tvst

Predicting Thioflavin Fluorescence of Retinal Amyloid Deposits Associated With Alzheimer's Disease from Their Polarimetric Properties

Yunyi Qiu[1], Tao Jin[2], Erik Mason[3], and Melanie C. W. Campbell[4,5]

[1] Department of Physics and Astronomy, University of Waterloo, Waterloo, Ontario, Canada
[2] Department of Physics and Astronomy, University of Waterloo, Waterloo, Ontario, Canada
[3] Department of Physics and Astronomy, University of Waterloo, Waterloo, Ontario, Canada
[4] Department of Physics and Astronomy, School of Optometry and Vision Science, Department of Systems Design Engineering, Centre for Bioengineering and Biotechnology, Waterloo Institute of Nanotechnology, University of Waterloo, Waterloo, Ontario, Canada
[5] Centre for Eye and Vision Research, Hong Kong

Correspondence: Melanie C.W. Campbell, Department of Physics and Astronomy, 200 University Ave W, Waterloo, ON, N2L 3G1, Canada. e-mail: mcampbel@uwaterloo.ca

Purpose: To use machine learning in those with brain amyloid to predict thioflavin fluorescence (indicative of amyloid) of retinal deposits from their interactions with polarized light.

Methods: We imaged 933 retinal deposits in 28 subjects with post mortem evidence of brain amyloid using thioflavin fluorescence and polarization sensitive microscopy. Means and standard deviations of 14 polarimetric properties were input to machine learning algorithms. Two oversampling strategies were applied to overcome data imbalance. Three machine learning algorithms: linear discriminant analysis, supporting vector machine, and random forest (RF) were trained to predict thioflavin positive deposits. For each method; accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve were computed.

Results: For the polarimetric positive deposits, using 1 oversampling method, RF had the highest area under the receiver operating characteristic curve (0.986), which was not different from that with the second oversampling method. RF had 95% accuracy, 94% sensitivity, and 97% specificity. After including deposits with no polarimetric signals, polarimetry correctly predicted 93% of thioflavin positive deposits. Linear retardance and linear anisotropy were the dominant polarimetric properties in RF with 1 oversampling method, and no polarimetric properties were dominant in the second method.

Conclusions: Thioflavin positivity of retinal amyloid deposits can be predicted from their images in polarized light. Polarimetry is a promising dye-free method of detecting amyloid deposits in *ex vivo* retinal tissue. Further testing is required for translation to live eye imaging.

Translational Relevance: This dye-free method distinguishes retinal amyloid deposits, a promising biomarker of Alzheimer's disease, in human retinas imaged with polarimetry.

Introduction

Alzheimer's disease (AD) is a neurodegenerative disease that leads to cognitive impairment and, ultimately, death. AD is the primary cause of dementia. In 2010, the worldwide prevalence of dementia was 35.6 million and this number is predicted to triple by 2050.[1] Currently, this disease can only be definitively diagnosed post mortem through the severity of 2 hallmarks: amyloid plaques (composed of misfolded amyloid-β protein) and neurofibrillary

1

tangles (composed of tau protein), both of which begin to accumulate in the brain before symptoms of cognitive impairment.[2] Detection of amyloid, tau, and neurodegeneration in the brain is recommended for detecting changes associated with the AD disease process in clinical trials of novel treatments.[3] Confirmation of amyloid in the brain is currently achieved using positron emission tomography (PET) scans[4] or analyzing the cerebral spinal fluid.[5] However, both methods are invasive: PET scans use a radiative isotope injection,[4] and a cerebral spinal fluid test requires extraction of fluid from the patient's spine.[5] Currently, there is no effective cure for AD. However, it has been suggested that earlier detection of brain amyloid in prodromal disease could enable more successful treatment before damage from amyloid and tau accummulates.[6] Furthermore, the amyloid, tau, and neurodegeneration framework allows for the incorporation of other biomarkers of amyloid in the brain.[5] Therefore, it is important to develop a noninvasive method to detect amyloid in association with AD before AD diagnosis. Potential peripheral biomarkers of amyloid in the brain include detection of amyloid in the retina.[7]

Many changes have been measured in the retina in association with AD.[8] An accumulation of amyloid deposits has been found in the retinas of AD patients by our group (M.C.W. Campbell, et al. *IOVS* 2010;51:ARVO E-Abstract 5778) and others,[6,9] including immunohistochemistry positivity for the presence of amyloid-β[10–12] (Yuchun Tsai, et al. *IOVS* 2014;55:ARVO E-abstract 523-524). We demonstrated that it is possible to detect amyloid deposits in the retina using polarized light in an imaging device with animal models (Michael Tokiyoshi Hamel, et al. *IOVS* 2016;57:ARVO E-Abstract 2216) and humans (Tao Jin, et al. *IOVS* 2017;58:ARVO E-Abstract 3367; David DeVries, et al. *IOVS* 2015;56:ARVO E-Abstract 2385), because the interactions of the deposits with polarized light are significantly different from those of the surrounding retina. Retinal deposits were confirmed to be amyloid via Thioflavin-S staining.[13] Koronyo et al.[14] have also demonstrated the feasibility of the *in vivo* detection of retinal amyloid deposits using curcumin staining. This method, however, requires the ingestion of curcumin for 2 to 10 days to bind with the deposits.[14] Our group has also shown that the number of amyloid deposits in the anterior retina predicts the severity of AD pathology in the brain, including the severity of brain amyloid[15] (Frank Corapi, et al. *IOVS* 2018;59:ARVO E-Abstract 1582).

Thioflavin is a fluorescent marker of amyloid, including amyloid-β, known to occur in AD. Thioflavin also stains other amyloids,[16] some of

which are associated with other neurodegenerative diseases.[17] In the amyloid, tau, and neurodegeneration diagnostic framework for AD, amyloid PET scans are an approved method of measuring brain amyloid.[3] Similarly, a dye used in amyloid PET scans, Pittsburgh compound B, is an uncharged analog derived from thioflavin and thus marks amyloid and may not be specific to amyloid-β.[18]

A number of studies have applied machine learning techniques to AD diagnosis. Two studies used a supporting vector machine (SVM) algorithm to distinguish those with AD from controls using neuroimaging data from magnetic resonance imaging.[19,20] Another used random forest (RF) analysis on magnetic resonance imaging data to separate those with AD from healthy controls and compared the performance of RF with SVM.[21] In optical coherence tomography, a SVM-based 3-dimensional segmentation of retinal layers has been reported in both diseased and normal retina.[22] In addition, images taken with polarized light of morphologically similar algae have been classified with a convolutional neural network.[23]

In those with brain amyloid, imaging the retina with polarized light (polarimetry) detects the majority of the deposits stained with Thioflavin-S (true positives), some deposits that are not stained by Thioflavin-S (false positives) and fails to detect a few Thioflavin-S–positive deposits (false negatives). The purpose of this study was to differentiate Thioflavin-S-positive amyloid deposits (true positives) from those deposits without Thioflavin-S fluorescence (false positives), based on the polarimetric images of the deposits to identify retinal amyloid deposits without the use of dye. To this end, we implemented and compared machine learning approaches, which can predict the presence of retinal amyloid fluorescence using only the information from images obtained with polarized light.

Methods

Sample Preparation

Retinal Samples

Eyes and brains ($N = 28$) were obtained post mortem from donors in compliance with the Declaration of Helsinki. Informed consent was obtained from the donors and the research was approved by the Human Research Ethics Committee of the University of Waterloo. Upon post mortem examination (National Institute on Aging–Alzheimer's Institute guidelines[2]), 23 donors had a high cumulative score of AD neuropathologic change, 4 had a moderate score, and 1 had a low score; all had

evidence of brain amyloid. Eyes were immersed in 10% formalin before dissecting, staining, and flat mounting the retina. Retinas were stained with 0.1% Thioflavin-S, counterstained with DAPI, and coverslipped. Each retina was imaged with fluorescence and polarimetric microscopy to determine presumed amyloid deposits. In total, 920 polarization-positive retinal deposits found in anterior retinal layers from 28 individuals were analyzed, including 789 deposits with fluorescence signals and 131 deposits without fluorescence signals. An additional 16 deposits had only fluorescent signals; 13 of these deposits had no polarimetric signals and 3 had unclear polarimetric signals.

