International Journal of Global Engineering (IJGE) E- ISSN: 2456-3099

VOL 3 ISSUE 2 (2018) PAGES 27 - 34

Received: 12/10/2018. Published: 26/10/2018

AUTOMATIC DETECTION OF MACULAR DEGENERATION.

R.Nishanthini, S.Muthuilakshmi, B.Muthamil
Department of Biomedical Engineering
Dhanalakshmi Srinivasan Institute of Technology, Samayapuram, Tamil Nadu, India

Abstract
The rising prevalence of age-related eye diseases, particularly age-related macular degeneration, places an ever-increasing burden on health care providers. As new treatments emerge, it is necessary to develop methods for reliably assessing patients' disease status and stratifying risk of progression. The presence of drusen in the retina represents a key early feature in which size, number, and morphology are thought to correlate significantly with the risk of progression to sight-threatening age-related macular degeneration. Manual labeling of drusen on color fundus photographs by a human is labor intensive and is where automatic computerized detection would appreciably aid patient care. We review and evaluate current artificial intelligence methods and developments for the automated detection of drusen in the context of age-related macular degeneration.
Keywords : Age-related, macular degeneration, age-related disorder, artificial intelligence, machine learning, deep learning

1. Introduction

With longer life expectancy, age-related disorders are increasing the burden placed on health care providers. In particular, age-related macular degeneration (ARMD) is one of the major causes of vision loss in the elderly.[1] ARMD currently affects 6 million people in the UK alone and was estimated to have cost the country's economy £155 million in 2011.49 By 2040, the number of people affected globally by the disease is projected to be 288 million.[2]

The earliest phase of ARMD is typically observed as the presence of (asymptomatic) macular drusen, often incidentally found on examination or fundus imaging. Drusen are small deposits of predominantly lipid, acellular debris that accumulate between the retinal pigment epithelium and Bruch's membrane. Although the presence of small drusen is not itself diagnostic of ARMD, as drusen frequently occur in normal aging, increasing number and size of drusen increase the risk of progression to visually symptomatic ARMD. Later signs of ARMD, such as pigmentary changes of the retinal pigment epithelium that occur before the development of geographic atrophy (so-called dry ARMD) and exudative abnormalities (so-called wet ARMD) enable more established gradings [3] and classification of ARMD.[4]

Drusen appear as clusters of white or yellow spots in color fundus photographs and broadly exist as two main types, hard and soft. Hard drusen are round, small, discrete lesions with defined edges, whereas soft drusen are less defined and often confluent. Drusen are rarely homogenous in their composition. Because of their yellow color and brightness on color fundus photographs, drusen are distinguishable by the human eye, but computer algorithms to automatically detect

International Journal of Global Engineering (IJGE) E- ISSN: 2456-3099
VOL 3 ISSUE 2 (2018) PAGES 27 - 34
Received: 12/10/2018. Published: 26/10/2018

them need to be robust to the presence of other similarly brightly appearing pathology such as hard exudates. Indistinct borders for drusen appearing in color fundus photographs are challenging for conventional image-processing techniques such as edge detection and morphological filtering and have been discussed in detail in an earlier review.[5] To the best of our knowledge, no reviews cover recent developments, involving the application of artificial intelligence (AI) and deep learning (DL) techniques.

AI is a long-standing field of computer science that aims to simulate human intelligence by perceiving its environment and taking appropriate action to achieve a set of goals, one of which is decision-making. Machine learning (ML) is an approach to AI, partially inspired by how humans learn. [6] Learning is achieved through examples. If a child is presented with a new object, they will use features such as color, shape, and texture so that when they observe the object again they will use what they have learned to identify or categorize it as something they have previously seen. Similarly, many ML classification algorithms use features from training examples to discover or confirm patterns that categorize subsets. When new, unseen data are presented, the algorithm can classify which category they belong to (Fig. 1). These features can be learned by either training from previous examples (i.e., supervised learning) or discovered by the algorithm (i.e., unsupervised learning).

