Artificial Intelligence in Medicine 60 (2014) 103–112

Contents lists available at ScienceDirect

Artificial Intelligence in Medicine

journal homepage: www.elsevier.com/locate/aiim

Improving predictive models of glaucoma severity by incorporating quality indicators

Lucia Sacchi[a,*], Allan Tucker[b], Steve Counsell[b], David Garway-Heath[c], Stephen Swift[b]

[a] Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Via Ferrata 1, 27100 Pavia, Italy
[b] Department of Information Systems and Computing, Brunel University, Uxbridge, Middlesex UB8 3PH, UK
[c] NIHR Biomedical Research Centre at Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, 162 City Road, London EC1V 2PD, UK

ARTICLE INFO

Article history:
Received 25 February 2013
Received in revised form 2 December 2013
Accepted 4 December 2013

Keywords:
Predictive modelling
Reliability indicators
Visual field testing
Glaucoma severity prediction

ABSTRACT

Objective: In this paper we present an evaluation of the role of reliability indicators in glaucoma severity prediction. In particular, we investigate whether it is possible to extract useful information from tests that would be normally discarded because they are considered unreliable.
Methods: We set up a predictive modelling framework to predict glaucoma severity from visual field (VF) tests sensitivities in different reliability scenarios. Three quality indicators were considered in this study: false positives rate, false negatives rate and fixation losses. Glaucoma severity was evaluated by considering a 3-levels version of the Advanced Glaucoma Intervention Study scoring metric. A bootstrapping and class balancing technique was designed to overcome problems related to small sample size and unbalanced classes. As a classification model we selected Naïve Bayes. We also evaluated Bayesian networks to understand the relationships between the different anatomical sectors on the VF map.
Results: The methods were tested on a data set of 28,778 VF tests collected at Moorfields Eye Hospital between 1986 and 2010. Applying Friedman test followed by the post hoc Tukey's honestly significant difference test, we observed that the classifiers trained on any kind of test, regardless of its reliability, showed comparable performance with respect to the classifier trained only considering totally reliable tests (p-value > 0.01). Moreover, we showed that different quality indicators gave different effects on prediction results. Training classifiers using tests that exceeded the fixation losses threshold did not have a deteriorating impact on classification results (p-value > 0.01). On the contrary, using only tests that fail to comply with the constraint on false negatives significantly decreased the accuracy of the results (p-value < 0.01). Meaningful patterns related to glaucoma evolution were also extracted.
Conclusions: Results showed that classification modelling is not negatively affected by the inclusion of less reliable tests in the training process. This means that less reliable tests do not subtract useful information from a model trained using only completely reliable data. Future work will be devoted to exploring new quantitative thresholds to ensure high quality testing and low re-test rates. This could assist doctors in tuning patient follow-up and therapeutic plans, possibly slowing down disease progression.
© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Quality control during medical testing is an important issue in healthcare. Introducing unreliable, low-quality and noisy information into the care process can ultimately lead to incorrect diagnoses and therapy plans. High quality test delivery is important, but it should also take into account patient comfort and compliance, balancing the benefits of high quality measurements with the costs and harm potentially caused by over-repeating a test [1–3]. Checking reliability in a test can increase examination duration causing patient fatigue and affecting test outcome. Finding a trade-off between high quality care and patients' comfort during the testing procedure is still an open question in the medical community.

One of the most frequently applied techniques for ensuring high quality testing is to discard those tests that are considered unreliable. That said, it is often difficult to define the criteria for stating which tests are really trustworthy. These criteria are usually defined by the manufacturers of the testing devices and are conservative and general in nature. In this paper, we investigate whether it is possible to extract useful information from tests that may be discarded because they are considered unreliable. As a matter of fact, during clinical practice a qualitative evaluation of unreliable tests is often carried out to identify patterns that might

* Corresponding author. Tel.: +39 0382 985981.
E-mail address: lucia.sacchi@unipv.it (L. Sacchi).

0933-3657/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.artmed.2013.12.002

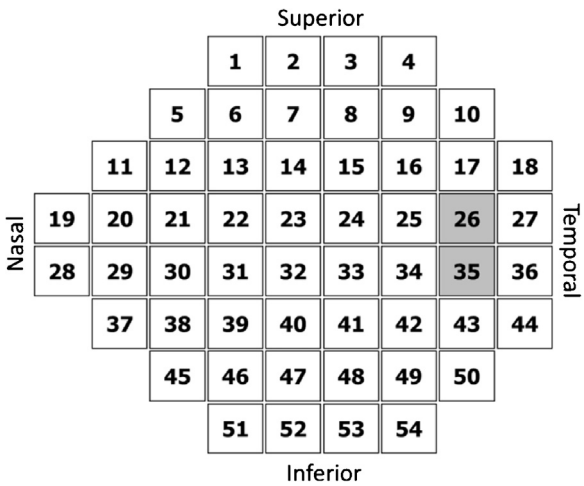

Fig. 1. Map of the visual field test locations for the right eye under the Humphrey Field Analyzer™ 24-2 program (the grey squares indicate the location of the blind spot).

suggest some useful information to the physician. This process, however, is not translated into a standard quantitative procedure and it is strongly dependent on the individual medical expert. The main objective of the present study is to evaluate the possibility of reducing the number of test repetitions while keeping the quality of the results and the patients' conditions at a high level. To this end, we consider an application of visual field testing (VF), a primary technique for detecting functional abnormalities related to glaucoma.