Pure Amyloid Samples

Amyloid-β (Aβ) 1-42 was purchased from rPeptide (Athens, GA). One milliliter of 10 mM HEPES and 150 mM NaCl (pH of 7.4) was added to 0.5 mg of Aβ (1-42) and incubated at 37°C for 72 hours without shaking. Stained Aβ was prepared by adding 500 µL of filtered 1.25 mM Thioflavin-T in 50 mM of phosphate-buffered saline and allowing it to sit for 5 minutes before gently rinsing the amyloid deposited on glass with 50 µL of distilled water (×3). Samples were then blown dry with a gentle stream of compressed N$_2$ and cover slipped.

Mueller Matrix Polarimetry

An inverted transmission Nikon microscope was modified for polarimetric imaging to detect the interaction of retinal deposits with polarized light. A polarization state generator (PSG) and a polarization state analyzer (PSA) were placed before and after the sample, respectively. Both the PSG and PSA are composed of a linear polarizer and a quarter wave plate (QWP), with the polarizer followed by a QWP in the PSG and the opposite in the PSA. The light travels from the PSG through the sample to the PSA before being collected by the camera (Fig. 1).

The QWP in the PSG is rotated to four angles (45°, 0°, −30°, and −60°) to generate four different input polarization states. The output at each setting was determined by rotating the QWP in the PSA to the same four angles. Thus, a combination of 16 measurements were performed for each sample. The images obtained were registered to compute the Mueller matrix of the sample at each image pixel. For a given pixel, the intensity, I_i recorded in the ith measurement is then,

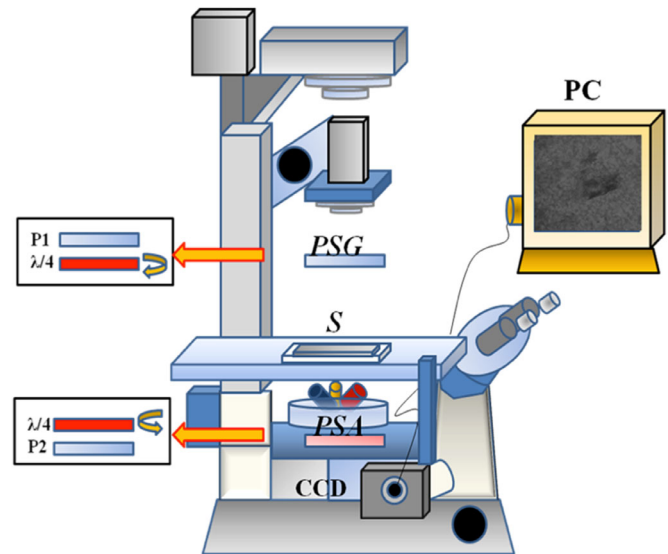$$I_i = M_{PSA,i} M_s M_{PSG,i} S_{in} \quad (i = 0, 1, 2, \ldots, 15) \quad (1)$$

Figure 1. A schematic diagram represents the setup of the microscope channel for polarimetric imaging. P1 and P2 are linear polarizers, $\frac{\lambda}{4}$ marks quarter wave plates. The CCD is a camera with a charge-coupled device. PSA, polarization state analyzer; PSG, polarization state generator.

where S_{in} stands for the Stokes vector of the incident light and $M_{PSA,i}$ and $M_{PSG,i}$ are the Mueller matrices of the PSA and PSG, respectively. Because the Mueller matrices of these elements are known, the Mueller matrix at each pixel of the sample, M_s, can be determined and polarimetric properties calculated.

Machine Learning Approach

Mean values and standard deviations of polarimetric properties (features) of the segmented deposits were used as feature inputs for machine learning algorithms which then learn to classify the deposits as members of the thioflavin positive or negative datasets labelled by their class. Thus, the algorithms are often referred to as classifiers. For definitions of machine learning terms please refer to Appendix A.

Segmentation and Calculation of Properties of Polarimetric Images

From the 16-element Mueller matrix calculated at each pixel, the state of polarized light exiting from the sample for each possible state of light input can be calculated. The way in which the input polarized states are changed to the exiting states is described by the Mueller matrix of each pixel of the sample from which polarimetric properties are calculated. For example, one state of incident polarization could be preferentially absorbed (the property of diattenuation) or the

refractive index seen by different polarizations could differ (the property of retardance) or polarized light could be changed to partially polarized light.[24]

For each deposit, the means and standard deviations of 14 polarimetric properties were extracted from the Mueller matrix values of the segmented pixels and used as feature inputs for machine learning algorithms. Linear retardance (LR), circular retardance, and depolarization power were extracted by the polar decomposition method.[25] Two polarimetric properties, linear anisotropy (LA) and circular anisotropy, were calculated from the Mueller matrix components.[26] Four polarimetric properties computed by the Mueller matrix transformation technique were: metric A (a measure of LA), metric b (sensitive to small sources of scattered light in samples), metric t (related to the magnitude of the anisotropy) and metric x (related to the direction of the aligned fibrous structures).[27,28] Linear and circular polarizance and diattenuation were directly obtained from the polarizance and diattenuation vectors of the Mueller matrix, respectively. The Q metric is a polarimetric property that contains information on depolarization, polarizance and diattenuation.[29] Table 1 in Appendix B summarizes all 14 polarimetric properties with equations. Polarimetric properties of presumed amyloid deposits (plotted as images) were segmented from the surrounding retinal background by a custom polarimetry segmentation method (Erik Mason, et al. *IOVS* 2019;60:ARVO E-Abstract 179).

Oversampling Methods for Solving Data Imbalance

As described in the Sample Preparation section, our 2 datasets do not have balanced deposit numbers (789 fluorescence positive and 131 fluorescence negative deposits). In this study, the minority (smaller) dataset contains samples with polarization signals but no fluorescence signals, and the majority (larger) dataset contains samples with both polarization and fluorescence signals. This imbalance, frequently observed in medical datasets, can lead to poor performance of most classification algorithms.[30] Two oversampling methods were used to resolve data imbalance while maintaining the information of the original dataset: (1) sampling fluorescence negative retinal background and (2) the borderline-SMOTE algorithm.[31]

Method 1: Retinal Oversampling—Adding the Surrounding Retinal Background

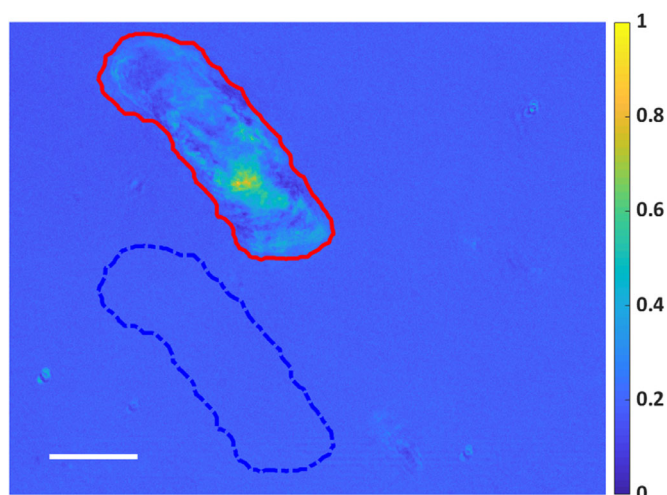To supply more samples with no fluorescence signal, regions with the same shape as a polarimetric-positive

Figure 2. An example of the retinal oversampling method. In the linear anisotropy image of a polarimetric-positive deposit, a region with the same shape (*blue dashed line*) as the deposit (*red line*) with no fluorescence signal is extracted as a fluorescence negative sample. Scale bar: 20 μm.

deposit in a nearby fluorescence and polarimetric-negative retinal region were extracted (Fig. 2). The extracted region has no fluorescence signal and a weak polarization signal. We then calculated the means and standard deviations of the polarimetric signals from these regions and labeled them as nonfluorescence deposits. Of these regions, 658 were randomly selected from the eligible retina areas. As a result, we have an equal number (789) of fluorescence-positive and fluorescence-negative samples after oversampling.