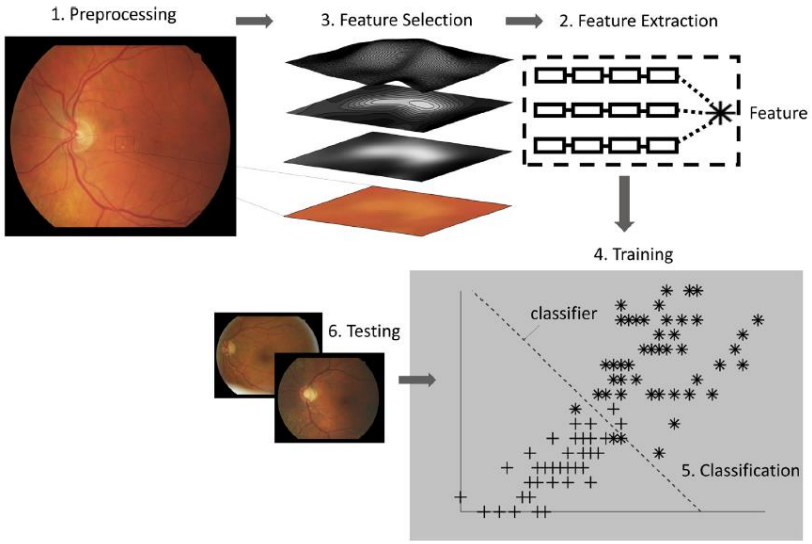

Fig. 1 Illustration of supervised machine learning pipeline. 1) Image preprocessing is performed to reduce noise and enhance image features. 2) Features such as measures of entropy, energy, color and texture of image intensities, and spatial or geometric properties are extracted. 3) Features are grouped as numerical vectors (forming the image representation) and often undergo a selection process to decide which features best represent the image. 4) Training phase builds a model that tries to separate the data into the target, distinct classes. 5) The classifier the mathematical function that implements classification and defines the classes. 6) Testing is performed by classifying unseen data belonging to know classes. Survey of ophthalmology

VOL 3 ISSUE 2 (2018) PAGES 27 - 34
Received: 12/10/2018. Published: 26/10/2018

DL is a subset of ML that is gaining prominence for medical imaging [7] and ophthalmology [8] because of increasing re-ports of high performance for clinical classification and decision-making. DL is based on neural networks, a class of algorithms inspired by the human brain. In a neural network, the neurons are organized in layers and implement simple operations on the input data or from the output of previous layers. In a deep neural network, the number of layers is much higher than that in conventional neural networks (indicatively10 or more as opposed to 2-3). The connections between the layers are assigned values, called weights, representing connection strengths. Learning the weights is the objective of the training process. Training and testing a deep neural network require large amounts of labeled data (i.e., known classes).In this review, we report and evaluate current AI strategies and developments for the automated detection of drusen in the context of ARMD (Fig. 2). Although some recent work has begun to explore the potential for automated drusen detection by optical coherence tomography, with varied methods and mixed results, [9] the focus of this review is on color fundus imaging of the retina.

2. Methods

2.1. Inclusion and exclusion criteria

We aimed to include all published studies applying AI to automatic drusen detection in color fundus photographs. Inclusion criteria were (1) original study, (2) those written in English, and (3) those that had validation by performance against at least one manual grader. The following studies wereexcluded: (1) reviews; (2) nonhuman research; (3) non-English (4) studies that involved methods other than color fundus photography (e.g., optical coherence tomography); (5) studies that did not feature robust validation, as outlined in the following paragraph.

Validation is the process of showing quantitatively that an algorithm performs correctly through comparison of its output to a reference standard, for example, manual grading of images by experts.[10]

Any article that did not include validation was excluded. The performance of an algorithm is typically measured using criteria such as accuracy, sensitivity, specificity, and area under receiver operating characteristic. Another important aspect is the size of the data set: the image set, algorithm tested, must be sufficiently large to be representative of the target population and to be suitable for the number of neural network parameters to be trained. AI methods are not immune to small sample size effects that can contaminate the evaluation of a proposed system. For instance, color fundus photographs can differ in appearance between patients, and disease manifestations are also of a varying nature. Considering this, articles that mentioned validation of less than 50 images were excluded.

