AI for Public Health: Self-Screening for Eye Diseases

Xiaohui Liu and Gongxian Cheng, University of London
John X. Wu, Moorfields Eye Hospital

A SOFTWARE-BASED VISUAL-FIELD TESTING SYSTEM INCORPORATES SEVERAL AI COMPONENTS, INCLUDING MACHINE LEARNING, AN INTELLIGENT USER INTERFACE, AND PATTERN DISCOVERY. THIS SYSTEM HAS BEEN SUCCESSFULLY USED FOR SELF-SCREENING IN SEVERAL DIFFERENT PUBLIC ENVIRONMENTS.

MEDICAL DIAGNOSTIC TESTS give doctors important clinical information. Such tests normally require a specially designed instrument and a medical operator's attention; a doctor then analyzes the test results. These testing instruments are often very expensive and not widely available for mass screening—for example, testing in large-scale field investigations, general practitioner clinics, and public halls. To make these tests easily accessible to the community, thus improving recognition of early signs of disease, we must seek an economical way of providing the tests without compromising reliability.

Developing a test capable of mass screening involves three important issues: *delivery*, *interface*, and *interpretation*. First, the test should be easy to deliver without too much cost or compromise in its reliability. Second, the test interface should adequately support the subject undertaking the test, preferably providing a self-testing environment where no instructions from a medical operator are necessary. Third, the test system should be able to interpret the test results, to give the subject a general warning of possible problems, without a doctor's involvement.

For the last 10 years, we've developed a self-screening test in which PCs without specialized hardware can examine the visual field. To address the three issues, the test system has three main AI components: machine-learning programs (for example, neural networks and decision-tree induction), an intelligent user interface, and a pattern-discovery model. Operating on portable or desktop computers, this system has been successfully used in several different public environments.

Perimetry

Visual-field testing, called *perimetry*, can provide early detection of eye diseases such as glaucoma and optic neuritis (see the "Glaucoma and optic neuritis" sidebar for a brief description of each). Perimetry requires a medical operator to explain and monitor the test, and a perimeter, a specially designed instrument that checks how wide the subject can see when focusing on the fixation point on the screen. The instrument can display a large number of stimuli varying in sizes and intensities and can record the subject's responses. It offers facilities for inspecting various reliability factors such as *fixation losses*, *false negative responses*, and *false positive responses* (see the "Reliability factors" sidebar).

Figure 1 depicts a standard perimeter in use. Perimeters of this type have been successfully tried in clinical environments. The problem is that they are specially designed, very expensive, and not widely available. Their use is restricted mostly to eye hospitals. Therefore, making them available for mass screening is difficult. On the other hand, people should undergo visual-field testing at the earliest possible stage. By the time a person has displayed overt symptoms and has been referred to a hospital for eye examination, the visual-field loss might already be at an advanced stage and not easily treated.

For many years, psychophysical researchers, influenced by the pioneer work of M. Flocks and his colleagues,[1] have been try-

1094-7167/98/$10.00 © 1998 IEEE

ing to apply CRT (cathode ray tube) technology as a stimulus-display device for visual-function testing. Flocks' group and his followers partially addressed the test-delivery issue—in theory, using television for eye testing lets many ordinary people perform the test without additional cost. However, this idea was not properly developed and realized, largely because human behavioral variants during testing made obtaining reliable test results difficult, especially in a self-testing environment. The collected data contained much noise, making the interpretation of the results particularly difficult. Also, such an approach did not address the issue of the interface between the subject and the machine; communication was essentially one-way. Finally, the machine did not record the test results, and interpretation of the data recorded on paper was tedious.

Computer-controlled video perimetry, the first visual-stimuli-generating program implemented on portable PCs, has demonstrated early success in detecting visual-field damage, especially under certain controlled test environments.[2] Like Flocks' work, CCVP partially addresses the delivery issue by using PCs as testing machines, and the data collected from individual subjects contain much measurement noise. Unlike Flocks' work, this software-based approach addresses the interface issue: communication between the subject and the PC is two-way. Moreover, on-line interpretation of test results is possible.

AI for visual-field self-screening

To develop an effective self-screening system, we integrated AI methods with CCVP. Figure 2 shows the transformation from a standard perimetry system to a self-screening system. Three software components made this transformation possible. First, in place of the specially designed perimeter and its associated testing methods, the self-screening system uses software-controlled perimetry that operates on PCs. Second, instead of having an experienced medical operator constantly monitoring the subject during the test, the system incorporates an intelligent user interface, which allows helpful interaction between the subject and the test. Third, rather than asking a doctor to make sense of the test results, the system attempts to provide the subject with important information regarding his or her visual function. This requires data-analysis capabilities, especially the understanding of how various diseases manifest themselves in the test data.