Method 2: Oversampling by Borderline-SMOTE

Another way of oversampling is to artificially generate data to add to the minority dataset. Here, an improved version of the SMOTE algorithm[32]— borderline-SMOTE[31] was applied. This algorithm first performs a *k*-nearest neighbor analysis on the thioflavin negative (minority) dataset where the nearest neighbors may be within either or both minority and thioflavin positive (majority) datasets. Based on the type of nearest neighbors, deposits in the minority dataset are ranked into 3 subsets: noise, danger, and safe. Deposits in the noise subset have all nearest neighbors in the majority dataset; deposits in the safe subset have more than one-half of their nearest neighbors in the minority dataset. These subsets are relatively far from the border between the minority and majority datasets. Deposits in the danger subset have more than one-half but not all of their nearest neighbors in the majority dataset. New synthetic data are generated by performing the SMOTE algorithm only on the danger subset.

Table 1. ACC, SEN, and SPE of the 3 Machine Learning Classifiers Under 3 Sampling Strategies

Method	Without Oversampling			Retinal Oversampling			Borderline-SMOTE Oversampling		
	ACC	SEN	SPE	ACC	SEN	SPE	ACC	SEN	SPE
LDA	90.3 ± 0.7%	96.6 ± 2.1%	47.5 ± 3.9%	90.9 ± 0.0%	91.1 ± 3.1%	93.4 ± 1.2%	80.5 ± 0.7%	76.8 ± 1.4%	85.4 ± 2.4%
SVM	85.9 ± 0.7%	100.0 ± 0.0%	0.0 ± 0.0%	92.5 ± 0.3%	95.6 ± 0.7%	91.4 ± 0.7%	93.1 ± 0.5%	91.0 ± 1.2%	95.0 ± 1.1%
RF	91.3 ± 0.8%	97.0 ± 2.2%	45.9 ± 1.5%	93.7 ± 0.2%	95.5 ± 1.3%	92.1 ± 2.1%	94.6 ± 0.8%	93.6 ± 2.1%	96.7 ± 0.77%

The means ± standard deviations are given for each of accuracy (ACC), sensitivity (SEN), and specificity (SPE) for each of linear discriminant analysis (LDA), supporting vector machine (SVM) and random forest (RF), without oversampling and with 2 different oversampling methods.

Applying Machine Learning Algorithms

Because different polarimetric properties have different scales due to their nature (e.g., the LR ranges from 0° to 180°, whereas linear diattenuation ranges from 0 to 1), a widely used method called min-max scaling, which converts all polarimetric properties to have the same scale using formula (2) was applied to standardize polarimetric property scales.

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} \qquad (2)$$

where X_{min} and X_{max} represent the minimum and maximum theoretical values of a given polarimetric property.

The rescaled combined thioflavin-negative and thioflavin-positive polarimetric datasets were then randomly split into 2 sets for training (80%) and testing (20%). During training, 10-fold cross-validation was performed to evaluate the accuracy (ACC) and robustness of the classifiers.[33] Three different algorithms were trained to automatically determine the existence of a fluorescence signal: linear discriminant analysis (LDA), SVM, and RF. The hyperparameters (input settings, Appendix B, Table 2) of SVM and RF were optimized by the randomized search function provided in the scikit-learn library,[34] which selects random combinations in a grid of hyperparameters to train the model and return the combination with the best ACC.

Apart from the mean ACC returned from 10-fold cross-validation, we also evaluated the performance of each trained algorithm (classifier) by computing sensitivity (SEN), specificity (SPE), and area under the receiver operating characteristic curve (AUC) on the test set. The evaluations were repeated five times to establish standard deviations. A receiver operating characteristic curve was calculated for each classifier to compare their classification performance.[35]

Finally, the variable importance[36] was obtained from the RF analysis to assess the importance of each polarimetric property in predicting deposit fluorescence. The variable importance is calculated from the average decrease of the Gini impurity when each property is considered. At each node τ, the Gini impurity is calculated as,

$$Gini(\tau) = 1 - \sum_k p_k{}^2 \qquad (3)$$

where k stands for the class (or label) of each dataset, fluorescence positive or fluorescence negative; and P_k stands for the proportion of the sample placed in class k before and after splitting at a node. The property with the smallest Gini impurity is chosen for node splitting.

Results

Visualization of Retinal Deposits

A Mueller matrix of a polarimetric and fluorescence positive deposit is shown in Appendix Figure A1. The deposit is clearly visible against the surrounding retina.

Four examples of the polarimetric properties of retinal deposits (2 fluorescence positive and 2 fluorescence negative) are displayed in Figure 3. In general, these polarimetric properties change from pixel to pixel as well as between deposits and surrounding retina. These deposits have signals in polarimetric properties including LR and LA. Retinal deposits 1 and 2 and both pure protein (amyloid-β) deposits are positive for thioflavin fluorescence. The distributions of the strength of polarimetric properties across retinal deposits differ between fluorescence-positive and -negative deposits, but are similar between fluorescence-positive retinal and pure protein deposits. Using machine learning algorithms, our aim was to identify the differences in polarimetric properties between fluorescence-positive and -negative retinal deposits and predict the existence of a fluorescence signal.

Initially, 16 deposits seemed to be fluorescent positive and polarization negative, but a comparison of the averages of polarization signals from each deposit and surrounding retina found that 3 deposits had weak polarization signals. None of the 16 deposits were included in the datasets.

Performance of the Three Algorithms in Classification

The performance of the 3 algorithms (LDA, SVM, and RF) in classification was assessed using 3 different methods (without oversampling, retinal oversampling method, and oversampling by borderline-SMOTE) and the outcomes are shown in Table 1. The receiver operating characteristic curves are shown in Figure 4 and their AUCs are summarized in Table 2.

ACC is the mean accuracy from 10-fold cross-validation. Shown are the mean values with standard deviations obtained from shuffling and recalculating the training and test sets five times. From Table 1, the SPEs of all classifiers are significantly improved after applying oversampling methods. Among the three classifiers, the performance of LDA is the most sensitive to the choice of oversampling method, with ACC, SEN and SPE decreasing by 11% on average in the borderline-SMOTE method compared with retinal oversampling. SVM had a lower SEN, but a higher SPE in borderline-SMOTE oversampling, giving a negli-
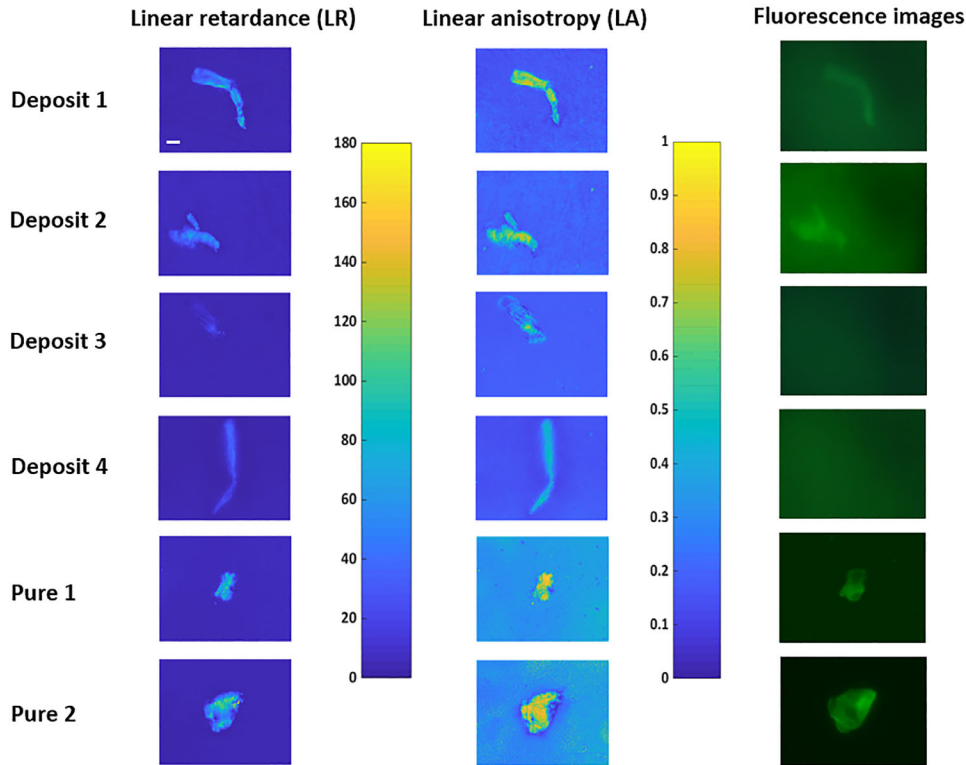
Figure 3.   Linear retardance (LR), linear anisotropy (LA), and fluorescence images of four retinal deposits (1 and 2 have fluorescence signals whereas 3 and 4 do not) and 2 pure Aβ-42 protein deposits with fluorescence signals. Range of LR: [0°, 180°], LA: [0, 1]. Scale bar in the upper left panel: 20 μm.

gible change in ACC. Both SVM and RF perform well and are relatively insensitive to the oversampling method with ACC of RF slightly above SVM at 94% to 95%.

