University of Dundee

Retinal Biomarker Discovery for Dementia in an Elderly Diabetic Population

Fetit, Ahmed; Manivannan, Siyamalan; McGrory, Sarah; Ballerini, Lucia; Doney, Alexander; MacGillivray, Thomas J.

Retinal Biomarker Discovery for Dementia in an Elderly Diabetic Population

Ahmed E. Fetit[1], Siyamalan Manivannan[1], Sarah McGrory[2], Lucia Ballerini[3,4],
Alexander Doney[1], Thomas J. MacGillivray[3], Ian J. Deary[2], Joanna M. Wardlaw[4],
Fergus Doubal[4], Gareth J. McKay[5], Stephen J. McKenna[1] and Emanuele Trucco[1]

[1] VAMPIRE project, CVIP, Computing (SSE), University of Dundee, UK.
[2] Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, UK.
[3] VAMPIRE project, Centre for Clinical Brain Sciences, University of Edinburgh, UK.
[4] Centre for Clinical Brain Sciences, University of Edinburgh, UK.
[5] Centre for Public Health, Queen's University Belfast, UK.

afetit@dundee.ac.uk

Abstract. Dementia is a devastating disease, and has severe implications on affected individuals, their family and wider society. A growing body of literature is studying the association of retinal microvasculature measurement with dementia. We present a pilot study testing the strength of groups of conventional (semantic) and texture-based (non-semantic) measurements extracted from retinal fundus camera images to classify patients with and without dementia. We performed a 500-trial bootstrap analysis with regularized logistic regression on a cohort of 1,742 elderly diabetic individuals (median age 72.2). Age was the strongest predictor for this elderly cohort. Semantic retinal measurements featured in up to 81% of the bootstrap trials, with arterial caliber and optic disk size chosen most often, suggesting that they do *complement* age when selected together in a classifier. Textural features were able to train classifiers that match the performance of age, suggesting they are potentially a rich source of information for dementia outcome classification.

Keywords: Retina, Dementia, Microvasculature, Classification, Biomarkers.

1   Introduction

Dementia is an umbrella term used to describe a set of brain disorders that trigger a loss of cognitive brain function. There are approximately 850,000 people with dementia in the UK, with numbers set to exceed one million by 2025 [1]. Alzheimer's Disease International reports an estimated 50 million people worldwide living with dementia in 2017, with the numbers affected expected to almost double every 20 years, reaching 75 million by 2030. The total worldwide estimated cost of dementia was put at US$818 billion in 2015, representing 1.09% of the global GDP.

Identifying individuals at an increased risk of developing dementia later in life earlier and detecting pre-clinical stages of cognitive decline may provide opportuni-

ties to preserve brain function and delay disease progression. The potential of retinal imaging to support early detection and risk stratification is under investigation, with studies relating the retina microvasculature with changes in the cerebral microvasculature (see [3, 11] for recent reviews). Anatomically and developmentally, the retina is an extension of the brain, hence the retinal microvasculature could work as an easily observed proxy reflecting the condition of the cerebral vasculature [2, 4, 5, 10, 11].

McGrory et al. [3] reviewed the application of fundus camera imaging to assess the associations between retinal microvascular changes and dementia, including various subtypes (i.e. Alzheimer's disease (AD), vascular and frontotemporal dementia). *Vis-á-vis* the heterogeneity among studies in terms of experimental design and the retinal parameters assessed, the most consistent finding identified was that a decreased fractal dimension (a global measure of branching complexity of the retinal vascular tree) tends to associate with AD, as reported by Williams et al. [4] who analyzed data from 507 participants, and Frost et al. [5] who used data from the Australian Imaging, Biomarkers and Lifestyle study of ageing.

We consider two categories of features computable from fundus camera imaging: **(i) clinically semantic features**, or measurements with direct clinical interpretation, e.g. optic disc radius, arterial/venular caliber, tortuosity; and **(ii) clinically non-semantic features**, or measurements that do not have a direct clinical interpretation, but may capture valuable patterns in terms of biomarkers. For these we use texture, a characterization of the spatial variation of pixel intensities, e.g. entropy and contrast computed from co-occurrence matrices [18].