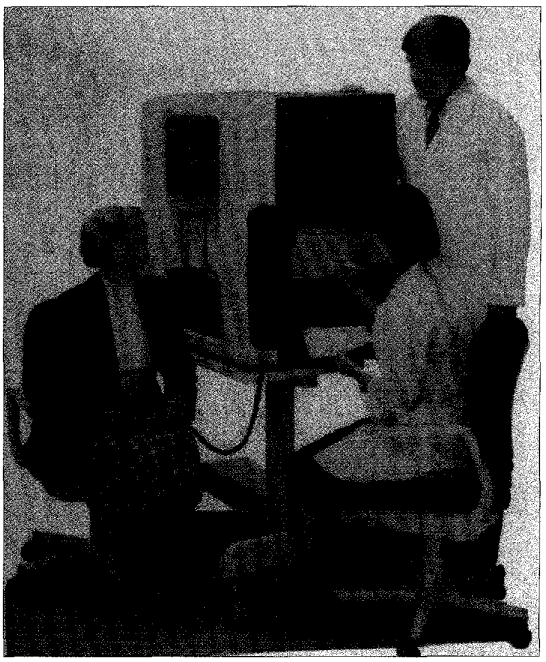

Figure 1. A standard perimetry, which tests a subject's visual field.

Because we are primarily interested in screening—identifying individuals who are most likely to develop serious eye diseases—

Glaucoma and optic neuritis

Glaucoma affects approximately one in 30 people over the age of 40 and is the second-largest cause of blindness in the developed world. Glaucoma is a condition, sometimes associated with high pressure in the eye, that over many years can damage the retinal nerve fibers at the back of the eye. A badly affected person only notices what she or he is directly looking at and misses things to the sides (a bit like wearing blinkers). Severe cases can eventually result in complete blindness. The earlier the condition is detected, the better the chance of preserving sight with treatment.

The diagnosis of glaucoma normally requires a variety of clinical information such as disk appearance, intraocular pressure, and visual field. However, only visual-field testing, called *perimetry*, provides early detection of the disease.

Optic neuritis is the most common optic-nerve disease to affect young people. The average age at the first attack is 31 years, but teenagers and people over 40 might develop this disease for the first time. It is an inflammatory disease that affects more women than men.

Its exact cause is unknown, but the symptoms are blurred central vision (the vision used to read and see fine detail), reduced color vision, and reduced sensation of light brightness. Commonly there is aching pain in the eye made worse by eye movement. The pain often goes away within a week or so, but the blurred vision lasts for weeks to several months. Optic neuritis can be a serious problem that leads to significant visual disability.

For decades, physicians have treated optic neuritis with steroids. However, because steroids can cause adverse side effects, the search continues for the best treatment of the disease.

Reliability factors

For perimetric results to be valid, the subject must focus on the fixation point. The subject's fixation can be checked by presenting stimuli in his or her blind spot. If the subject responds, a *fixation loss* has resulted.

Occasionally during each test, the projector moves as if to present a stimulus but does not do so. If the subject responds, a *false positive response* has occurred.

At other times, the perimeter presents a stimulus that is much brighter than normal, but the subject does not respond. This is a *false negative response*.

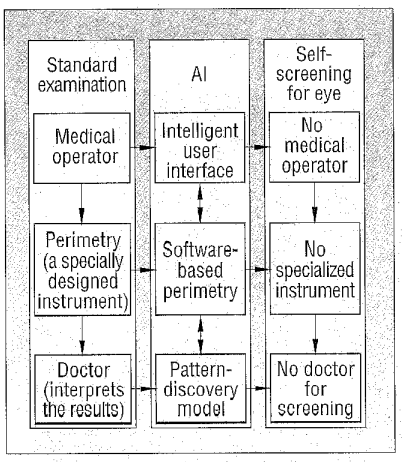

Figure 2. AI transforms standard perimetry to self-screening perimetry.

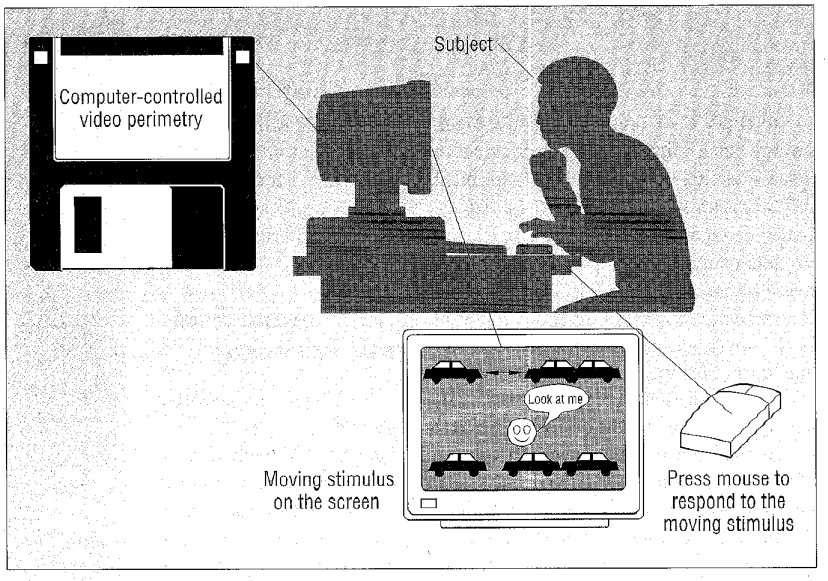

Figure 3. Software-based visual-field examination.

providing the subject with warning messages and advice regarding the possible danger is sufficient. Therefore, "No doctor for screening" in Figure 2 means there is no need for a doctor to get involved in identifying the danger; it does not mean that no doctor is required in the diagnosis and treatment.

