

## Web-StrabNet: a web-based expert system for the differential diagnosis of vertical strabismus (squint)

A.C. Fisher<sup>a\*</sup>, S.P. Lake<sup>a</sup>, I.P. Cunningham<sup>b</sup> and A. Chandna<sup>b</sup>

<sup>a</sup>Department of Medical Physics and Clinical Engineering, Royal Liverpool University Hospital, Liverpool, L7 8XP, UK; <sup>b</sup>Vision Assessment Unit, Department of Paediatric Ophthalmology, Royal Liverpool Children's Hospital, Liverpool, L12 2AP, UK

(Received 7 October 2008; final version received 20 April 2009)

A squint, also known as strabismus, is a condition where the eyes are misaligned because of incorrect balance in the controlling eye muscles. This may result from muscular, neuromuscular or purely mechanical factors. An affected eye will have either predominating vertical or horizontal deviation. Vertical deviations are usually classified into eight classes (diagnoses) and horizontal into 10. The present work considers only the former but extension to the latter is straightforward.

The differential diagnosis of strabismus is usually achieved in the prism cover test (PCT). A range of test prisms is presented to the eye and the resulting deviation in a particular direction of gaze is observed. In the full PCT, 10 positions of gaze are considered: in each position there are, say, 40 prisms of plus and minus power to investigate. The problem can be expressed as the inference 1-of-8 diagnoses in the output space from an input space of 10 parameters each with a resolution of 1-in-80. However, in the majority of clinical examinations, the corner-most positions of gaze are difficult to assess, particularly in children. Therefore, frequently the 6-position subset is reported requiring a corresponding reduction in the input space dimensions of 10 to 6. Web-StrabNet<sup>®</sup> is an expert system for the differential diagnosis of strabismus based on parallel instances of multi-layer perceptrons trained on exemplar data generated in consensus by two clinical experts. This machine expert is programmed in MatLab<sup>™</sup> and is freely available as an Internet website ([www.strabnet.com](http://www.strabnet.com)) which uses MatSOAP<sup>®</sup>, an XML/SOAP accessible automation server running a number of simultaneous MatLab instances.

StrabNet achieves diagnostic accuracies of 100 and >94% with artificial data and typically ~99 and ~99% in clinical data-sets for the 10-position and 6-position subset PCT's, respectively.

**Keywords:** StrabNet; strabismus; squint; expert system; differential diagnosis; prism cover test

### 1. Introduction

The motivation for this work was to provide an accessible tool in support of training and clinical diagnosis in what is commonly regarded as a challenging area of ophthalmology to the non-specialized clinical practitioner. A previous study of the differential diagnosis of strabismus by an expert system from prism cover test (PCT) data has previously been reported for the full 10-position PCT [2]: here this model is extended to the 6-position subset.

---

\*Corresponding author. Email: [a.c.fisher@liv.ac.uk](mailto:a.c.fisher@liv.ac.uk)

Strabismus or squint refers to a misalignment of the visual axes between the two eyes that maybe in the horizontal or vertical plane or both. In addition, the visual axes may be rotated (torsion) either in isolation or in combination with other planes [5]. It affects 3–4% of the population. The presence of strabismus is confirmed by a combination of tests, which ultimately leads to a measurement of the strabismus in degrees of deviation or prism dioptres ( $\Delta D$ ). For all tests the individual is asked to look and hold gaze in a particular direction. Distance and near fixation to a target are the most important positions. However, strabismus is examined and may be measured in up to eight other gaze positions to fully characterise the pattern of the deviation and achieve the differential diagnosis (see Figure 1) [1].

Inspection of the position of the eyes may reveal an obvious squint by either assessment of light reflexes in the pupil, which would be asymmetrical, or the Cover–Uncover Test which is the gold standard. The Cover–Uncover Test involves covering one eye while the subject views a target at either distance (6 m) or near (33 cm) and the tester records any movement of the other eye. Such movement required to achieve fixation indicates the presence of manifest strabismus.

