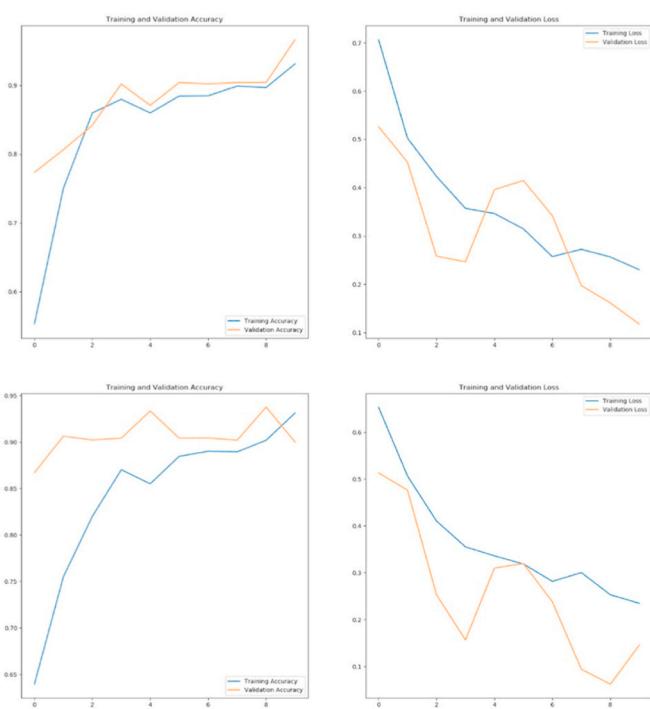
### 3.2. Approach 2

The model was trained during 100 epochs to allow the use of all subsets exactly once as a test set. Ten-fold cross-validation was implemented to assess the robustness of the model, by separating the test set into 10 subsets. The classifier was trained and tested 10 times, with a mean accuracy of  $0.79 \pm 0.06$  (confidence interval: 0.75–0.83). The

**Table 2**

Results for every training on the test set, with different numbers of epochs. GAP stands for Global Average Pooling, SD: standard deviation, CI: confidence interval.

Model features	Epochs	Accuracy (%)	AUC-ROC
Baseline (GAP + Dense(1)) + sigmoid	100	0.889	0.981
Baseline (GAP + Dense(1)) + sigmoid	100	0.841	0.983
Baseline (GAP + Dense(1)) + sigmoid	100	0.889	0.990
Baseline (GAP + Dense(1)) + sigmoid	100	0.873	0.982
Baseline (GAP + Dense(1)) + sigmoid	20	0.841	0.980
Baseline (GAP + Dense(1)) + sigmoid	20	0.857	0.978
Baseline (GAP + Dense(1)) + sigmoid	20	0.841	0.959
Baseline (GAP + Dense(1)) + sigmoid	20	0.841	0.980
Baseline (GAP + Dense(1)) + sigmoid	10	0.921	0.981
Baseline (GAP + Dense(1)) + sigmoid	10	0.905	0.979
Baseline (GAP + Dense(1)) + sigmoid	10	0.921	0.969
Baseline (GAP + Dense(1)) + sigmoid	10	0.857	0.963
	<b><math>87.3015 \pm 2.969</math></b>	<b><math>0.97708 \pm 0.0085</math></b>	<b>Mean <math>\pm</math> SD</b>
	<b>85.6213–88.9816</b>	<b>0.9722–0.9818</b>	<b>95% CI</b>



**Fig. 4.** Ten epochs. Learning curves: training and validation accuracy (right), training, and validation loss (left). Note that the model begins to overfit around 10 epochs in the first training (upper panels). The validation accuracy in the first training with 10 epochs was 0.903. In the second training, the model slightly underfits.

results of the 10 iterations are presented in [Table 3](#) below.