We contribute to the ongoing debate on the value of retinal vascular features for assessing dementia and predicting its risk [3] with the results of a bootstrap analysis with regularized logistic regression and cross-validation in a cohort of elderly diabetic patients. This analysis determines which semantic feature *sets* are selected most often in building sparse logistic regression models, indicating their value for associations. The dementia outcome is defined as whether or not a patient record indicates dementia within a certain time frame relative to the retinal scan.

2 Materials and Methods

2.1 Dataset

Fundus camera images (3504x2336 pixels) were obtained from the Genetics of Diabetes Audit and Research Tayside (GoDARTS) bio-resource, Scotland [12]. Images of 2,103 diabetic individuals were available, of which 1,742 patients matched our quality control criteria reported elsewhere [13]. Association with dementia was determined through linkage with prescription, hospital admission and other medical record repositories, using the date of the first recorded event of dementia in the patient records. Of the 1,742 participants included, 237 were identified as having developed dementia by the date the records were inspected. These patients' earliest recorded dementia event ranged from 2,144 days (5.87 years) before the retinal image capture to 2,929 days (8.02 years) after. The mean and median times to dementia event were 952 and 1033 days post-capture (2.6 and 2.8 years), respectively; standard deviation was 1,055 days

(2.9 years). Sex and age at retinal photograph were included in our pilot analysis. The mean and median age at scan were 75.7 and 76.1 years for dementia-associated patients, and 69.3 and 71.4 years for the remaining subjects. Of the dementia-associated patients, 105 were female and 132 were male; of the remaining subjects 646 were female and 859 were male. The study was carried out in full accordance with the current data governance and ethical approval regulations in the UK.

2.2   Feature extraction

Clinically semantic features. A single operator (LB), trained with an established protocol, used the semi-automatic VAMPIRE 3.0 (Vascular Assessment and Measurement Platform for Images of the Retina) [14, 15] to measure optic disc (OD) radius, central retinal arteriolar equivalent (CRAE) and central retinal venular equivalent (CRVE), retinal arterio-venule-ratio (AVR), tortuosity of arteries (tortA) and veins (tortV), as well as maximum tortuosity (tortAmax and tortVmax) in right-eye images, after verifying sufficient correlation ($r > $ ~0.6) of AVR, CRAE, CRVE with the left eye as done by others, e.g. [20].

Clinically non-semantic features. Using MATLAB R2016b (MathWorks, Massachusetts, USA), Grey-Level Co-Occurrence Matrix (GLCM) [18] and Grey-Level Run-Length Matrix (GLRLM) [19] features were computed. Right-eye images were used as above (high-contrast green channel). Each image was divided into four standard quadrants and 95 textural features were extracted from each quadrant. Additionally, global features summarizing the overall image texture were extracted, making it possible to build multi-scale feature pyramids.

In terms of GLCM, 1-pixel distance was chosen for displacement along the 0°, 45°, 90° and 135° directions. The number of grey-levels was quantized to 256 (standardizing the sizes of the GLCMs). The features were those provided by the library [6] and included autocorrelation, correlation, energy, and entropy. For the GLRLM, the number of grey-levels was also quantized to 256. Features extracted were those provided by library [9] and included short run emphasis, long run emphasis and grey-level non-uniformity. The full texture pyramid was a concatenation of the global features (95) with the four quadrants' features (380), yielding a total of 475 features per image.

2.3   Biomarker discovery by regularized logistic regression

Our goal was to deliver a pipeline for (i) assessment of the utility of retinal features for classifying dementia outcome, and (ii) automatic identification of subsets of the available semantic retinal features which give rise to models that are parsimonious and that predict dementia outcome effectively, if any.