The PCT determines the magnitude of strabismic deviation. Prisms of increasing strength are placed with the apex of the prism in the direction of the deviation and combined with a cover test. The endpoint is reached when the deviation just reverses direction. The strength of the prism required is the magnitude of the deviation. A graphical description of the full PCT is available at: [www.strabnet.com](http://www.strabnet.com) (see Details of the PCT) [4]. The differential diagnoses for vertical strabismus and their StrabNet classes are shown in the Table 1.

## 2. Method

The StrabNet expert system uses an array of multi-layer perceptrons using back-propagation learning. The array comprises three neural networks (see Figures 2 and 3):

- (1) Net 1: classifies PCT data into either Hypertropic (upward) or Hypotropic (downward) squints;
- (2) Net 2: classifies Hypertropic PCT data into StrabNet Classes 1–4 (see Table 1);
- (3) Net 3: classifies Hypotropic PCT data into StrabNet Classes 5–8 (see Table 1).

Inputs to each net are the 10 PCT prism values (6 in the 6-position subset) in the range  $-40$  to  $+40 \Delta D$ , with each net configured with three layers with the hidden layer having

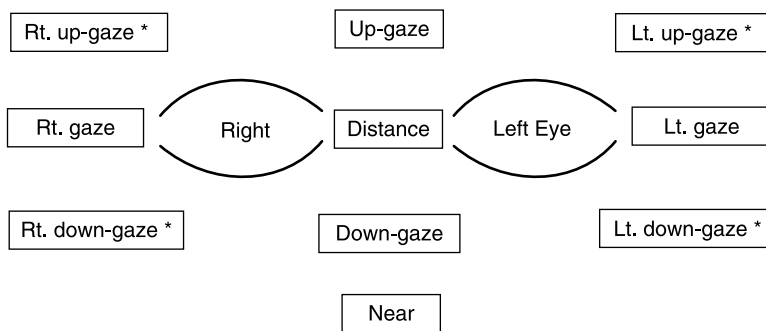


Figure 1. 10-position graphic representation of strabismus: positions marked \* are absent in the 6-position subset [5].

Table 1. Classification of vertical strabismus.

Hypertropic (upward deviation)		Hypotropic (downward deviation)	
StrabNet class	Type	StrabNet class	Type
1	Unilateral superior oblique palsy with inferior oblique over-action	5	Superior oblique tendon sheath syndrome
2	Bilateral superior oblique palsy	6	Superior oblique tendon sheath syndrome plus
3	Unilateral superior oblique palsy with lax superior oblique tendon	7	Thyroid eye disease (inferior rectus involvement)
4	Longstanding superior oblique palsy	8	Orbital floor blow-out fracture

five log-sigmoid nodes. Net 1 had a single binary hard-limiting output (*viz* either Hypertropia or Hypotropia). Nets 1 and 2 had four linear outputs from a four node log-sigmoid layer corresponding to StrabNet Classes 1–4 and 5–6 for Hypertropia and Hypotropia, respectively. A *winner-takes-all* strategy was used by employing a *post hoc* 4-input-one-output competitive (*voting*) layer.

An artificial data-set of 400 PCT records spanning the range  $\pm 40\Delta D$  as 8 groups of 50 (corresponding to StrabNet Classes 1–8) was generated by two clinical experts (a consultant ophthalmologist and an experienced orthoptist) in consensus using the graphical strabismus representation (see Figure 1). The data-set was divided randomly into four sets (a *group* corresponding to a single StrabNet diagnostic class [1, 2...8]) as follows:

- (1) training set: 160, as 8 groups of 20
- (2) validation-in-training set: 120, as 8 groups of 15
- (3) validation-and-test set: 120, as 8 groups of 15.

All data were represented as *involved right eyes*. Left eye data were simply laterally inverted prior to analysis.