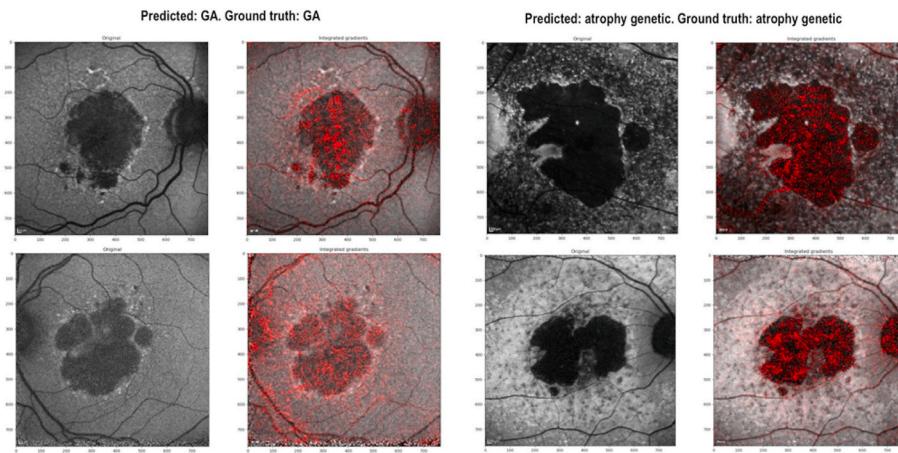
## 4. Discussion

This study reveals that atrophy secondary to dry AMD also called GA and its masqueraders (in this case, atrophy of genetic causes in STGD1 and PSPD, [Fig. 1](#)) can be distinguished with good accuracy by a deep learning model.

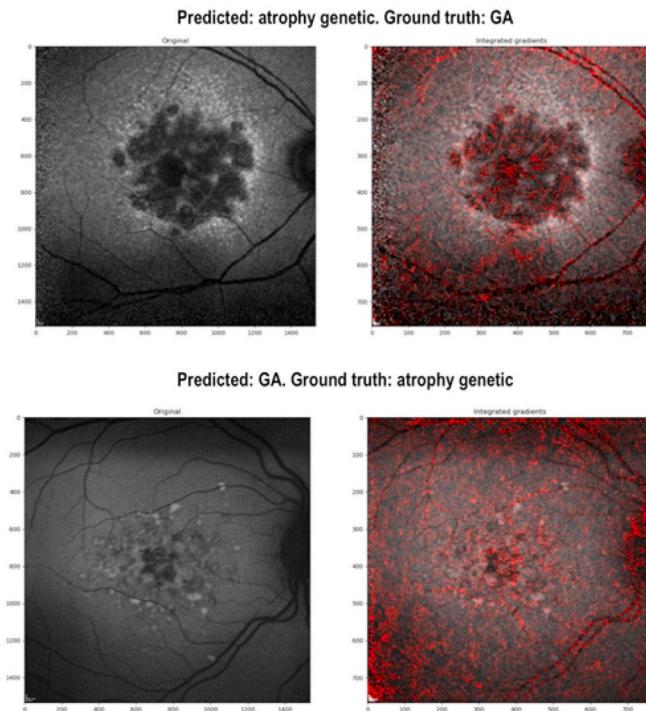
In our study, we used a pre-trained ResNet101 to perform the classification task, with pre-trained weights from the ImageNet dataset ([Fig. 2](#)). In the first approach, both GAP layers and variations of the number of epochs were used to avoid overfitting of the model. However, when testing the model with 100 epochs, obvious overfitting occurred before the model reached epoch 20 ([Fig. 3](#)). The best performance of the model was obtained using 10 epochs for Approach 1, with an accuracy of 0.921 and an AUC-ROC of 0.981 ([Table 2](#)). In Approach 2, we used 10-fold cross-validation to limit overfitting and to assess the robustness and repeatability of the classifier, obtaining a mean accuracy of  $0.792 \pm 0.064$  and an AUC-ROC of  $0.9371 \pm 0.016$  ([Table 3](#)). Our results show that, by using solely FAF images, with both a ‘hold out’ validation technique (training/validation/test) and k-fold cross-validation technique, a pretrained deep learning classifier can accurately distinguish between the 2 etiologies of atrophy with good accuracy and high AUC-ROC values.