 Logistic regression was used for ease of model interpretation; model coefficients can be directly linked to feature importance. Regularization was used to perform shrinkage and drive feature coefficients towards zero. This approach is well suited for biomarker identification; it performs feature selection simultaneously with model estimation. For a particular choice of regularization parameters, λ and $\alpha \in [0, 1]$,

model estimation given N training pairs $(\mathbf{x}_i, y_i)$, consists of minimizing the penalized negative log-likelihood,

$$\min_{(\beta_0, \boldsymbol{\beta}) \in \mathbb{R}^{p+1}} -\left[\frac{1}{N}\sum_{i=1}^{N} y_i \, (\beta_0 + \mathbf{x}_i^T\boldsymbol{\beta}) - \log\left(1 + e^{(\beta_0 + \mathbf{x}_i^T\boldsymbol{\beta})}\right)\right] + \lambda\left[\frac{(1-\alpha)\|\boldsymbol{\beta}\|_2^2}{2} + \alpha\|\boldsymbol{\beta}\|_1\right] \qquad (1)$$

where β_0 and $\boldsymbol{\beta} = [\beta_1, \beta_2, ..., \beta_N]$ are the logistic model coefficients. The parameter λ controls the strength of regularization. The parameter α is the elastic-net regularization mixing parameter; $\alpha = 1$ corresponds to the lasso (also known as L_1 regularization) whereas $\alpha = 0$ corresponds to ridge regression (L_2 regularization) [8]. For highly correlated features, the lasso tends to pick one of the features and discard the others, whereas the ridge shrinks the feature coefficients towards each other. Elastic net mixes the two; for example, $\alpha = 0.5$ tends to select or exclude groups of highly correlated features together [8].

The *R* package *glmnet* [8] was used to perform the regularized logistic regression experiments. *glmnet* implements an efficient algorithm for computing entire regularization paths, showing the effect of varying λ on the classification error and the number of features retained in the model. Cross-validation was used to select λ; two values of interest are reported by *glmnet*: $\lambda_{\mathbf{min}}$, the value of λ at which the lowest validation set error is achieved, and $\lambda_{\mathbf{1SE}}$, the λ for the most regularized model whose validation error is within one standard error of λ_{min}. This is of interest because the difference in error obtained by λ_{min} and λ_{1SE} is unlikely to be significant but the number of features returned by the classifier identified by the latter is likely to be lower.

We report experiments with lasso and elastic net regularizers. Each drives some model coefficients to zero and thus performs feature selection. As different data samples give rise to different feature sets being selected, we perform a bootstrap analysis, measuring how likely features and feature subsets are to be selected as the data are perturbed [16]. Bootstrap was used similarly, for example, by Park and Hastie [17] to investigate gene interactions albeit using a different feature selection method.

Analysis of semantic features. Each bootstrap analysis comprised 500 bootstrap trials. The feature vector used comprised VAMPIRE measurements, sex and age at scan. The number of patients associating with dementia was less than 14%, resulting in a significant imbalance of the *dementia* and *no dementia* classes. Therefore, sampling with replacement was carried out to extract 100 dementia and 100 non-dementia samples from the cohort in each trial. Regularized logistic regression was implemented on the 200 samples. The regularization path was computed. Model selection used 10-fold cross-validation (CV) to choose λ. For a λ of interest (e.g. λ_{min}), the corresponding feature coefficients ($\boldsymbol{\beta}$) and classification error obtained were reported.

We then computed the proportion of times within the 500 bootstraps that each feature had a non-zero weight, thereby providing a measure of how likely the features were to be selected. To detect interesting or frequently occurring patterns amongst groups of features across the bootstraps, we computed the number of times each possible feature-selection permutation took place (e.g., in, say, 200 of the 500 bootstraps, tortuosity and width of arteries were chosen by the model as important, while the rest of retinal features were assigned zero weights). The average classification error was

calculated across the 500 bootstraps, together with the corresponding 95% CIs. This process was carried out under four different regularization settings: (a) Lasso and λ_{min}; (b) Elastic-net and λ_{min}; (c) Lasso and λ_{1SE}; (d) Elastic-net and λ_{1SE}.