Software-based perimetry. Figure 3 demonstrates how this perimetry works. The test examines several locations on the test screen that correspond to crucial positions in the visual field. The test screen consists of a number of objects of the same type; at any stage of the test, only one of them (the stimulus) is moving. Prompted by the test, the subject, using one of his arms supporting his chin in front of the fixation point (the smiling face on the screen), clicks on the mouse to respond to the stimulus. Figure 4 shows a sample test screen, with six test locations (numbered).

The CCVP program obtains repeated measurements over test locations. This should ensure that the system reliably estimates the subject's visual function even if one or more measurement cycles include noise in the form of false positive or false negative responses.

However, software-based perimetry poses significant problems. We no longer have the dedicated hardware capable of monitoring the subject's behavior and providing reliability indicators. Also, we no longer have the luxury of a standard test environment where we can strictly control important factors such as light and viewing distance. Therefore, the software-based test will naturally be less reliable in that the data collected from subjects will contain more measurement noise.

The principle we have adopted for identifying noise is that the interesting properties in data are more stable than noise.[3] For example,

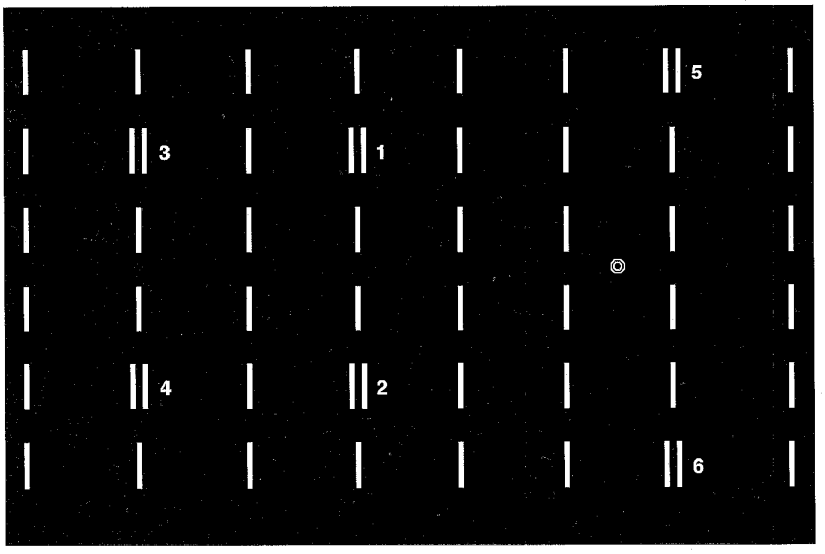

Figure 4. A computer-controlled video perimetry (CCVP) screen layout. The numbers indicate the test locations.

the property that a normal person with no visual-function loss can usually see the stimuli on the test screen is more stable than the fluctuation in data caused by occasional false negative responses for whatever reasons. The key to implementing this principle is to identify the more stable parts of the data. The less stable parts of the data, which are inconsistent with those features, will then be exposed.

Using the SOM. We have developed a computational method for identifying measurement noise primarily using Teuvo Kohonen's *self-organizing maps*.[4] An SOM consists of two layers of nodes. The input layer is a vector of N nodes that present the input patterns to the network, and the output layer is often a 2D array of M output nodes (the output map). Figure 5 shows a simple Kohonen network with two input and nine output nodes. Each input node fully connects to every output node through a connection weight, so each output node has an associated *weight vector*.

An SOM defines a mapping from the input data space onto a set of nodes on the output map and can map similar input patterns onto geometrically close output nodes. When the SOM receives an input vector, the algorithm computes the distance between it and each of the weight vectors. The output node whose

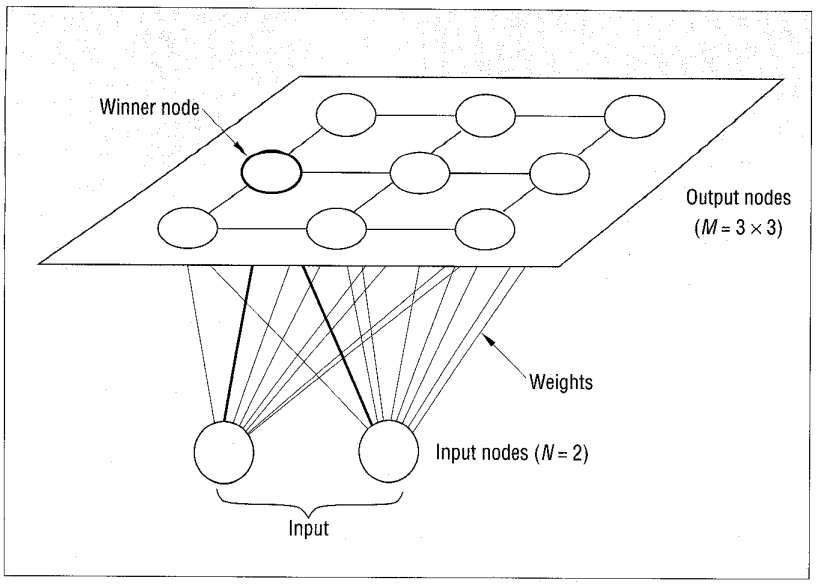

Figure 5. A simple Kohonen network.