During a VF test, a subject is shown spots of light of varying brightness at several locations on the VF map. The number of locations on the map depends on the specific test program adopted in the clinical practice. VF tests are currently delivered through computerised automated perimetry. In this paper, we consider tests taken using second generation Humphrey Field Analyzers™ under the 24-2 program (HFA, Carl Zeiss Meditec, Inc., Dublin, CA). This protocol tests a total of 54 locations. At each location, the intensity of the stimulus that the subject is able to just see can be directly related to the sensitivity of retina to light. Each test location can be identified through use of a fovea-centred coordinates system or, more conveniently, a unique identification number, as shown for the right eye in Fig. 1.

VF testing is a good example of a testing paradigm where the reliability issue plays a crucial role. The process of defining retinal sensitivity needs to be performed using a small number of stimuli with the ultimate goal of minimising test duration. Algorithms devoted to the evaluation of patient reliability are included in the majority of testing programs. Besides sensitivity values, a VF test also collects global indices aimed at evaluating a patient's condition and the reliability of the test. Among these indicators, those intended to monitor reliability are the false positives rate (FP), the false negatives rate (FN) and fixation losses count (FL). In full-threshold automated perimetry, which is the technique we will consider in this study, an estimate of false positive and false negative rates is usually derived from a catch-trial strategy [4]. According to this technique, responses given during a pause in stimulus presentation are recorded as FP. On the other hand, when no response is given to a stimulus substantially brighter than the established threshold at a VF location, an FN response is recorded. FLs are determined by the count of positives responses to random stimuli at the location for the physiological blind spot [5–7].

Presently, tests results are not automatically adjusted for reliability and it is still up to the clinician to interpret an unreliable test. The challenge of determining which tests should be discarded is still open. Manufacturers of VF analysers have proposed certain thresholds on the reliability indices which, if exceeded, can be associated with a significant alteration of the visual field test results due to an unreliable patient. Once these thresholds are exceeded, a clinician is advised to discard the test and repeat it.

VF test reliability assessment has been addressed by many studies in the literature. In their early work, Katz and Sommer [8] described how manufacturer-defined reliability parameters are distributed into a clinic-based population. The effect of missed catch trials in normal subjects was studied in the work by Cascairo et al., where a population of normal controls was analysed when asked to purposely miss an increasing percentage of questions [9]. Bengtsson [5] considered reliability of VF test results expressed as threshold reproducibility, showing that it can be reasonably predicted by the amount of field loss with only a small contribution from reliability indices. Work by the same group [10] analysed the meaning of FN responses in a VF test, concluding that they were more related to the damage of the eye than to the patient's reliability characteristics. FP responses were addressed in both [6] and in [11]. In the first of these studies, the authors evaluated the accuracy of the catch-trial method for estimating the FP rate, proposing that tests showing less than 20% of FP should not be discarded and should be considered reliable instead. In the second study, two different test programs, full-thresholds and the Swedish interactive threshold algorithms (SITA standard) [12] were compared on their capability for estimating FP error frequency, with results showing that SITA suffered from FP underestimation on normal patients.

We address the problem of understanding the impact that test reliability indices have on the task of determining future disease severity. Many studies have investigated the appropriateness of reliability cut-offs for determining whether those thresholds reflect sensitivity alterations due to patient unreliability. To the best of our knowledge, the issue of how reliability indicators affect the results of the problem of predicting future disease severity has not yet been tackled.

We particularly focus on the following questions: is the information contained in unreliable tests totally impractical for determining the future evolution of the disease? How do different reliability indicators impact on prediction results? Are these contributions significantly different?

To answer these questions we setup a traditional predictive modelling framework [13,14], where we try to forecast future disease severity on the basis of visual field threshold sensitivities during a sample visit. In the literature, the problem of forecasting glaucoma evolution has been tackled by many authors [15]. It has been shown that machine learning classifiers (MLCs) out-perform traditional statistical analysis packages in glaucoma diagnosis [16–18]. Also, progression detection was shown to benefit from predictive modelling [19–21]. In these works, VF sensitivities have been used either to distinguish between glaucomatous and normal patients or to predict future disease progression based on specific criteria. In this paper we will consider disease severity, and we attempt to predict it in different reliability scenarios. To account for disease severity, we focus on scoring metrics used to assess visual loss.

A number of clinical trials have proposed scoring strategies to assess the severity of visual field loss for a specific test and to define disease progression during follow-up [22–24]. Advantages and limitations of such tests have been highlighted in several studies, mainly focusing on progression characterisation [25–28]. Even though effective visual field scoring and accurate progression definitions are highly desirable, the choice of standardised criteria is still very controversial in the clinical community. We selected the Advanced Glaucoma Intervention Study (AGIS) scoring system to assess VF test results. This choice comes from the current

Download English Version:

https://daneshyari.com/en/article/10320555

Download Persian Version:

https://daneshyari.com/article/10320555

Daneshyari.com