2.2. Data extraction

For all identified studies, an independent reviewer (E.P.) screened the titles and abstracts. Irrelevant and duplicate articles were removed, and the remaining articles were assessed for agreement with the inclusion and exclusion criteria by full-text review. Data extracted from studies at this stage included title, year of publication, authors, study aim, study type, number of images (training and test), diagnostic criteria, participant selection criteria, method of fundus imaging, algorithm, performance metric(s) results, and conclusions. The most recent articles

International Journal of Global Engineering (IJGE) E- ISSN: 2456-3099
VOL 3 ISSUE 2 (2018) PAGES 27 - 34
Received: 12/10/2018. Published: 26/10/2018

were hand searched following the same strategy, filtered for the current year (i.e., 2018), and subjected to the same inclusion criteria. A similar strategy was followed for articles cited within the bibliographies of the results.

3. Results

A total of 2236 articles were identified in the initial search performed in 2017. After filtering for ARMD, 1318 articles were excluded, such as those featuring diabetic retinopathy (n = 42) and glaucoma (n = 42). From the remaining 918 articles, 834 were excluded because they did not use color fundus photographs (n = 18), did not use imaging (n = 770), or were not reviews (n = 34). Seventy-three articles did not meet the selection criteria, such as articles not reporting performance (n = 9) or featuring software optimization (n = 3), hardware reports (n = 2), or fewer than 50 images for validation (n = 12). At the end, 8 articles met all inclusion criteria. One additional article was included after searching bibliographies, and 5 articles were found by hand searching for this current year (2018). The resulting 14 articles were considered in this review. They all used ML and DL techniques for drusen detection in color fundus photographs.

3.1. Study designs and populations

The 14 studies involve 4 publicly available data sets (i.e., automatic retinal image analysis, STructured Analysis of the REtina,[11] Age Related Eye Disease Study [AREDS], and RetinaGallery12), private data sets and other sourced from a telemedicine platform and a cohort from an independent study. Some studies contained overlapping report analyses on the same data sets, but used different methods. Four articles aimed to achieve disease or no-disease classification. Six articles aimed to classify ARMD severities according to AREDS2 or in-house grading criteria (Cologne Image Reading Center and Laboratory [CIRCLE]). Two articles aimed to classify dry ARMD vs. normal images and 1 wet ARMD vs. dry ARMD or normal.

3.2. Preprocessing and feature extraction

In automatic detection, preprocessing is a commonly used step to enhance an image to better facilitate the extraction of features relating to objects of interest. The human eye distinguishes “features” of disease in an image (such as geographic atrophy and drusen), but AI algorithms need to extract “features” measured from the pixels pertaining to an object (i.e., drusen). In addition, a color fundus photograph typically contains a black border that needs either to be avoided or eliminated because these pixels will not be of any relevance. Retinal landmarks (e.g., the optical nerve boundary, blood vessels, and macula) may obstruct features of small objects, so their removal may further improve automatic detection by reducing sources of false targets for drusen detection. A color fundus photograph might also contain artifacts (e.g., from dust particles on the lens) and display areas of uneven illumination that preprocessing can eliminate. The type of preprocessing used in the studies included depended on the particular features used.

3.3. Classification

Classification uses the features selected to identify the model that best separates the data into the desired classes. A collection of images is typically separated into training and testing sets, of which the former is used to develop the model and the latter is used to test it. In the context of ARMD, this would test the model's ability to classify disease/no-disease or dry/wet ARMD. To evaluate the accuracy of the classifier, cross-validation is often performed.[12] The algorithm

performance is commonly reported in terms of statistics of measures, comparing the classifiers' decisions against those of one or more human experts.

3.3.1. Disease/no disease

Hijazi and coworkers[13] proposed a case-based reasoning system to develop an automated screening tool to classify 144 color fundus photographs into ARMD or normal categories. Case-based reasoning is a problem-solving technique based on the observation of how humans use previous examples or information to solve new, but similar, problems. If a case-based reasoning system is given a new case, it will use the previous most similar cases in its case base to solve the problem. Each image histogram was conceptualized to a set of curves, called a time series, and used to generate a 2-step case-based reasoning classification. The first case consisted of enhanced green channel images, with the blood vessel pixels replaced with null values. The second case contained the same but with the further process of removing the optic disc. Histograms and their time series of a collection of unseen graded images were passed to the first case for comparison to the training images. An algorithm called dynamic time warping was used to measure the similarity between the histograms and time series of the testing and training images. If the unseen image was below a certain similarity measure, it was then passed to the second case for reassessment. The output is whether the input image is similar to either the learned time series of an ARMD image or a healthy image in the case base. A specificity of 82% was reported for the effectiveness of the classifier in identifying ARMD images, 65% specificity for the classifier identifying normal images, and 75% accuracy in classifying images as ARMD or normal. This two-pass approach offered a system whereby isolation and segmentation of drusen was not required; however, removal of vessels and the optic disc was needed to improve the accuracy.