Analysis of non-semantic features. A bootstrap analysis with textural features was run using the aforementioned methodology. Here, we aimed to explore whether any texture information would associate with dementia. To this end, we experimented with two settings: (a) Using the full textural pyramids; and (b) Using the local quadrant features only. We did not explore different regularization settings but ran the framework using lasso and reported results obtained with λ_{min}.

3   Results

Clinically semantic features. Fig. 1 shows an example of a bootstrap trial. As can be seen in Table 1, the most powerful feature in predicting dementia association was the patient age at scan, which was assigned a non-zero weight in almost all bootstrap runs. This was expected given the elderly cohort, and that the risk of developing dementia increases with age. Retinal features were to a great extent complementary to age, given that at least one retinal feature was chosen 81-82% of the time, using λ_{min}. This percentage was lower when λ_{1SE} was used, as expected. OD radius and CRAE were highly ranked after age. In terms of classification performance (Table 2), the mean classification error (MCE) ranged between 35% and 38%, suggesting a reasonable set of discriminatory patterns available in the models used. CIs were reasonably narrow, providing a good level of certainty in the classification results obtained.

We computed the frequency of unique feature patterns across the bootstraps, shown in Figure 2 (lasso and λ_{min}). The peak of the plot (94/500 times, 18.8%) is age only. The graph tail covers the large number of permutations possible. In 35 trials (7%), CRAE alone was selected in addition to age at scan. All retinal measurements were selected together 26 times (5.2%). All tortuosity measures tended to be selected or discarded together in the highly ranked counts. The above observations were in line with what emerged in all four regularization settings used (complete permutation reports not included for compactness).

Using a classifier trained with semantic measurements only (without age) returned a high MCE of 46% (CI: 38.5%-55%, lasso and λ_{min}). We then tested the MCE of age-only classifiers to predict the outcome with unregularized logistic regression Using the *CARET* [22] package in R (*GLM*-based implementation [7]). The MCE was 36.8% across the trials (95% CIs: 31.5%-43%). Using CRAE and age together (the second most prevalent feature sebset in the feature selection experiment) gave an MCE of 35.3% (CIs: 29.5%-41%).

Clinically non-semantic features. (a) Using the textural feature pyramid without age, logistic regression models were able to classify dementia association with a MCE of 37% (CI: 30%-43%). The following quadrant features appeared in the majority of trials (at least 300): *low grey-level run emphasis (top and bottom right quadrants), grey-level non-uniformity (bottom left) and inverse difference moment normalized (45° direction, bottom right, top right and top left; 135° direction, bottom left).* Global

features also appeared in at least 300 trials: *inverse difference moment normalized (0° and 45° directions)* and *grey-level non-uniformity.*

(b) Using a feature vector that only comprised quadrant features achieved a comparable MCE of 38% (CI 32%-44%). The same features that appeared in (a) were also selected here in at least 300 trials, with the exception of grey-level non-uniformity. Additionally *short run emphasis (bottom left)* featured as important.
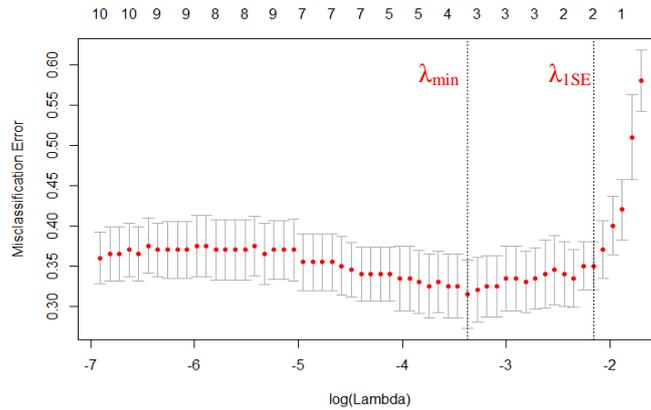

Fig 1. Example of an individual bootstrap trial (Lasso). Varying λ affects the MCE (with 10-fold CV). Numbers above figure are the numbers of features retained in the regularized model. Interval bars are standard deviations.