Treder et al. recently used a deep learning classifier to differentiate between GA and a class entitled ‘other retinal diseases’, which consisted of AMD without GA, adult-onset foveomacular vitelliform dystrophy, central serous chorioretinopathy, and epiretinal membranes. Their model achieved a training accuracy of 98% and a validation accuracy of 91%, with reported sensitivity, specificity, and accuracy on the test dataset of 100% [10]. However, the authors used 500 epochs, with both training and validation accuracy reaching a plateau early on.

FAF is classically used in both clinical practice and clinical trials to assess GA. In the “Beacon” study ([ClinicalTrials.gov](#) Identifier:



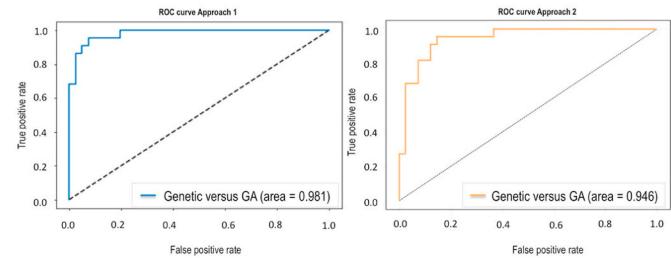
**Fig. 5.** Examples of correct attributions with integrated gradient visualization. Left panels: Geographic atrophy (GA) fundus autofluorescence images (FAF) correctly classified as ‘GA’. Right panels: Atrophy of ‘genetic cause’ (Stargardt disease) correctly classified as such.



**Fig. 6.** Examples of incorrect attributions with integrated gradient visualization. Upper panel: Geographic atrophy (GA) fundus autofluorescence image (FAF) incorrectly classified as ‘atrophy genetic’. Lower panel: Atrophy of genetic cause incorrectly classified as ‘GA’.

NCT02087085), and in the “Chroma” study ([ClinicalTrials.gov](#) Identifier: NCT02247479) the primary outcome measures were the changes from baseline using FAF. Image processing techniques, such as level set and other region-based techniques have been integrated into instruments to segment GA on FAF images. One such example is the commercial software Region finder® (Heidelberg Engineering), based on region growing techniques, which was validated for GA progression assessment [16]. Then, machine learning and, more recently, deep learning methods, were also used to detect and classify GA in particular in recent years in FAF or OCT imaging [17–22].

In a recent paper part of a technical conference by Hu et al., three deep learning CNNs were used to segment GA in 50 randomly chosen FAF images of 50 GA patients [21]. The automated segmentation results were then compared with the manual delineation performed by a



**Fig. 7.** ROC curves in the two approaches, generated by plotting the true-positive rate and the false-positive rate, and adjusting the threshold on the output neuron from 0.5. Left panel: ROC curve for the first training with 10 epochs in approach 1. In this case, the AUC-ROC is 0.981, showing the capacity of the model in distinguishing atrophy of genetic causes (class 0) from geographic atrophy (class 1). Right panel: ROC curve for the fourth training in approach 2 with 10 folds cross-validation. In this case, the AUC-ROC is 0.946, showing the capacity of the model in distinguishing atrophy of ‘genetic causes’ (class 0) from ‘geographic atrophy’ GA (class 1).

**Table 3**

Results for the test set using k-fold cross-validation (10 folds). GAP stands for Global Average Pooling, SD: standard deviation, CI: confidence interval.

Model features	Epochs	Accuracy (%)	AUC-ROC
Baseline (GAP + Dense (1)) + sigmoid	100	0.809	0.958
Baseline (GAP + Dense (1)) + sigmoid	100	0.825	0.939
Baseline (GAP + Dense (1)) + sigmoid	100	0.873	0.937
Baseline (GAP + Dense (1)) + sigmoid	100	0.873	0.946
Baseline (GAP + Dense (1)) + sigmoid	100	0.841	0.949
Baseline (GAP + Dense (1)) + sigmoid	100	0.730	0.902
Baseline (GAP + Dense (1)) + sigmoid	100	0.825	0.938
Baseline (GAP + Dense (1)) + sigmoid	100	0.730	0.929
Baseline (GAP + Dense (1)) + sigmoid	100	0.682	0.955
Baseline (GAP + Dense (1)) + sigmoid	100	0.730	0.919
		<b>0.792 ± 0.064</b>	<b>0.9371 ± 0.016</b>
		<b>0.752–0.832</b>	<b>0.927–0.947</b>
			<b>Mean ± SD</b>
			<b>95% CI</b>