3.3.2. ARMD severity

Phan and coworkers[14] attempted to classify ARMD severity according to their AREDS categories5 using visual words, also known as “bag of words.” The most salient features in the image were detected and their frequencies counted and binned in to a histogram. This forms a so-called vocabulary that can be used for automated detection of the same words in an unseen image. The authors used Speeded Up Robust Features to build the vocabulary from different color spaces (red, green, blue and a color space describing lightness, green-red, and blue-yellow, called L*a*b) of 279 images, including poor-quality images, to build the vocabulary. SVM and random forest classifiers were tested with and without feature-selection steps. They report the best performance for ARMD screening with SVM classifier.

3.3.3. Wet/dry/no disease

Using entropy measures as features from wavelet coefficients and from green channel CLACHE-enhanced images, detection of dry ARMD using SVM, Naïve Bayes, probabilistic neural networks, k-nearest neighbors, and decision trees was proposed by Mookiah and coworkers.[15] This system was trained and tested separately on three data sets (automatic retinal image analysis, STARE, and a private data set). The best performance was reported for an SVM classifier where Gabor, local pixel intensity changes, and entropy features ranked best. The highest performances were observed in automatic retinal image analysis and STARE, with an accuracy of correctly classifying between dry ARMD and normal of 95.7% and 95%, respectively.43 Statistical moments, energy, entropy, and Gini index features extracted from discrete wavelet transform (a well-known image denoising technique) also presented the best

accuracy for SVM (93.70%).41 This system did not require prior segmentation of retinal landmarks and drusen, and the use of multiple classifiers provided a degree of discrimination ability of the extracted features.

4. Result and Conclusion

ML is highlighted as the predominant technique for ARMD detection and classification, with most recent articles reporting DL techniques. The primary aim of drusen-related automated image analysis is to support decision-making in the clinic. Rather than detecting individual drusen, image-level classification was more common with the aim of computerizing ARMD screening and grading systems. Only a single article reported discrete drusen measurement and quantification. Manually outlining individual drusen to provide ground truth for algorithm training is very labor intensive and motivates the shortage of ML approaches to individual drusen segmentation. AREDS categories,5 class 1 and class 2 ARMD, are the most difficult to separate because grading relies on drusen counts and measurements that cannot be obtained automatically without the reference data. ML is particularly susceptible to this paradox because they are driven by examples that are assumed to be representative of the population. A newly obtained image may not be similar to any of the examples used to train the model, and therefore, it may fail to classify it. This effect of data variability was also observed when the model was evaluated on an independent data set containing color fundus photographs with retinopathies not present in the training set and removal improved performance. This raises questions as to how ML would generalize to the clinic.

The results of our search identified a number of articles reporting algorithms for detection of DR and glaucoma where drusen can also be present. Fundus imaging has also been used to derive biomarkers for systemic conditions, such as hypertension and diabetes. Recently, there are an increased number of reports linking ARMD to Alzheimer disease (AD). AD is diagnosed using medical history, psychiatric examination, brain imaging, and biomarkers in cerebrospinal fluid. Definitive classification requires neuropath logical changes as seen on postmortem examination. Characteristic retinal changes have previously been identified in AD, such as a sparser retinal vascular network (inferring altered cerebral vasculature) and thinning of the retinal nerve fiber layer (a marker of axonal loss). A key component of AD-related deposits in the brain, amyloid β, is also found in drusen. Amyloid β is an aggregate-prone peptide family that aggressively targets neurons, and there are an increasing number of reports of amyloid plaques in the retina in patients with AD. As the retina is anatomically, embryologically, and physiologically linked to the central nervous system, it is perhaps not surprising that these depositions may have implications to neurodegenerative disease of the brain. Indeed, the progression of drusen formation in the peripheral retina has been found to be more prevalent in patients with AD than in the age-matched control. These findings were in a small cohort but suggest a promising biomarker for disease-related plaque formation in the brain.