Overall, the AUCs of the three classifiers in Table 2 increased after using oversampling methods, from previously less than 90% to more than 95%, except for the LDA classifier in the borderline-SMOTE method, which was not different from before oversampling. The AUC values for the RF and SVM classifiers were not significantly different between the 2 oversampling methods. The RF classifier has the highest AUC in comparison to LDA and SVM using the borderline-SMOTE oversampling (AUC = 0.986).

Importance of Polarimetric Properties

We ranked the polarimetric properties by their importance as feature inputs in RF analysis in Figure 5 for retinal oversampling and in Figure 6 for the borderline-SMOTE oversampling method.

In retinal oversampling (Fig. 5), LR and LA were the two dominant polarimetric properties (features) in predicting Thioflavin-S fluorescence. Their means and standard deviations together accounted for

74.8% of the total variable importance. Therefore, other feature inputs with lower importance were excluded, and the analysis was rerun with only mean and standard deviation of these two polarimetric properties as feature inputs. Table 3 summarizes the results of retinal oversampling before and after this selection.

From Table 3, there was a decrease in the ACC and SEN for LDA and a small decrease in ACC for RF, but no other significant changes after performing feature selection. Although only four polarimetric properties were used as features, SVM and RF classifiers still achieved more than 90% in ACC. Thus, other polarization metrics can be excluded to speed up the analysis with little loss in performance in predicting deposit fluorescence.

In contrast, the variable importance distribution of each polarization feature was much different under borderline-SMOTE oversampling in Figure 6 than in retinal oversampling with a much smaller range of contribution values (contributions range from 1.5% to 10.1%). Therefore, feature selection was not applied for the borderline-SMOTE oversampling method because there was no evidence of dominant polarimetric properties.
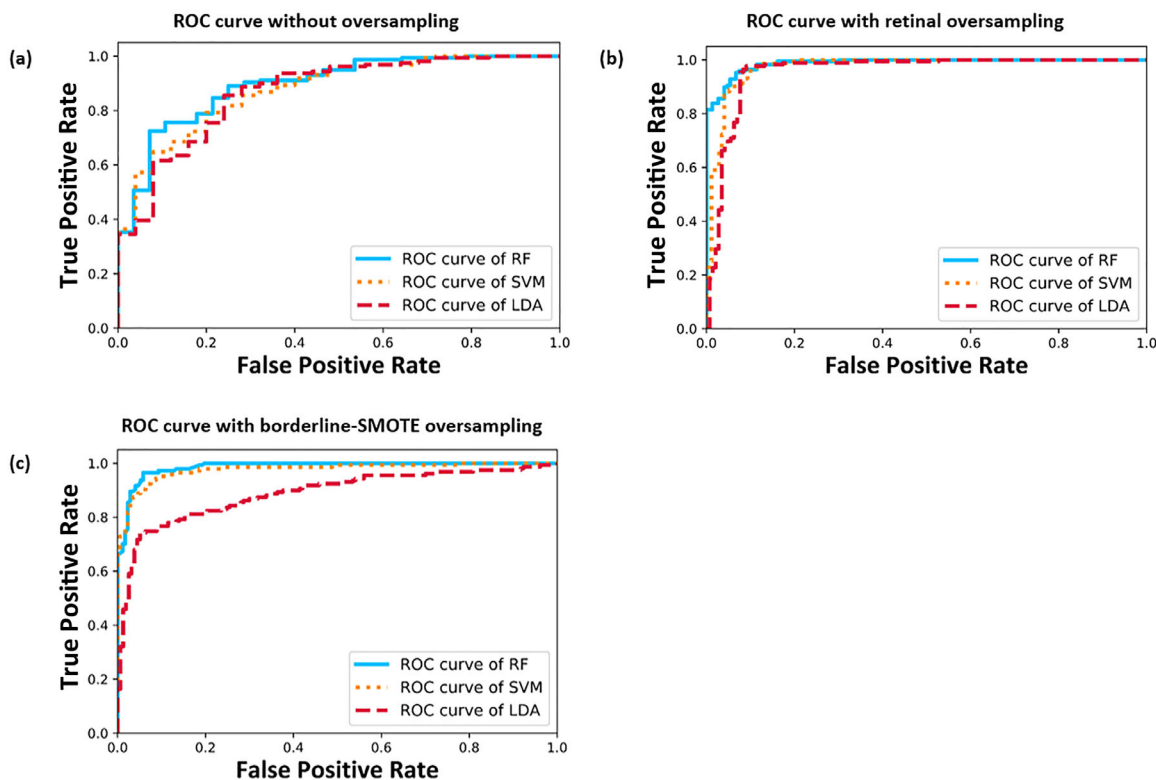
Figure 4. Receiver operating characteristic (ROC) curves for fluorescence signal prediction by the three classifiers under three sampling strategies. The true-positive rate is the sensitivity and the false-positive rate is 1 – specificity. LDA, linear discriminant analysis; RF, random forest; SVM, supporting vector machine.

Table 2. Summary of AUC of the ROCs of the 3 Machine Learning Classifiers Under 3 Sampling Strategies

Method	Without Oversampling AUC	Retinal Oversampling AUC	Borderline-SMOTE Oversampling AUC
LDA	0.863 ± 0.015	0.958 ± 0.003	0.885 ± 0.011
SVM	0.859 ± 0.019	0.967 ± 0.008	0.975 ± 0.003
RF	0.880 ± 0.019	0.981 ± 0.006	0.986 ± 0.007

The area under the receiver operating characteristic curve (AUC of the ROC) is shown $\pm$ standard deviation for each of the 3 classifiers: linear discriminant analysis (LDA), support vector machine (SVM) and random forest (RF) without and with 2 oversampling methods, retinal oversampling and borderline-SMOTE oversampling. The corresponding curves are shown in Figure 4.

Discussion

There are some limitations in this study. First, it is difficult to quantitatively evaluate the similarity between an oversampled dataset and the actual dataset. This factor may affect the generalization and robustness of our models. Because retinal deposits are found sparsely distributed across the retina, retinal oversampling is justified. Although more data from retinal deposits with polarimetric but no fluorescence signals would further validate the reliability of the oversampling methods, because polarimetry predicts thioflavin

positivity with high ACC, such false-negative deposits are scarce.

Second, the limited number of retinas ($n = 28$) and deposits ($n = 920$, before oversampling) involved in this study could bias our models. Thus, it would be helpful to analyze more retinas to enhance the generalization of the models before moving to a clinical setting. However, the number of paired retinas and brains available post mortem is limited. *In vivo* studies could compare retinal polarimetry and *in vivo* brain amyloid measurements in larger sample sizes.