certified grader. The authors applied two-fold cross-validation with a mean GA segmentation accuracy of  $0.97 \pm 0.02$  and an overlap ratio between the automatic and manual segmentations of  $0.79 \pm 0.12$ . Another recent technical conference paper from the same group aimed to detect and classify atrophy in GA versus STGD1 on FAF images using residual networks for the classification task and U-Net for the segmentation task [22]. The dataset consisted here of 320 FAF images from healthy subjects, 320 FAF images of atrophic AMD (GA), and 100 with atrophic STGD1. The screening accuracy was 0.98 for GA and 0.95 for atrophic STGD1. There are several differences to the present work: while Wang et al. [22] used 320 FAF GA images and 100 atrophic STGD1 FAF images, in the present work we used 110 GA FAF images and 204 FAF images of atrophic stages of IRDs, of note not only of STGD1 (ABCA4 mutation) but also PSPD (PRPH2 mutation). Both IRDs included in this study had a confirmed genetic molecular diagnosis. Hence, while the previous chapter focused on distinguishing between juvenile STGD1 atrophy from GA, we expanded the focus on the later-onset PSPD and late-onset STGD1, which from a clinical perspective are harder to differentiate. Another particularity of our study is the use of integrated gradient visualization, which allows us to ascertain the clinical comprehensiveness of the model's prediction and should help the clinician avoid the automation bias.

Besides FAF, multimodal imaging techniques, and in particular invasive imaging techniques, such as FA and ICGA, allow a clearer distinction between GA and IRDs. In the particular case of STGD1, 'dark atrophy' is prevalent (92%) in late frames of ICGA, compared to GA, where dark atrophy is present in only 13% of cases [23]. The advent of high resolution, non-invasive imaging techniques, from FAF to OCT and OCTA, has led to a paradigm shift in terms of diagnostic imaging. To date, the American Academy of Ophthalmology recommendations for imaging in patients with IRDs consists of non-invasive imaging techniques [24]. Therefore, in late-onset forms of STGD1, as well as in PSPD, a deep learning classifier can distinguish between the genetic or degenerative cause of the atrophy, without requiring any invasive examinations.

Other IRDs, besides STGD1 and PSPD, may masquerade as GA, such as cone dystrophy, adult vitelliform dystrophy, North Carolina macular dystrophy, Doyne honeycomb dystrophy, Sorsby macular dystrophy, X-linked retinoschisis, as well as maternally inherited diabetes and deafness (MIDD). Further studies should focus on a broader spectrum of IRDs for the differential diagnosis with GA.

Last but not least, disease progression is distinct between late-onset forms of STGD1 or PSPD and dry AMD, concerning the yearly rate of atrophy growth, atrophy kinetics, foveal involvement, and the consequent loss of visual acuity [1]. Therefore, refined phenotyping in these cases is essential, to correctly assess the disease course in patients presenting with RPE atrophy. A review of the patient's family history, as well as diagnostic studies such as ERG and genetic testing, in conjunction with FAF, can help make a clear diagnosis.

In conclusion, an accurate differential diagnosis between GA and late-onset IRDs masquerading as GA on FAF can be performed using a deep learning model, with excellent accuracy and excellent AUC-ROC values.

## Funding/support

No funding was received.

## Financial disclosures

Alexandra Miere is a consultant for: Optovue (Freemont, California, USA), Allergan Inc (Irvine, California, USA), Novartis (Basel, Switzerland) and Bayer Shering-Pharma (Berlin, Germany). Vittorio Capuano is consultant for ZEISS (Dublin, USA). Arthur Kessler: none; Olivia Zambrowski: none; Camille Jung: none; Donato Colantuono: none; Carlotta Pallone: none; Oudy Semoun is a consultant for Allergan

Inc (Irvine, California, USA), Bayer Shering-Pharma (Berlin, Germany), Novartis (Basel, Switzerland). Eric Petit: none; Eric H. Souied is a board member and consultant for: Allergan Inc (Irvine, California, USA), Bausch + Lomb (Rochester, New York, USA), Bayer Shering-Pharma (Berlin, Germany), Novartis (Basel, Switzerland).