4 Discussion and Conclusions

Findings. The role of fundus camera imaging as a means of identifying changes associated to dementia remains disputable [3]. This pilot used a bootstrap analysis based on regularized logistic regression to investigate the association of retinal vascular features computed with the VAMPIRE 3.0 software in fundus camera images (semantic features) as well as of textural features (nonsemantic) with a dementia outcome in an elderly population of diabetics. In the specific cohort, age was the strongest predictor; retinal features were selected in up to ~81% of the trials (both Lasso-λ_{min} and Elastic-net-λ_{min}), suggesting that they do *complement* age when selected together in a classifier. Arterial caliber and OD size were the retinal measurements selected most often, indicating highest discriminative power within the semantic set, when semantic retinal features only (not age) were considered. Textural features were able to match the performance of age and are potentially a rich source of information for dementia outcome classification. Specific features appeared in over 300 of the 500 trials of each experiment, hence further analysis of these textural features will drive our future efforts.

Study limitations and future work. *Analysis.* Non-linear models might be able to reveal further associations compared to linear logistic regression. Classification results were obtained with 10-fold cross-validation; a held-out test set could be adopted, although the focus of this work was investigating discriminative sets of retinal features, not maximizing classification performance. Using one eye only assumes left-right

symmetric measurements, an assumption *sub judice* in the recent literature [21]. Given promising initial results using textural features, it will be interesting to also try feature representations obtained from deep learning using large retinal datasets. *Cohort.* The average age at scan of dementia-class subjects was 6.4 years higher than that of non-dementia subjects, making age the most powerful predictor in this cohort. The significant time lapses between scans and dementia-defining events make it necessary to confirm results with a tighter inclusion criterion for the dementia class. We shall also include further parameters beyond age and sex, used in this pilot. Finally, strongly longitudinal data are needed to evaluate the strengths of retinal measurements for predicting risk of dementia.
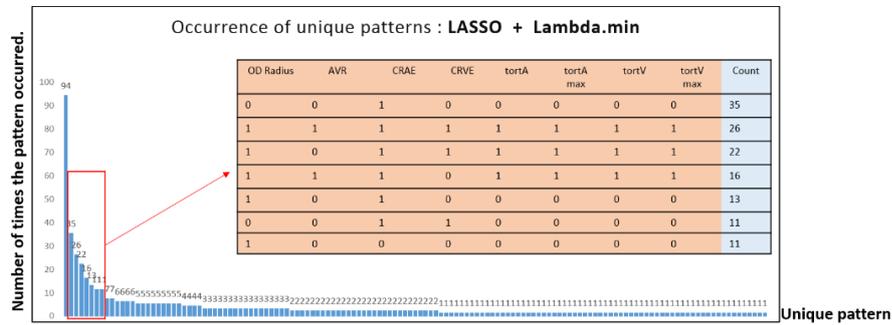
Fig. 2. Unique feature subset patterns that were selected, together with the counts of the highly ranked patterns. A '1' indicates that the feature was selected.

Table 1. Relative feature importance as the number of times a feature was selected (non-zero weight) in the 500 bootstraps. *% ret* is the percentage of times that at least one retinal feature was selected. (a) Lasso and λ_{min}, (b) Elastic-net and λ_{min}, (c) Lasso and λ_{1SE}, (d) Elastic-net and λ_{1SE}. The three most frequently selected features in each analysis are displayed in bold.

Variation	OD Radius	AVR	CRAE	CRVE	tortA	tortA max	tortV	tortV max	Age at scan	Sex	% ret
(a)	**243**	156	**392**	151	192	172	180	193	**500**	188	81%
(b)	**263**	195	**310**	168	189	217	215	211	**500**	211	82%
(c)	**61**	34	**113**	36	37	33	40	30	**499**	44	36%
(d)	66	50	**138**	61	42	46	**67**	47	**500**	58	43%

Table 2. Classification error output from the bootstrap analysis. The feature vector comprised VAMPIRE semantic measurements, sex and age at scan.