Third, other classification models such as convolutional neural networks should be tested to see if
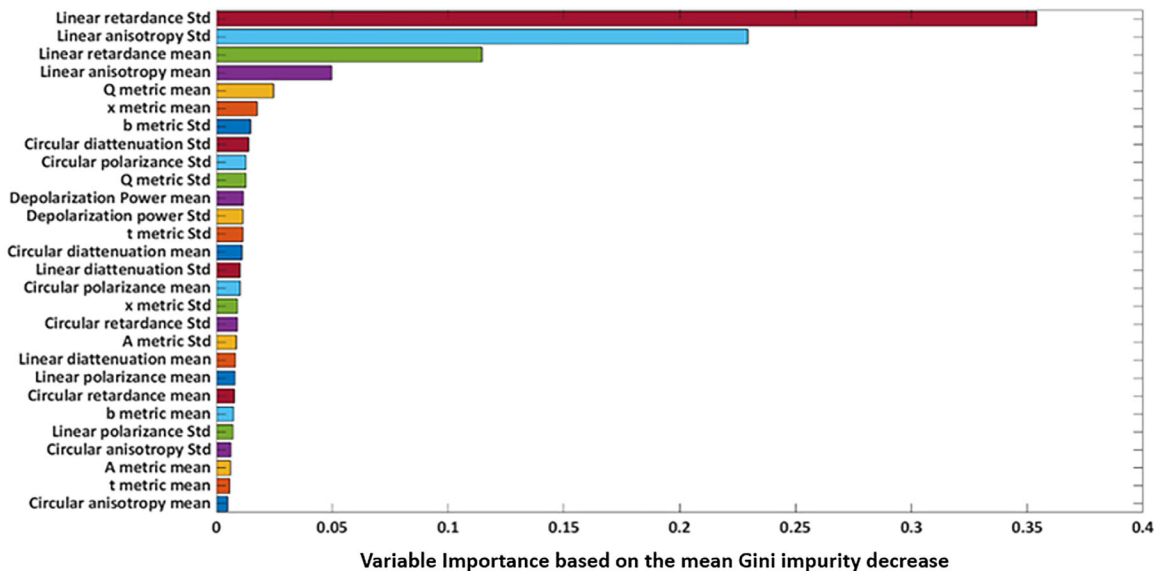
Figure 5. The variable importance of 28 polarimetric properties as feature inputs (means and standard deviations [Std] of 14 polarimetric properties) from a random forest analysis with retinal oversampling. The sum of the variable importance of all features is 1.
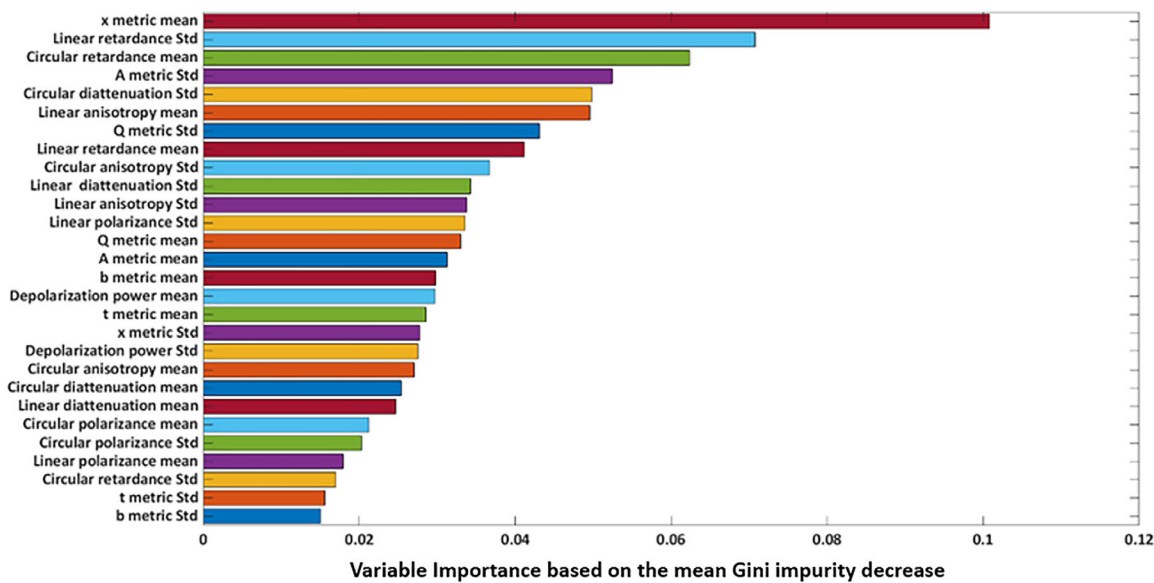

Figure 6. The variable importance of 28 feature inputs (means and standard deviations [Std] of 14 polarimetric properties) for a random forest analysis with borderline-SMOTE oversampling. The sum of the variable importance of all feature is 1.

they can further improve the already excellent ACC of the classification of thioflavin positive (presumed amyloid) versus thioflavin-negative deposits. Considering the ACC as well as the computation/memory and time costs of the models, we could then select the one that would be best for clinical application.

Amyloid stained with Congo red is known to interact with polarized light producing an apple green birefringence signal.[37,38] We have shown that unstained pure amyloid-β and thioflavin positive retinal amyloid deposits also interact with polarized light and show

very similar variations of LR and depolarization across deposits (Corapi F, et al. *IOVS* 2018;59:ARVO E-Abstract 1582). These similarities and thioflavin positivity[39] are presumably due to the fibrillary, well-ordered structure of these deposits.[39] Thus, it is not unexpected that LR and polarimetric properties which are related to LR (such as LA) are important to distinguishing thioflavin positive (amyloid) deposits from those that are not thioflavin positive.

Amyloid deposits show a range of polarimetric property strengths, both within and between deposits,

Table 3. Classification Performance by Retinal Oversampling Before and After Applying Feature Selection

Method	Before Feature Selection			After Feature Selection		
	ACC	SEN	SPE	ACC	SEN	SPE
LDA	$90.9 \pm 0.0\%$	$91.1 \pm 3.1\%$	$93.4 \pm 1.2\%$	$89.1 \pm 0.3\%$	$84.5 \pm 3.4\%$	$94.4 \pm 1.6\%$
SVM	$92.5 \pm 0.3\%$	$95.6 \pm 0.7\%$	$91.4 \pm 0.7\%$	$92.4 \pm 0.3\%$	$94.0 \pm 1.4\%$	$92.0 \pm 2.1\%$
RF	$93.7 \pm 0.2\%$	$95.5 \pm 1.3\%$	$92.1 \pm 2.1\%$	$92.8 \pm 0.5\%$	$94.9 \pm 0.7\%$	$91.4 \pm 2.4\%$

The average accuracy (ACC), sensitivity (SEN), and specificity (SPE) $\pm$ standard deviation before feature selection (with all polarimetric properties included) and after feature selection (including a small subset with high variable importance shown in Figure 5) for each of the 3 classifiers: linear discriminant analysis (LDA), support vector machine (SVM) and random forest (RF) for the retinal oversampling method.

potentially owing to a variety of factors, including variations in thickness (thinner regions have weaker signals, usually at the edge of deposits (e.g., deposit 2 in Fig. 3), in the uniformity of orientation of the fibrils and potential differences in amyloid species. Thioflavin has been known to stain amyloid species other than amyloid-β, such as alpha-synuclein.[16] By choosing retinas from donors with a brain pathology of amyloid deposits of the Alzheimer's type, we expect the majority of thioflavin positive deposits to contain amyloid-β. The lack of polarization signals for a small number ($n = 16$) of thioflavin positive deposits may indicate one or more of the following: thinner deposits (Tao Jin, et al. *IOVS* 2017;58:ARVO E-Abstract 3367), more disordered deposits, or the presence of other impurities (analogous to neuritic material found in brain plaques).[40]

Despite the range of polarimetric properties of deposits in the retina, which indicates differing interactions with polarized light, in the present study, we trained machine learning classifiers to precisely and robustly distinguish polarimetry-positive retinal deposits with and without thioflavin fluorescence signals. In turn, thioflavin positivity defines a presumed amyloid deposit. In this study, the RF and SVM classifiers exhibited high ACC, SEN, and SPE. Two oversampling methods proved to be useful for dealing with the larger numbers of fluorescence positive versus negative deposits and improved the SPE, AUC and, in all but one case, the ACC of the classifications.