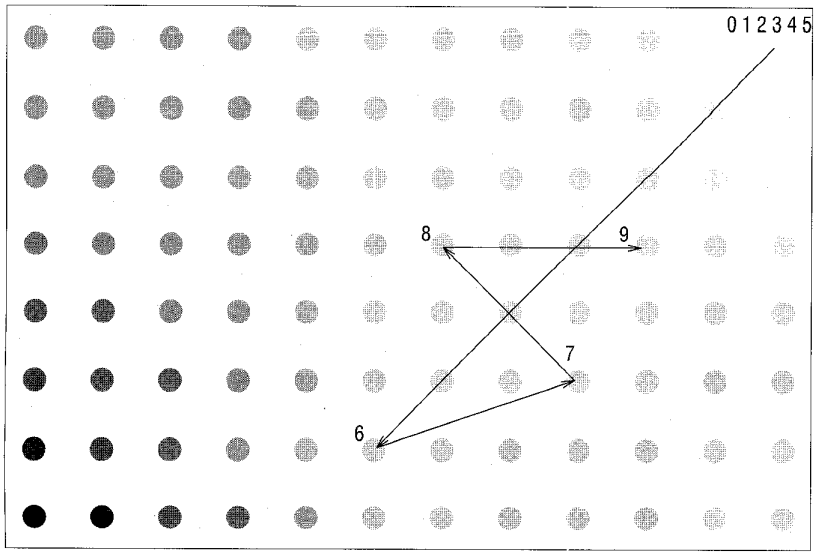

Figure 6. Reflecting fatigue on the self-organizing map (SOM) output map.

weight vector is closest to the input vector is called the *winner node* or the *location of the response*. The response's exact magnitude need not be determined: the algorithm simply maps the input vector onto this location on the output map.[4] It then defines a symmetric neighborhood of nodes surrounding the winner node. The algorithm updates the winner node and those in the neighborhood such that their weight vectors gradually get closer to the current input vector.

Adding visual-field test data. Applying the SOM to the visual-field test data involves these steps:[5]

(1) *Define the terms*: First, V is the input data set such that each $\mathbf{v}^i \in V$, a vector of N dimensionality with values over the set {0, 1}, corresponds to the response pattern from one measurement cycle such that $\mathbf{v}^i_k$ is 1 if the subject can see the kth stimulus in the cycle, and 0 otherwise. A is the output space and M is the number of output nodes. Second, W is a set of connection-weight vectors where each output node j ($1 \le j \le M$) is associated with a connection weight vector of N dimensionality of the form $\mathbf{w}_j = (w_{j1}, \ldots, w_{jN})$, where w_{jk} is the connection weight between the input node k and the output node j. Third, η is the gain value that affects the rate of adjustment of the connection-weight vectors; N_c is a round neighborhood of the output node c; and r is the radius of N_c. For any output node j, $j \in N_c$ if the distance between c and j in the output map is not greater than r.

(2) *Initialize the output map's topology and size*. Set the value for M.

(3) *Initialize the weights*. Initialize the connection weights to random values over the interval [0.0, 1.0], and normalize both the input vectors and the connection-weight vectors. Initialize η and r.

(4) *Present the new input*. Set i to $i + 1$ and present input vector $\mathbf{v}^i$.

(5) *Select the minimum distance*. Compute the distance between the input vector $\mathbf{v}^i$ and each output node j:

$$d(\mathbf{v}^i, w_j) = \sum_{k=1}^{N} \left(\mathbf{v}^i_k - w_{jk}\right)^2$$

Designate the winner node with the minimum distance as c.

(6) *Update the weights and neighborhood*. Adjust the connection-weight vectors of c and N_c—that is, for each node $j \in N_c$ perform $w_j^{(\text{new})} = w_j^{(\text{old})} + \eta\,[\mathbf{v}_i - w_j^{(\text{old})}]$. Decrease both r and η.

(7) *Repeat by going to Step 4*. This iteration process continues until it produces a stable network. In our experiments, we used several thousand input vectors to train the network and iteratively submitted them 100 times in random orders to achieve convergence.

Once this method obtains a stable network, the network operates in *recall* mode, in which the map responds to an input vector without modifying its weights. As we mentioned before, the system computes the distance between each input vector and its weight vector, and the competition among nodes follows.

In our application, a single input vector, representing the results of a single measurement cycle, consists of a 2D array of measurements taken at six locations; one or more stimuli test each location. A single field test consists of 10 such cycles and therefore generates 10 input vectors. Because each input vector produces a winner node, a single field test produces 10 winner nodes; they constitute a transition trajectory on the output map (see Figure 6).