Variation	Average classification error across bootstraps	95% CI of classification error
Lasso and λ_{min}	0.35	0.29-0.40
Elastic-net and λ_{min}	0.35	0.30-0.41
Lasso and λ_{1SE}	0.37	0.32-0.43
Elastic-net and λ_{1SE}	0.38	0.32-0.44

References

1. https://www.alzheimers.org.uk/info/20025/policy_and_influencing/251/dementia_uk.
2. Patton, N. et al.: Retinal Vascular Image Analysis as a Potential Screening Tool for Cerebrovascular Disease: a Rationale based on Homology between Cerebral and Retinal Microvasculatures. Journal of Anatomy (206), 319-348 (2005).
3. McGrory, S. et al.: The Application of Retinal Fundus Camera Imaging in Dementia: a Systematic Review. Alzheimer's and Dementia (6), 91-107 (2016).
4. Williams, M. A. et al: Retinal Microvascular Network Attenuation in Alzheimer's disease. Alzheimer's and Dementia 1(2), 229-235 (2015).
5. Frost, S. et al.: Retinal Vascular Biomarkers for Early Detection and Monitoring of Alzheimer's Disease, Translational Psychiatry 3(2):e233 (2013).
6. https://uk.mathworks.com/matlabcentral/fileexchange/22187-glcm-texture-features
7. https://stat.ethz.ch/R-manual/R-devel/library/stats/html/glm.html
8. Friedman, J. et al.: Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software, 33 (1), 1-22 (2010).
9. https://uk.mathworks.com/matlabcentral/fileexchange/52640-gray-level-run-length-image-statistics
10. Thomson, K.L. et al.: A systematic review and meta-analysis of retinal nerve fiber layer change in dementia, using optical coherence tomography. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 1(2):136-43 (2015).
11. Heringa, S.M. et al.: Associations between retinal microvascular changes and dementia, cognitive functioning, and brain imaging abnormalities: a systematic review, J Cerebr Blood F Met. 33(7):983-95 (2013).
12. http://medicine.dundee.ac.uk/godarts
13. MacGillivray, T.J. et al.: Suitability of UK Biobank retinal images for automatic analysis of morphometric properties of the vasculature: a VAMPIRE study. PlosONE, (2015).
14. MacGillivray, T.J. et al.: VAMPIRE: Vessel Assessment and Measurement Platform for Images of the Retina. Human Eye Imaging and Modeling. CRC Press, (2012).
15. Trucco, E. et al.: Novel VAMPIRE algorithms for quantitative analysis of the retinal vasculature. In: Proceedings from the 4th ISSNIP/BRC.1–4 (2013).
16. Efron, B. & Tibshirani, R.: An Introduction to the Bootstrap, Chapman & Hall/CRC, Boca Raton (1993).
17. Park, M. Y & Hasite, T.: Penalized Logistic Regression for Detecting Gene Interactions, Biostatistics, 9(1):30-50, (2008).
18. Haralick, R. M. et al.: Textural Features for Image Classification, IEEE Transactions on Systems, Man and Cybernetics, SMC-3(6):610-621 (1973).
19. Galloway, M. M.: Texture Analysis using Gray Level Run Lengths, Computer Graphics and Image Processing, 4(2):172-179 (1975).
20. Jie D et al.: Retinal vascular caliber and the development of hypertension: a meta-analysis of individual participant data. Journ. Hypertension, 32(2): 207–215 (2014).
21. Cameron J. R. et al.: Lateral thinking: inter-ocular symmetry and asymmetry in neurovascular patterning, in health and disease. Progress in Retinal Eye Research, Apr 2017, epub ahead of print, doi: 10.1016/j.preteyeres.2017.04.003.
22. https://cran.r-project.org/web/packages/caret/caret.pdf