Using the RF algorithm with the retinal oversampling method, we found that the means and standard deviations of two polarimetric properties—LR and LA—have superior performance in predicting fluorescence positivity. We can use only these four properties as feature inputs to produce excellent classification performance. From the perspective of the physical meanings of these two polarimetric properties, LR is directly related to the linear birefringence which results from the ordered alignment of the fibrils. LA is a measure of the polarimetric anisotropy of samples that can be interpreted as the different optical responses to light polarized in different directions caused by the samples' LR and linear diattenuation.[26,41] LA is postulated to have a periodic variation that reflects the geometric structure of samples,[26] such as the orientation of fibrillary structures. The high importance of these two polarimetric properties suggests that the ordered arrangement of fibrils may be the key to distinguishing between fluorescent and nonfluorescent deposits. Given the similarities between LR and LA patterns in deposits in Figure 3, for amyloid, retardation seems to contribute more to LA than diattenuation.

In the case of the second (borderline-SMOTE) oversampling method, the variable importance returned from RF analysis indicates that LA and LR are no longer dominant variables. The contributions of all variables are within one order of magnitude. This difference in variable contribution is likely caused by the differing data introduced by the two oversampling methods. We performed a t-distributed stochastic neighbor embedding (t-SNE)[42] (see Appendix A) to show the distributions of the polarimetric properties of each deposit for each oversampling method. The t-SNE algorithm maps the local distance information from the original high-dimensional space (of 14 properties) to a low-dimensional space (two properties). Perplexity, whose value is varied, provides a smooth measurement of the number considered as neighbors, which alters the algorithm's attention on local and global information (local variations dominate at a small perplexity). The t-SNE visualization plots for differing values of perplexity are displayed in Figure 7a for the retina oversampling method and Figure 7b for the borderline-SMOTE oversampling method.

We can see in the t-SNE perplexity plots (Figs. 7a and 7b), that the data points of fluorescent negative and fluorescent positive areas are somewhat more separated in the retinal oversampling method than
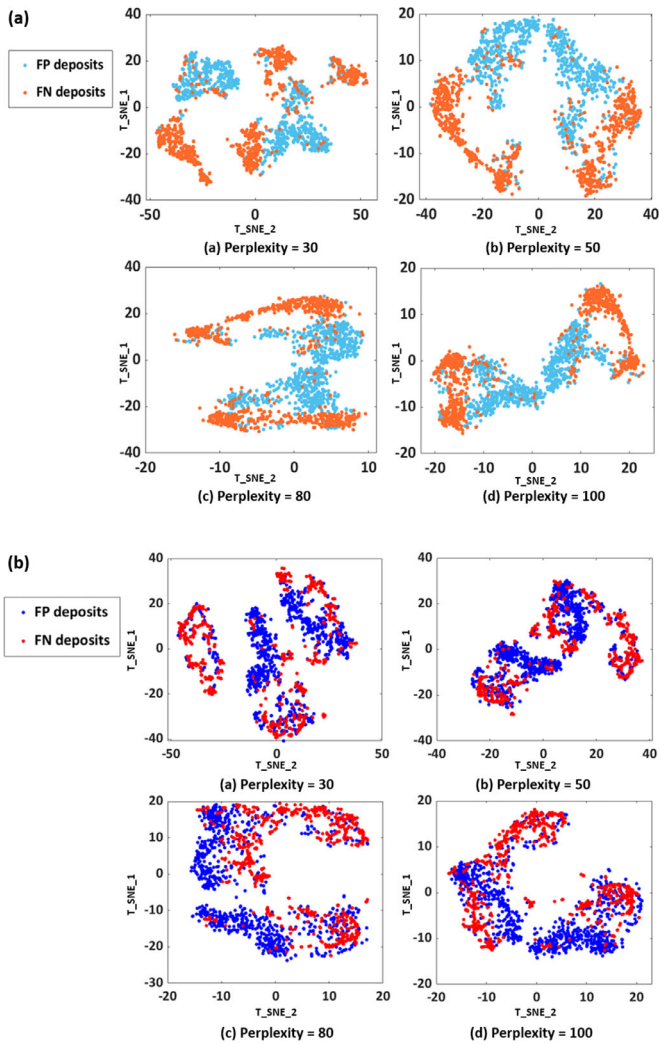
Figure 7. Two-dimensional t-distributed stochastic neighbor embedding (t-SNE) visualization of the polarization properties of the fluorescent positive and fluorescent negative datasets from (a) the retinal oversampling method and (b) the borderline-SMOTE oversampling method under different values of perplexity. The distribution focus moves from the local to the global as the perplexity increases. FP and FN are fluorescent positive and negative deposits, respectively. T_SNE_1 and T_SNE_2 are the axis of the 2-dimensional space to which the t-SNE mapped the higher dimensions.

in the borderline-SMOTE oversampling method. As the perplexity increases, the sample distribution with negative fluorescence in retinal oversampling tends to aggregate at both ends of the dataset, whereas the borderline-SMOTE oversampling method produces a more complex data distribution. This finding implies that the properties of deposits identified in the borderline-SMOTE oversampling are closer to those of the fluorescent-positive deposits than in retinal oversampling. The differing data distributions are consistent with lower ACC for LDA with borderline-SMOTE. It is also not surprising that, for the two

oversampling methods, differing polarimetric properties differentiate fluorescent positive from fluorescent-negative deposits.

The methods described herein predict from the polarimetric properties of retinal deposits, with high ACC, the existence of amyloid with thioflavin positive fluorescence signals. We have shown that, in combination with machine learning algorithms, imaging using Mueller matrix polarimetry can detect amyloid positive deposits in the *ex vivo* retina without using a dye. We have previously reported that the polarimetric signals from retinal amyloid deposits measured in double pass, analogous to the live eye imaging method proposed in Campbell's patents,[43] give polarimetric signals which are twice as large as those seen in single pass measurements (Photonics North presentation, 2016; M.C.W. Campbell, et al. *IOVS* 2020; 61:ARVO E-Abstract 1850). Our preferred live eye implementation uses a confocal scanning laser ophthalmoscope combined with polarimetry. Using the methods described here, we expect that thioflavin-positive retinal deposits could be identified with high ACC in live eye imaging. In turn, we have previously shown that the number of retinal amyloid deposits predicts the severity of amyloid in the brain.[15] Therefore, *in vivo* dye-free polarimetric imaging of deposits in the retina could determine brain amyloid noninvasively, in contrast with current invasive methods. The results reported here are an important step towards the development of a novel, noninvasive, clinical method of imaging retinal amyloid as a predictor of brain amyloid.

Acknowledgments

Disclosure: **Y. Qiu,** P; **T. Jin,** None; **E. Mason,** None; **M.C.W. Campbell,** LumeNeuro (I,S); University of Waterloo (P)

References

1. World Health Organization., Alzheimer's Disease International. *Dementia: a public health priority*. Genera: World Health Organization; 2012.

2. Hyman BT, Phelps CH, Beach TG, et al. National Institute on Aging–Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease. *Alzheimer's Dement*. 2012;8:1–13, doi:10.1016/J.JALZ.2011.10.007.

3. Jack CR, Bennett DA, Blennow K, et al. NIA-AA Research framework: toward a biological definition of Alzheimer's disease. *Alzheimer's Dement*. 2018;14:535–562, doi:10.1016/j.jalz.2018.02.018.

4. Klunk WE, Engler H, Nordberg A, et al. Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound-B. *Ann Neurol*. 2004;55:306–319, doi:10.1002/ana.20009.

5. Anoop A, Singh PK, Jacob RS, Maji SK. CSF biomarkers for Alzheimer's disease diagnosis. *Int J Alzheimers Dis*. 2010;2010:606802, doi:10.4061/2010/606802.

6. Ising C, Stanley M, Holtzman D. Current thinking on the mechanistic basis of Alzheimer's and implications for drug development. *Clin Pharmacol Ther*. 2015;98:469–471, doi:10.1002/cpt.200.

7. Sabbagh MN, Blennow K. Peripheral biomarkers for Alzheimer's disease: update and progress. *Neurol Ther*. 2019;8:33–36, doi:10.1007/s40120-019-00171-6.