Figure 6 demonstrates the results of one

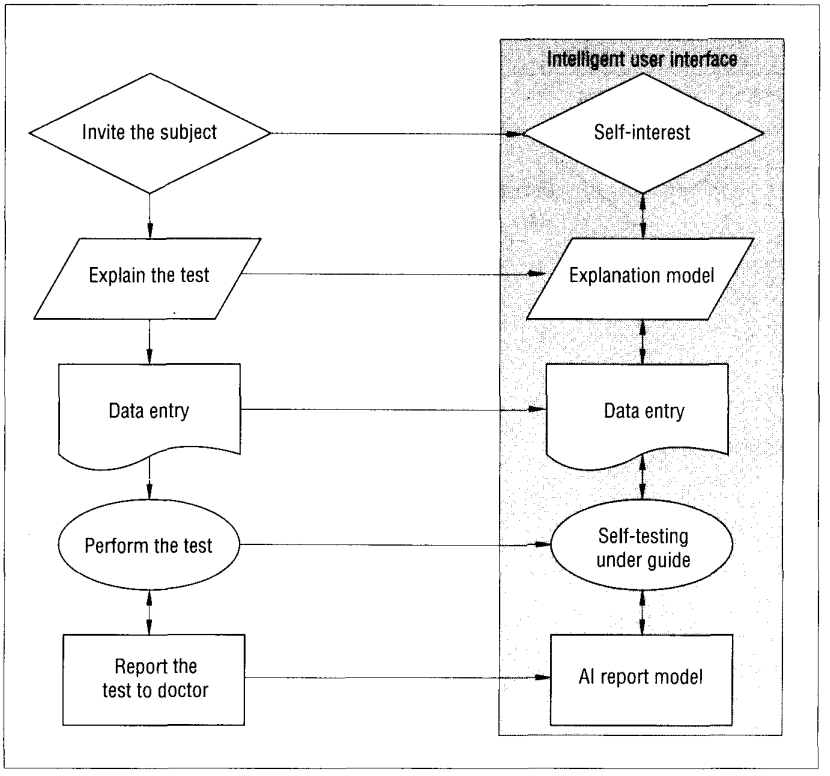

Figure 7. Functions of the intelligent user interface, compared to a standard visual-field test.

particular test performed on a subject's right eye. One of the map's physical meanings, produced by calculating the average *sensitivity* (the proportion of positive responses) of all the input vectors associated with each output node,[5] is as follows. The top-right region of the map is most sensitive, and the sensitivity gradually fades toward the bottom-left region. Therefore, we can easily see that this subject could see most of the time during the first six measurement cycles. The nodes of the next four cycles move away from the first node to some nodes with lower sensitivity values, which indicates that the subject cannot see as clearly. This is most likely a case of fatigue; we can then delete the data vectors associated with nodes 6 through 9.

This method can identify the more stable part of the data (and therefore measurement noise) and provide information about the subject's behavior during the test. For example, it can indicate whether the subject had a reliable test or demonstrated behaviors that adversely affect the test's accuracy.[5]

A rule for identifying stable data. Because SOM maps similar input vectors onto geometrically close winner nodes, we can form a general rule for identifying the more stable part of the data: If most of the winner nodes center around one particular region, the input data vectors associated with these nodes constitute the more stable part of the data. One way of implementing this rule would be to search for the maximum set of neurons that occupies the smallest topographical area on the output map. In particular, we can use an evaluation function (see Equation 1) for this purpose; the objective is to find a subset of winner nodes, S, that minimizes the value of $F(S)$.

$$F(S) = A(S(k))/k^2 \quad (k = P, P-1, \ldots, \lfloor P/2 + 1 \rfloor) \quad (1)$$

P is the total number of winner nodes (10); A denotes the topographical area in the map occupied by a subset of winner nodes; $S(k)$ represents a subset of winner nodes with k members.

Intelligent user interface. Figure 7 illustrates how the medical operator normally manages the standard visual-field test and how the IUI helps provide self-screening. Normally, the operator invites a subject for a test, explains the test to the subject, and enters the subject's details. The operator then monitors and controls the test, and reports the results to the doctor.

For self-screening, a subject examines the visual field out of his or her interest in the early detection of possible diseases. An explanation model explains the test and procedure, and the subject enters the data. Self-testing then proceeds under the system's guidance. Finally, the system issues a report, informing the subject whether he or she has passed the test and, if not, which part of the visual field appears to have problems. The report is based on results from the pattern-discovery model, which we'll discuss later.

One of the key considerations in the development of the test interface was how to handle human behavioral instability during the test so that the subjects would make fewer mistakes (false positive or false negative responses). The major behavioral factors are learning effects, inattention, and fatigue.

Learning effects. Subjects are more likely to make mistakes the first time they take the test. Subjects who have taken the test before, but who are retested after a considerable time interval, might also make mistakes.

The IUI provides self-training sessions for such users before the actual test. The program keeps the subjects in the training session until it judges them to be familiar with the test. It bases this judgment on response time (from when the screen displays a stimulus to when the subject responds). If the response times for most of the responses collected during the test are close to each other and no great irregularity exists, the subject can move on to the actual test.

Inattention. To keep the subject alert during the test, we've taken care to ensure that taking the test is an interesting experience rather than something that the subject "has to do." Techniques include

- a feedback system that uses sound and text to indicate the subject's performance,
- an adaptive fixation point that uses smiling and frowning faces to attract the subject's attention,
- employing interesting test stimuli, and
- customized test strategies for individuals.