## Other disclosures

None.

All authors attest that they meet the current ICMJE requirements to qualify as authors.

## Declaration of competing interest

The authors have no competing interests to declare.

## References

- [1] M. Lindner, S. Lambertus, M.M. Mauschitz, et al., Differential disease progression in atrophic age-related macular degeneration and late-onset Stargardt disease, *Invest. Ophthalmol. Vis. Sci.* 58 (2) (2017) 1001–1007, <https://doi.org/10.1167/iovs.16-20980>.
- [2] L.S. Lim, P. Mitchell, J.M. Seddon, F.G. Holz, T.Y. Wong, Age-related macular degeneration, *Lancet Lond Engl* 379 (9827) (2012) 1728–1738, [https://doi.org/10.1016/S0140-6736\(12\)60282-7](https://doi.org/10.1016/S0140-6736(12)60282-7).
- [3] F.G. Holz, E.C. Strauss, S. Schmitz-Valckenberg, M. van Lookeren Campagne, Geographic atrophy: clinical features and potential therapeutic approaches, *Ophthalmology* 121 (5) (2014) 1079–1091, <https://doi.org/10.1016/j.ophtha.2013.11.023>.
- [4] J.R. Sparrow, T. Duncker, R. Woods, F.C. Delori, Quantitative fundus autofluorescence in best vitelliform macular dystrophy: RPE lipofuscin is not increased in non-lesion areas of retina, in: C. Bowes Rickman, M.M. LaVail, R. E. Anderson, C. Grimm, J. Hollyfield, J. Ash (Eds.), *Retinal Degenerative Diseases*, vol. 854, Springer International Publishing, 2016, pp. 285–290, [https://doi.org/10.1007/978-3-319-17121-0\\_38](https://doi.org/10.1007/978-3-319-17121-0_38).
- [5] C.J.F. Boon, Retinal dystrophies associated with the PRPH2 gene, in: B. Puech, J.-J. De Laey, G.E. Holder (Eds.), *Inherited Chorioretinal Dystrophies*, Springer Berlin Heidelberg, 2014, pp. 213–233, [https://doi.org/10.1007/978-3-540-69466-3\\_19](https://doi.org/10.1007/978-3-540-69466-3_19).
- [6] S. Schmitz-Valckenberg, M. Pfau, M. Fleckenstein, et al., Fundus autofluorescence imaging, *Prog. Retin. Eye Res.* (August 3, 2020) 100893, <https://doi.org/10.1016/j.preteyeres.2020.100893>. Published online.
- [7] M.D. Abramoff, Y. Lou, A. Erginay, et al., Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning, *Invest. Ophthalmol. Vis. Sci.* 57 (13) (2016) 5200–5206, <https://doi.org/10.1167/iovs.16-19964>.
- [8] M. Treder, J.L. Lauermann, N. Eter, Automated detection of exudative age-related macular degeneration in spectral domain optical coherence tomography using deep learning, *Graefes Arch Clin Exp Ophthalmol Albrecht Von Graefes Arch Klin Exp Ophthalmol* 256 (2) (2018) 259–265, <https://doi.org/10.1007/s00417-017-3850-3>.
- [9] D. Le, M. Alam, C. Yao, et al., Transfer learning for automated OCTA detection of diabetic retinopathy, *ArXiv191001796 Eess Q-Bio*. Published online October 4, 2019, <http://arxiv.org/abs/1910.01796>. (Accessed 8 September 2020).
- [10] M. Treder, J.L. Lauermann, N. Eter, Deep learning-based detection and classification of geographic atrophy using a deep convolutional neural network classifier, *Graefes Arch. Clin. Exp. Ophthalmol.* 256 (11) (2018) 2053–2060, <https://doi.org/10.1007/s00417-018-4098-2>.
- [11] Masqueraders of Age-Related Macular Degeneration. *Retina Today*. <http://fyra.io>. (Accessed 8 September 2020). <https://retinatoday.com/articles/2011-ap-masqueraders-of-age-related-macular-degeneration>.
- [12] Vanwinckelen G, Blockeel H. On estimating model accuracy with repeated cross-validation. In: BeneLearn 2012: Proceedings of the 21st Belgian-Dutch Conference on Machine Learning. 20120101:39–44. Accessed November 2, 2020. <https://lirias.kuleuven.be/retrieve/186558>.
- [13] Kim J-H. Estimating classification error rate: repeated cross-validation, repeated hold-out and bootstrap." *Comput. Stat. Data Anal.*, 53(11), 3735-3745. *Comput Stat Data Anal.* 2009;53:3735-3745. doi:10.1016/j.csda.2009.04.009.
- [14] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, *ArXiv14126980 Cs*. Published online January 29, 2017, <http://arxiv.org/abs/1412.6980>. (Accessed 13 September 2020).
- [15] Sundararajan M, Taly A, Yan Q. Axiomatic Attribution for Deep Networks. *ArXiv170301365 Cs*. Published online June 12, 2017. Accessed September 4, 2020. <http://arxiv.org/abs/1703.01365>.
- [16] S. Schmitz-Valckenberg, C.K. Brinkmann, F. Alten, P. Herrmann, N.K. Stratmann, A.P. Göbel, M. Fleckenstein, M. Diller, G.J. Jaffe, F.G. Holz, Semiautomated image processing method for identification and quantification of geographic atrophy in age-related macular degeneration, *Invest. Ophthalmol. Vis. Sci.* 52 (10) (2011 Sep 29) 7640–7646, <https://doi.org/10.1167/iovs.11-7457>.
- [17] Z. Hu, S. Sadda, Chapter 17 - Image Analysis Tools for Assessment of Atrophic Macular Diseases. *Computational Retinal Image Analysis*, Academic Press, 2019,