When ARMD progresses asymmetrically, patients risk remaining asymptomatic due to maintaining good visual acuity in their healthy eye. The resulting delay in presentation and treatment impacts visual prognosis.

For automated drusen assessment to be applied in the clinic, it must go beyond cross-sectional phenotyping and instead relate to real patient visual outcomes. Longitudinal studies will be

required to determine if automated image grading, based on drusen detection, can accurately predict disease progression.

Future algorithms involving drusen detection should aim to provide useful quantification to aid screening for ARMD. A screening program should stratify patients according to optimal follow-up pathway. For automated drusen detection to contribute to the cost-effectiveness of a screening program for ARMD, it must separate individuals with drusen associated with normal aging from patients whose drusen load progresses and stratify patients with mild ARMD into those at low risk and at high risk of progression to severe ARMD. This would enable the ophthalmologist to select relevant patients for regular follow-up, thus improving the efficiency of patient care.

References

[1] N. Joachim, P. Mitchell, G. Burlutsky, A. Kifley, J.J. Wang The incidence and progression of age-related macular degeneration over 15 years: the Blue Mountains eye study 1229 (12) (2015), pp. 2482-2489doi:10.1088/0031-9155/58/13/R97.
[2] T. Yoshida, K. Ohno-Matsui, S.J. Ichinose The potential role of amyloid beta in the pathogenesis of age-related macular degeneration Clin Invest, 115 (10) (2005), pp. 2763-2800
[3] A.C. Bird, N.M. Bressler, S.B. Bressler, I.H. Chisholm An international classification and grading system for age-related maculopathy and age-related macular degeneration Surv Ophthalmol, 49 (5) (1995), pp. 367-374
[4] AREDS Group A randomized, placebo-controlled, clinical trial of high-dose supplementation and vitamins C and E, beta carotene, and zinc for age-related macular degeneration and vision loss: AREDS report no 8 Arch Ophthalmol, 119 (2001), pp. 1417-1436
[5] C. Duanggate, B. Uyyanovara A review of automatic detection and segmentation from retinal images The 3rd Int Symp Biomed Eng (isbme) (2008), pp. 222-225
[6] Y. LeCun, Y. Bengio, G. Hinton Deep learning Nature, 521 (2015), pp. 436-444
[7] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A.S. Setio A survey on deep learning in medical image analysis Med Image Anal, 42 (2017), pp. 60-88
[8] J. De Fauw, J.R. Ledsam, B. Romera-Paredes, S. Nikolov Clinically applicable deep learning for diagnosis and referral in retinal disease Nat Med, 24 (9) (2018), pp. 1342-1350
[9] R. Zhao, A. Camino, J. Wang, A.M. Hagag Automated detection in dry age-related macular degeneration by multiple depth, enface optical coherence tomography Biomed Opt Express, 8 (11) (2017), pp. 5049-5064
[10] E. Trucco, A. Ruggeri, T. Karnowski, L. Giancardo Validating retinal fundus image analysis algorithms: Issues and a Proposal IOVS, 54 (5) (2013), pp. 3546-3559
[11] Hoover Structured Analysis of the Retina Available at: http://cecas.clemson.edu/~ahoover/stare/, Accessed 29th Jun 2018
[12] M. Stone Cross-validity choice and assessment of statistical predictions J R Stat Soc Ser B (methodological), 36 (2) (1974), pp. 111-147

VOL 3 ISSUE 2 (2018) PAGES 27 - 34

Received: 12/10/2018. Published: 26/10/2018

[13] M. Hijazi, F. Coenen, Y. Zheng Retinal image classification using histogram based approach The 2010 International Joint Conference on Neural Networks (IJCNN). Barcelona (2010), pp. 1-7

[14] T. Phan, L. Seoud, H. Chakor, F. Cheriet Automatic screening and grading of age-related macular degeneration from texture analysis of fundus images J Ophthalmol, 8 (2016), pp. 1-11

[15] M. Mookiah, U. Acharya, J. Koh, C.K. Chua Decision support system for age-related macular degeneration using discrete wavelet transform Med Biol Eng Comput, 52 (2014), pp. 781-796