For example, one of the original test screens consists of a number of vertical bars with a central circle for the fixation point (see Figure 4). The program tests a fixed number of bars, using several different stimuli, including bar movement and flicker. Some subjects might find this screen layout and

these test stimuli monotonous, and therefore have difficulty concentrating on the test. So, we developed alternative screen layouts and test stimuli. For example, we've used cars and faces rather than vertical bars. Instead of the simple movements of bars as stimuli, we have headlights flashing and windshield wipers moving.[6] These developments have led to a test that allows several different screen presentations, each providing a different test image and presenting different degrees of stimuli.

Also, we introduced a special function to check whether the subject is really concentrating on the test. The display of the stimuli on the screen is now irregular—for example, stopping for a while after a fixed number of regular displays. This strategy is particularly effective for dealing with subjects who anticipate the frequency of presentation of the stimuli.

Fatigue. We've also concentrated on reducing fatigue, which often causes noise in the data. Not every subject needs to go through the same number of measurement cycles. The test results from the first few cycles might provide sufficient information. Therefore, determining whether the system has obtained sufficient information at a certain stage of the test is important. If it has, we can stop testing.

This kind of dynamic visual-field testing strategy has these key steps:

(1) Calculate the sensitivities of a few initial measurement cycles.
(2) Use these calculated sensitivities to predict the sensitivity of all 10 cycles.
(3) Calculate the squared error between the sensitivity of a few initial cycles and the predicted sensitivity of all 10 cycles.
(4) Compare the squared error with a predetermined tolerance. If the squared error is less than the acceptable tolerance, stop testing. Otherwise, perform another measurement cycle, and repeat Steps 1 to 4.

We tested this strategy on a set of visual-field data involving six test locations, four test stimuli, and 10 measurement cycles. We used techniques such as neural networks, multiple regression, and decision-tree induction to implement Step 2.[7] The strategy kept the number of measurement cycles relatively low without much adverse effect on diagnostic accuracy. In particular, Quinlan's decision-tree induction program, C4.5,[8] achieved approximately 95% accuracy while saving on average two measurement cycles per test. We built the prediction (classification) models in the form of decision trees, using four attributes, eleven classes, and several hundred cases.

To illustrate these attributes and classes, here are the clinical test results for one subject at location 3 (see Figure 4):

- Stimulus 1 0 1 1 1 0 1 1 1 1 0
- Stimulus 2 1 1 1 0 1 1 1 1 1 0
- Stimulus 3 0 1 1 1 1 1 1 1 1 1
- Stimulus 4 0 1 1 1 1 1 1 1 0 0

TO KEEP THE SUBJECT ALERT DURING THE TEST, WE'VE TAKEN CARE TO ENSURE THAT TAKING THE TEST IS AN INTERESTING EXPERIENCE RATHER THAN SOMETHING THAT THE SUBJECT "HAS TO DO."

The first column of digits shows the results for the first cycle; this subject made no response to Stimulus 1, responded to Stimulus 2, and made no responses to Stimuli 3 and 4. The second column shows that, on the second cycle, the subject responded to all four stimuli, and so on.

Suppose we try to use the sensitivity values of the first five measurement cycles to predict the sensitivity values of all 10 cycles for Stimulus 1. In this case, 0.6 (3/5), 0.8 (4/5), 0.8 (4/5), and 0.8 (4/5) are the sensitivity values for the four stimuli after the first five cycles, and 0.7 (7/10) is the sensitivity value for Stimulus 1 after all 10 cycles. [0.6, 0.8, 0.8, 0.8, 0.7] can then become one of the several hundred cases for building the classification (prediction) model in Quinlan's C4.5. We use the sensitivity values for all four stimuli after five cycles to predict the sensitivity value of Stimulus 1 after 10 cycles (and likewise for the prediction of the other three). The first four elements are the attribute values, and the last element (0.7 in this example) is the value of the class variable. The class variable has eleven possible values (0, 0.1, 0.2, ..., 1), because we are concerned about the number of positive responses out of a total of 10 (0, 1, 2, ..., 10).

Pattern-discovery model. At the end of a visual-field self-testing, subjects are naturally interested in knowing whether their vision is abnormal. The average sensitivity values for different testing locations give an overall assessment of how well the subject has done. However, an indication of the likely diagnosis based on these sensitivity values and on the relationships between different test locations would be desirable. For example, if the sensitivity values for a subject are low for testing locations 3 and 4 (see Figure 4) but high for other testing locations, might the subject have a certain eye disease?

To determine which test results indicate possible eye disease, we compiled over 3,000 clinical test records from patients with visual-field loss from glaucoma and optic neuritis. We then analyzed them using interactive data exploration where a data analyst steered the discovery process.[9]

The analyst performed three important tasks. First, the analyst used relevant domain knowledge to form initial hypotheses regarding what to explore in the data. Second, the analyst organized and preprocessed different data sets according to various criteria such as disease types and sampling considerations before using those sets to extract the behavioral relationships between test locations. Third, after the relevant features were extracted by neural networks and were displayed by visualization techniques, the analyst organized, observed and analyzed them. This was very much an interactive, iterative process where the analyst made numerous decisions—for example, whether to select more specific data or to take alternative action, whether any patterns detected had any significant meanings, and whether they could be validated.