8. Lim JKH, Li QX, He Z, et al. The eye as a biomarker for Alzheimer's disease. *Front Neurosci*. 2016;10:536, doi:10.3389/fnins.2016.00536.

9. Begum A, Kvanta A, Seregard S, et al. Retinal amyloid pathology and proof-of-concept imaging trial in Alzheimer's disease. *J Pharmacol Exp Ther*. 2017;326:196–208, doi:10.1172/JCI.INSIGHT.93621.

10. Hart NJ, Koronyo Y, Black KL, Koronyo-Hamaoui M. Ocular indicators of Alzheimer's: exploring disease in the retina. *Acta Neuropathol*. 2016;132:767–787, doi:10.1007/s00401-016-1613-6.

11. Koronyo-Hamaoui M, Koronyo Y, Ljubimov AV., et al. Identification of amyloid plaques in retinas from Alzheimer's patients and noninvasive *in vivo* optical imaging of retinal plaques in a mouse model. *Neuroimage*. 2011;54:S204–S217, doi:10.1016/j.neuroimage.2010.06.020.

12. den Haan J, Morrema THJ, Verbraak FD, et al. Amyloid-beta and phosphorylated tau in post-mortem Alzheimer's disease retinas. *Acta Neuropathol Commun*. 2018;6:147, doi:10.1186/s40478-018-0650-x.

13. Westermark G., Johnson K., Westermark P. Staining methods for identification of amyloid in tissue. *Methods Enzymol*. 1999;309:3–25.

14. Koronyo Y, Biggs D, Barron E, et al. Retinal amyloid pathology and proof-of-concept imaging trial in Alzheimer's disease. *JCI insight*. 2017;2:e93621, doi:10.1172/jci.insight.93621.

15. Campbell MCW, Emptage L, Redekop R, et al. Amyloid deposits imaged in postmortem retinas using polarimetry predict the severity of a postmortem brain based diagnosis of Alzheimer's disease. *Alzheimer's Dement*. 2018;14:P774–P775, doi:10.1016/j.jalz.2018.06.947.

16. Wördehoff MM, Hoyer W. α-Synuclein aggregation monitored by Thioflavin T fluorescence assay. *Bio-protocol*. 2018;8:e2941, doi:10.21769/BioProtoc.2941.

17. Eberling JL, Dave KD, Frasier MA. α-synuclein imaging: a critical need for Parkinson's disease research. *J Parkinsons Dis*. 2013;3:565–567, doi:10.3233/JPD-130247.

18. Kepe V, Moghbel MC, Långströmd B, et al. Amyloid-β positron emission tomography imaging probes: a critical review. *J Alzheimer's Dis*. 2013;36:613–631, doi:10.3233/JAD-130485.

19. Magnin B, Mesrob L, Kinkingnéhun S, et al. Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI. *Neuroradiology*. 2009;51:73–83, doi:10.1007/s00234-008-0463-x.

20. Zhang Y, Dong Z, Phillips P, et al. Detection of subjects and brain regions related to Alzheimer's disease using 3D MRI scans based on eigenbrain and machine learning. *Front Comput Neurosci*. 2015;9:1–15, doi:10.3389/fncom.2015.0006.

21. Lebedev AV, Westman E, Van Westen GJP, et al. Random Forest ensembles for detection and prediction of Alzheimer's disease with a good between-cohort robustness. *NeuroImage Clin*. 2014;6:115–125, doi:10.1016/j.nicl.2014.08.023.

22. Fuller AR, Zawadzki RJ, Choi S, Wiley DF, Werner JS, Hamann B. Segmentation of three-dimensional retinal image data. *IEEE Trans Vis Comput Graph*. 2007;13:1719–1726, doi:10.1109/TVCG.2007.70590.

23. Li X, Liao R, Zhou J, Leung PTY, Yan M, Ma H. Classification of morphologically similar algae and cyanobacteria using Mueller matrix imaging and convolutional neural networks. *Appl Opt*. 2017;56:6520, doi:10.1364/ao.56.006520.

24. Michael B, Casimer D, Jay E, et al. *Handbook of Optics*. 3rd ed. New York: The McGraw-Hill Companies, Inc.; 2009.

25. Lu S, Chipman RA. Interpretation of Mueller matrices based on polar decomposition. *J Opt Soc Am A*. 1996;13:1106–1113.

26. Arteaga O, Garcia-Caurel E, Ossikovski R. Anisotropy coefficients of a Mueller matrix. *J Opt Soc Am A*. 2011;28:548, doi:10.1364/josaa.28.000548.

27. Sun M, He H, Zeng N, et al. Characterizing the microstructures of biological tissues using Mueller matrix and transformed polarization parameters. *Biomed Opt Express*. 2014;5:329–340, doi:10.1364/BOE.5.004223.

28. Zeng N. A possible quantitative Mueller matrix transformation technique for anisotropic scattering media/Eine mögliche ... A possible quantitative Mueller matrix transformation technique for anisotropic scattering media. *Photonics & Lasers in Medicine*. 2013;2:139–137, doi:10.1515/plm-2012-0052.

29. Espinosa-Luna R, Bernabeu E. On the Q(M) depolarization metric. *Opt Commun*. 2007;277:256–258, doi:10.1016/j.optcom.2007.05.051.

30. Rahman MM, Davis DN. Addressing the class imbalance problem in medical datasets. *Int J Machine Learn Comput*. 2013;3:224–228, doi:10.7763/IJMLC.2013.V3.307.

31. Han H, Wang W, Mao B. Borderline-SMOTE: a new over-sampling method. *Adv Intell Comput*. 2005;3644:878–887.

32. Chawla N V, Bowyer KW, Hall LO. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res*. 2002;16:321–357.

33. Stone M. Cross-validatory choice and assessment of statistical predictions (with discussion). *J R Stat Soc Ser B*. 1976;38:102–102, doi:10.1111/j.2517-6161.1976.tb01573.x.

34. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12:2825—2830.

35. Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit*. 1997;30:1145–1159, doi:10.1016/S0031-3203(96)00142-2.

36. Menze BH, Kelm BM, Masuch R, et al. A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC Bioinformatics*. 2009;16:1–16, doi:10.1186/1471-2105-10-213.

37. Yakupova EI, Bobyleva LG, Vikhlyantsev IM, Bobylev AG. Congo Red and

38. Howie AJ, Brewer DB. Optical properties of amyloid stained by Congo red: history and mechanisms. *Micron*. 2009;40:285–301, doi:10.1016/j.micron.2008.10.002.

39. Krebs MRH, Bromley EHC, Donald AM. The binding of thioflavin-T to amyloid fibrils: localisation and implications. *J Struct Biol*. 2005;149:30–37, doi:10.1016/J.JSB.2004.08.002.

40. Atwood CS, Martins RN, Smith MA, Perry G. Senile plaque composition and posttranslational modification of amyloid-β peptide and associated proteins. *Peptides*. 2002;23:1343–1350, doi:10.1016/S0196-9781(02)00070-0.

41. José JGP, Razvigor O. *Polarized Light and the Mueller Matrix Approach*. Boca Raton, FL: CRC Press; 2016.

42. Van Der Maaten L, Hinton G. Visualizing data using t-SNE. *J Machine Learn Res*. 2008;9:2579–2605.

43. Campbell MCW. Method and system for imaging amyloid beta in the retina of the eye in association with Alzheimer's disease. United States patent US 9,149,184. 2015.

amyloids: history and relationship. *Biosci Rep*. 2019;39:BSR20181415, doi:10.1042/BSR20181415.

Appendix A: Definitions of Machine Learning Terms not Defined in the Text

Words and phrases in bold are used in the main text.

Feature: A **feature** is a property input to the **machine learning algorithm**. In this article we use polarimetric properties which describe interactions of the sample with polarized light.

Datasets of each class: The groups into which the data is classified, each of which has a unique label, are referred to as **classes**. In our case, we have 2 datasets of deposits labelled with the **classes**, thioflavin positive and thioflavin negative, both members of which are described by polarimetric properties. When datasets contain unequal numbers of data, this is referred to as **data imbalance**.