- ISBN 9780081028162, pp. 353–378, <https://doi.org/10.1016/B978-0-08-102816-2.00018-6>.
- [18] G. Ometto, G. Montesano, S. Sadeghi Afgeh, G. Lazaridis, X. Liu, P.A. Keane, D. P. Crabb, A.K. Denniston, Merging information from infrared and autofluorescence fundus images for monitoring of chorioretinal atrophic lesions, *Transl Vis Sci Technol* 9 (9) (2020 Aug 25) 38, <https://doi.org/10.1167/tvst.9.9.38>.
  - [19] Z. Hu, G.G. Medioni, M. Hernandez, S.R. Sadda, Automated segmentation of geographic atrophy in fundus autofluorescence images using supervised pixel classification, *J. Med. Imaging* 2 (2015), 014501.
  - [20] Z. Hu, G.G. Medioni, M. Hernandez, A. Hariri, X. Wu, S.R. Sadda, Segmentation of the geographic atrophy in spectral-domain optical coherence tomography and fundus autofluorescence images, *Invest. Ophthalmol. Vis. Sci.* 54 (2013) 8375–8383.
  - [21] Automated segmentation of geographic atrophy using deep convolutional neural networks, in: Z. Hu, Z. Wang, S.R. Sadda (Eds.), *Medical Imaging 2018: Computer-Aided Diagnosis*, International Society for Optics and Photonics, 2018.
  - [22] Deep learning for automated screening and semantic segmentation of age-related and juvenile atrophic macular degeneration, in: Z. Wang, S.R. Sadda, Z. Hu (Eds.), *Medical Imaging 2019: Computer-Aided Diagnosis*, International Society for Optics and Photonics, 2019.
  - [23] A. Giani, M. Pellegrini, E. Carini, A. Peroglio Deiro, F. Bottoni, G. Staurenghi, The dark atrophy with indocyanine green angiography in Stargardt disease, *Invest. Ophthalmol. Vis. Sci.* 53 (7) (2012) 3999–4004, <https://doi.org/10.1167/iovs.11-9258>.
  - [24] Recommendations on Clinical Assessment of Patients with Inherited Retinal Degenerations, American Academy of Ophthalmology, Published February 6, 2016, <https://www.aao.org/clinical-statement/recommendations-on-clinical-assessment-of-patients>, 2016. (Accessed 8 September 2020).