Assuming that true features are more repeatable than false ones, we verified the features found from the left-eye group using those from the right-eye group. With glaucomatous data, we found a strong correlation among locations in the upper hemifield (1, 3, and 5 in Figure 4), and in the lower hemifield (2, 4, and 6). This finding is consistent with that of early research on conventional visual-field test methods. For optic-neuritis data, we have also found a strong correlation between certain pairs of retinal locations, but

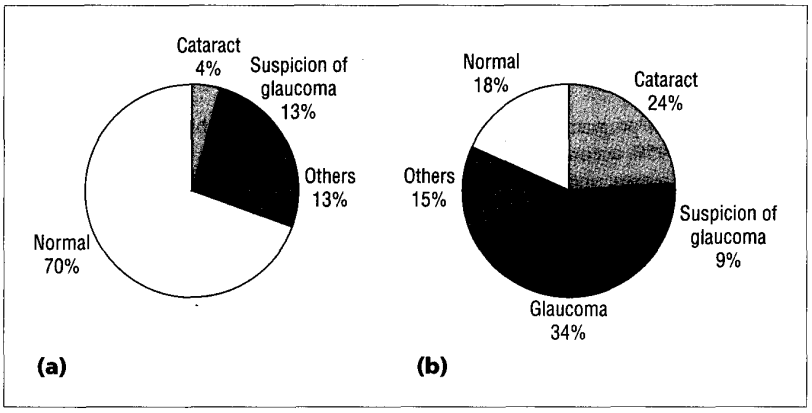

Figure 8. The distribution of clinical diagnosis by an ophthalmologist: (a) selected patients who passed the test (the control group, $n = 45$); (b) patients who failed the test ($n = 33$).

this time across the two hemifields.[9] Topographic analysis of clinical chorioretinal changes (changes to the choroid, a vascular membrane near the retina, or to the retina) related to the sensitivity at these six test locations has confirmed this finding. These discovered patterns then become useful in providing warnings to the subject regarding the possible danger.

Self-screening in the community

Our self-screening system has been used in a World Health Organization program for preventing optic neuritis,[10] in occupational health screening,[11] and, most recently, in a pilot study by the UK's Medical Research Council to detect people with glaucoma. These studies have been conducted to evaluate the test's acceptance (that is, how many people agree to take the test) in different primary-care settings.

In the WHO study—mass screening for visual-field loss caused by optic neuritis—the subjects were from a farming community in Africa. These subjects were largely computer-illiterate; the test was conducted in farmers' houses, using portable PCs. In the occupational health study—screening for ocular abnormalities and vision defects—the subjects were the employees of a large telecommunication company. These subjects were young (the average age was about 30) and healthy, and they used computers daily; the test was performed in the company's offices using PCs connected by a local-area network. In the glaucoma study, the subjects were patients of a general practice, and the test was conducted in the waiting room, using a desktop PC. Each study examined a large number of subjects and reported good acceptance of the test.

Let's look at the third study in more detail. The test was offered during routine attendance at a large urban general practice in North London and was conducted by the subjects themselves in the waiting room. For a three-month period during the pilot study, all patients aged 40 or older who routinely attended the practice were offered the test. Upon entering the clinic, each patient received an information sheet explaining the purpose of the pilot study, the nature of glaucoma and the visual-field test, what to expect during and after the test, and information about whom to contact if they wished to know more about the test in general or were concerned about their own results. Each interested patient then signed a consent form.

Of the 925 people tested during the three-month period, 33 failed the test. That is, the results indicated an abnormality in their visual field. These 33 people, together with 45 of those who passed the test (controls), were later assessed clinically in the practice by an ophthalmologist.

Figure 8 summarizes the results. The overall picture is clear: the group who failed the test had many more eye problems than the group who passed the test. For example, 70% of those who passed the test (controls) had a normal visual field, and none of them was a confirmed glaucoma case. On the other hand, 82% of the people who failed the test had various visual defects, including 34% confirmed glaucoma cases and 9% glaucoma suspects (those who had increased intraocular pressure or whose optic disc appeared abnormal, but who had no visual-field loss).

These findings are particularly encouraging. An overwhelming majority of the subjects did not consider the possibility of having any eye problems when they visited the clinic. This opportunistic test has shown high sensitivity and has allowed for the early detection of eye diseases such as glaucoma.

Acceptance is also one of the most important issues in evaluating a screening test or an opportunistic test. Although little was done to increase the number of patients participating (no advertising, and little stimulation from the clinic staff), this study showed higher-than-expected acceptance: of the 1,215 people who were offered the test, 925 (76%) accepted. (The acceptance of opportunistic tests in city practices generally ranges from 50% to 70%.) This is an encouraging finding in such an elderly population during a very short period.

We've concluded that testing patients in a general practitioner's waiting room offers a good opportunity to screen patients efficiently and without additional cost. Moreover, it provides patients with an interesting alternative to the usual activities in the waiting room—reading newspapers and magazines, chatting with others, and so on.

PUBLIC HEALTH IS THE COLLECTIVE action taken by society to protect and promote the health of entire populations.[12] Medical screening, with its focus on the prevention of disease at the population level, is one of the most important tools contributing to public health.