Machine learning algorithm: Computer programs (algorithms) that can "learn" from **training data** and improve their performance in predicting the **class** of input **testing data**. Our 2 datasets are labelled thioflavin positive and thioflavin negative. After learning **training**, these programs are given **testing data**, which
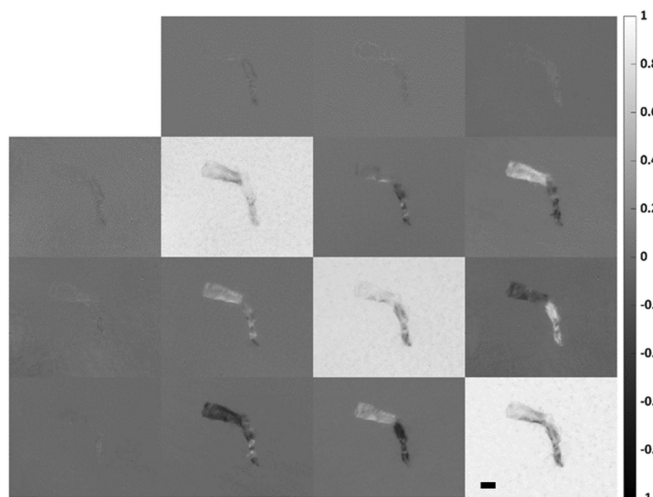
Figure A1. A spatially resolved Mueller matrix (MM) of a polarimetric and fluorescence positive deposit. Elements are coded by row and column number (M00 to M33) and each pixel position has 16 elements associated with it. The matrix elements are normalized by M00. Scale bar: 20 μm.

they classify into the **classes**. These algorithms are also known as **classifiers**. The algorithms we test are:

Linear discriminant analysis (LDA): Uses **training data** to learn linear equations that produce values that are close for members of the same **dataset**, and far apart for members of different **datasets**. These equations are then used to classify **testing data** into **classes**.

Supporting vector machine (SVM): When **training, SVM** finds a decision boundary in the **feature** space to separate the different **datasets** as far as possible. **Testing data** are classified based on which side of the boundary they fall on. **SVM** heavily weights data close to the decision boundary.

Random forest (RF): Uses many decision trees to generate a consensus decision. A decision tree classifies data points by using multiple properties to differentiate data into separate **classes**. When **training**, many trees fit a random subset of the **training** set. Each data point is available to be sampled multiple times during training. This is called bootstrapping. When testing, bootstrapping again generates many subsets, and each is classified. This classification uses a decision tree based on **Gini impurity**. Data points are then classified into the **class** where it was most frequently placed.

Gini impurity: is a commonly used metric in **RF** decision trees to decide the optimal property to split the data into subsets. In general, as the splitting continues, the amount of data in branch nodes of a decision tree which belong to the same **class** should increase, this implies that the "purity" of the nodes is increasing. The purity of a **dataset** varies inversely with Gini

impurity (Equation 3). The property with the smallest Gini impurity is chosen for node splitting. The decrease in Gini impurity at each node over the tree is summed to calculate **the variable importance of each property**.

Oversampling strategy: If the **datasets** of each **class** are not balanced in their number of data points, these methods can be applied to increase the data points in the dataset with lower numbers (**minority**, in our case thioflavin negative deposits). Our **majority dataset** consists of thioflavin positive deposits.

***K*-nearest neighbor analysis, used in oversampling:** Determine the k points nearest a given point (from the nearest to the *k*th nearest). The points are defined by the values of the mean and standard deviation of each of the polarimetric properties. The nearest points have the most similar properties.

SMOTE algorithm: A commonly used algorithm to generate synthetic samples (**oversample**) in the **minority dataset** to balance the number in the **majority dataset**. The algorithm begins by searching for the *k*-nearest neighbors of the same class (*k* is usually set to 5) for every sample in the **minority dataset**, then randomly generates a number of new synthetic data points along the line between the minority datapoint and its nearest neighbors. The number of synthetic data points matches the total number in the **minority dataset** to the **majority dataset**.

Borderline-SMOTE oversampling: Based on SMOTE oversampling, **borderline-SMOTE** only performs the **SMOTE algorithm** on examples in the **minority dataset** that have nearest neighbors in the **majority dataset** as these are easily misclassified.

Cross-validation: (10-fold): Divides the data into 10 subsets with similar size. One subset is used as the **testing data** and the other 9 are the **training data**. This is repeated for the other 9 subsets and the average of the 10 results is output.

t-Distributed stochastic neighbor embedding (t-SNE)[42] **test:** A machine learning method for dimension reduction that can help to identify the patterns of properties in the **datasets**. The main advantage of t-SNE over other dimension reduction methods (such as principal component analysis) is the ability to maintain local patterns in the dataset. This means that similar points in high-dimensional space are still close to each other in a lower-dimensional projection.

Perplexity is a parameter that controls the dimension reduction in **t-SNE**. It can be interpreted as the nearest neighbors considered when matching the original (high-dimensional space) and the fitted (low-dimensional space) of each point. A smaller

perplexity means only a few points are considered as the nearest neighbors when doing fitting, which highlights the information local to the data points. A higher perplexity means that more points are considered as the nearest neighbors, giving a more "global view."

Appendix B:

Table B1. Equations for the Polarimetric Properties Used as Features in the Machine Learning Algorithms

Feature Name	Equation
Linear retardance (LR)	$\dfrac{\sqrt{(M_{R23}-M_{R32})^2+(M_{R31}-M_{R13})^2}}{2sin\{cos^{-1}[\frac{tr(M_R)}{2}-1]\}}$
Circular retardance	$\dfrac{M_{R12}-M_{R21}}{2sin\{cos^{-1}[\frac{tr(M_R)}{2}-1]\}}$
Linear diattenuation	$\sqrt{\left(M_{01}^2+M_{02}^2\right)}$
Circular diattenuation	M_{03}
Depolarization power	$1-\dfrac{|tr(M_\Delta)-1|}{3}$
Metric b	$\dfrac{M_{22}+M_{33}}{2}$
Metric t	$\dfrac{\sqrt{(M_{22}-M_{33})^2+(M_{23}+M_{32})^2}}{2}$
Metric A	$\dfrac{2b*t}{b^2+t^2}$
Metric x	$\dfrac{tan^{-1}(\frac{M_{31}}{M_{21}})}{2}$
Linear polarizance	$\sqrt{M_{10}^2+M_{20}^2}$
Circular polarizance	M_{30}
Q metric	$\dfrac{\frac{\sum_{i,j=0}^{3}(M_{ij}^2-M_{00}^2)}{M_{00}^2}-M_{01}^2-M_{02}^2-M_{03}^2}{1+M_{01}^2+M_{02}^2+M_{03}^2}$
Linear anisotropy (LA)	$\dfrac{\sqrt{(M_{01}+M_{10})^2+(M_{02}+M_{20})^2+(M_{23}-M_{32})^2+(M_{13}-M_{31})^2}}{\sqrt{\Sigma}}$
Circular anisotropy	$\dfrac{\sqrt{(M_{03}+M_{30})^2+(M_{12}-M_{21})^2}}{\sqrt{\Sigma}}$

In this table, $\Sigma=(3M_{00}^2-M_{11}^2-M_{22}^2-M_{33}^2)+2(M_{01}M_{10}+M_{02}M_{20}+M_{03}M_{30}-M_{23}M_{32}-M_{13}M_{31}-M_{12}M_{21})$. M_{ij} is the Mueller matrix element in the ith row and jth column. M_Δ and M_R are the depolarization and retardance matrices obtained from polar decomposition.[25]

Table B2. Hyperparameters of SVM and RF Where the Parameters Used Are Defined in[34]

Classifier	Hyperparameter Distributions
SVM	Kernel: [radial basis function, polynomial, sigmoid] Degree of polynomial kernel function: [1,2,3] Gamma: [0.001, 0.01, 0.1] Shrinking: [True, False] Cost parameter: 1 Search iteration:30
RF	Number of trees: a random integer in [100,1000] Bootstrap: [True, False] Max features of splitting: a random integer in [1,28] Minimal sample of splitting: a random integer in [2,27] Criterion: [Gini impurity, information gain] Maximum tree depth: unlimited Search iteration: 100

translational vision science & technology