As computers become ever more accessible, software-based tests are an obvious approach to mass screening in the community without additional hardware costs. Although an overwhelming majority of general practices in UK have PCs, no solid evidence indicates that this has led to significant clinical improvement.[13] One of the problems is the lack of clinically orientated software for general practitioners. By making such software available in GPs' clinics, we could efficiently use existing computing resources and provide additional medical care to the community at minimal capital cost.

Widespread implementation of clinically oriented software in primary-care systems might ultimately help improve health care and its cost-effectiveness. We plan to use this test system in more public environments to gain more experience, which will allow us to improve the system further. ■

Acknowledgments

The British Council for Prevention of Blindness, the International Glaucoma Association, the UK's Medical Research Council, and the World Health Organization supported this work in part. We thank Barrie Jones, community health scientist; Roger Hitchings, leading glaucoma expert; Richard Wormald, ophthalmic epidemiologist; Fred Fitzke, visual scientist; and Stephen Corcoran, general practitioner, for their cooperation. We are grateful to Roger Mitton for reading and commenting on early drafts of the article and to other members of the Intelligent Data Analysis Group at Birkbeck for their contributions, especially Kwa-Wen Cho on dynamic testing and Jo Collins on the user interface. Finally, we thank Dan O'Leary and the anonymous referees for their constructive and helpful comments, and the staff of *IEEE Intelligent Systems* for their careful style editing.

References

1. M. Flocks, A.R. Rosenthal, and J. Hopkins, "Mass Visual Screening via Television," *Ophthalmology*, Vol. 85, 1978, pp. 1141–1149.
2. J.X. Wu, *Visual Screening for Blinding Diseases in the Community Using Computer Controlled Video Perimetry*, PhD thesis, Inst. of Opthalmology, Univ. of London, 1993.
3. S. Becker and G.E. Hinton, "Self-Organizing Neural Network That Discovers Surfaces in Random-Dot Stereograms," *Nature*, Vol. 355, Jan. 1992, pp. 161–163.
4. T. Kohonen, *Self-Organizing Maps*, Springer-Verlag, Berlin, 1995.
5. X. Liu, G. Cheng, and J.X. Wu, "Identifying the Measurement Noise in Glaucomatous Testing: An Artificial Neural Network Approach," *Artificial Intelligence in Medicine*, Vol. 6, 1994, pp. 401–416.
6. J. Collins, *EYETEST: A Perimetry Test Developed to Aid the Detection of Glaucoma*, MSc dissertation, Dept. of Computer Science, Birkbeck College, Univ. of London, London, 1993.
7. K.W. Cho, X. Liu, and G. Loizou, "Decision Making in Dynamic Visual Field Testing by Backpropagation: C4.5 and Multiple Regression," *Proc. Sixth Int'l Conf. Information Processing and Management of Uncertainty in Knowledge-Based Systems*, Proyecto Sur de Ediciones, Granada, Spain, 1996, pp. 401–406.
8. J.R. Quinlan, *C4.5: Programs for Machine Learning*, Morgan Kaufmann, San Francisco, 1993.
9. G. Cheng et al., "Discovering Knowledge from Visual Field Data: Results in Optic Nerve Diseases," *Medical Informatics Europe '96*, J. Brender et al., eds., IOS Press, Amsterdam, 1996, pp. 629–33.
10. G. Cheng et al., "Establishing a Reliable Visual Function Test and Applying It to Screening Optic Nerve Disease in Onchocercal Communities," *Int'l J. Bio-Medical Computing*, Vol. 41, 1996, pp. 47–53.
11. L. Wright et al., "Motion Sensitivity Testing in Occupational Health Screening," *Perimetry Update 1994/95*, Kugler Publications, Amsterdam, 1994, pp. 335–338.
12. R. Beaglehole and R. Bonita, *Public Health at the Crossroads*, Cambridge Univ. Press, Cambridge, UK, 1997.
13. J.C. Wyatt, "Clinical Data Systems, Part 1: Data and Medical Records," *Lancet*, Vol. 344, 1994, pp. 1682–1688.

Xiaohui Liu is a senior lecturer in the Department of Computer Science at Birkbeck College, University of London. His research interests are in AI and intelligent data analysis, particularly their application to challenging real-world problems. He received his PhD in Computer Science from Heriot-Watt University, Edinburgh. He is a member of the AAAI, ACM, British Computer Society, and IEEE Computer Society. Contact him at the Dept. of Computer Science, Birkbeck College, Malet St., London WC1E 7HX, UK; hui@dcs.bbk.ac.uk.

Gongxian Cheng is a college associate research fellow at Birkbeck College, University of London. His research interests include artificial and computational intelligence, intelligent data analysis, and pattern recognition. He received his Bachelor of Computer Science from Fudan University and his Master of Computer Engineering from the Beijing Information Technology Institute. Contact him at the Dept. of Computer Science, Birkbeck College, Malet St., London WC1E 7HX, UK; g.cheng@dcs.bbk.ac.uk.

John X. Wu is a senior research fellow at Moorfields Eye Hospital. His main research interest is in the screening of optic nerve diseases. He received his PhD in public health from the University of London. Contact him at the Glaxo Dept. of Ophthalmic Epidemiology, Moorfields Eye Hospital, Inst. of Ophthalmology, Bath St., London EC1V 9EL, UK; john.wu@ucl.ac.uk.