All networks were trained using the Levenberg–Marquardt back-propagation (L–Mbp) algorithm available in the MatLab Neural Network Toolbox. L–Mbp was chosen in preference to the alternative gradient descent algorithms as it converges efficiently, its requirement for large memory being readily accommodated by modest modern PCs. Net 1 was trained with the training set arranged as a Hypertropic subset (80, as 4 groups combined) and Hypotropic subset (80, as 4 groups combined). Similarly, two validation-in-training subsets were constructed each of 60 as 4 groups combined. Learning was monitored at intervals of 20 iterations and stopped if the accuracy tested against the validation-in-training subsets  $\geq 96\%$  (accuracy defined as percentage of correct differentiations into Hypertropic/Hypotropic classes: 96% represents  $\sim 58$  correct from 60).

Net 2 was trained with the Hypertropic subset of training data (80, as 4 Hypertropic groups of 20). The outputs (*i.e.* StrabNet Classes 1–4) were represented as a 4-element vector: *e.g.* Class 2, Bilateral Superior Oblique Palsy as [0 1 0 0], and Class 4, Longstanding Superior Oblique Palsy as [0 0 0 1] *etc.* Learning was monitored at intervals of 50 iterations and allowed to proceed until an accuracy was achieved with the validation-

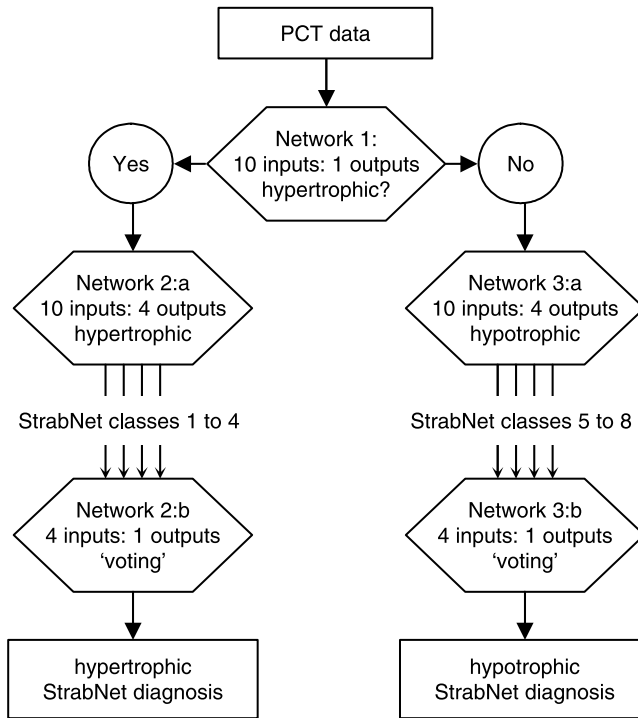


Figure 2. Arrangement of the three networks used in StrabNet to perform the 1-of-8 differential diagnoses of vertical strabismus from the 10-position PCT record.

in-training subsets of  $\geq 97\%$  *i.e.* at least 55 correct diagnoses from 60. Net 3 was trained for Hypotrophic data similarly to Net 2.

This method of training until a given target performance in the validation-in-training is achieved is known as *early stopping* and is employed to restrain the learning progress from accommodating the learning data specifically, or becoming *over-trained*. This way the network retains the property of *generalisability* and has classification potential that is useful across a range of *as yet not seen* data rather than just the specific (a unique sub-sample of all theoretically available PCT data) details of the learning data-set. Unfortunately, this technique can limit the eventual performance of the classifier: *generalisability* is at the expense of ultimate accuracy *wrt* the training and validation-in-training data-sets. As the convergence of a network proceeds from an initial random and undefined state, it is unlikely that two training runs will ever produce exactly the same network with exactly the same balance between accuracy and generalisability. In this study, 10 candidate versions of Nets 1–3 were derived and the best selected on the basis of performance with the validation-and-test data-set as follows.

The accuracy achieved by each candidate net was determined using validation-and-test data. The data-set was perturbed by the addition Gaussian random noise of incrementally increasing amplitude to PCT values and assessing the accuracy achieved with data from the validation-and-test data-set. The *best* candidates for Net 1, Net 2 and Net 3 were identified as the most robust under this operation. The most robust solutions were found to achieve an accuracy of 100% for the entire ensemble data-set (400 examples).

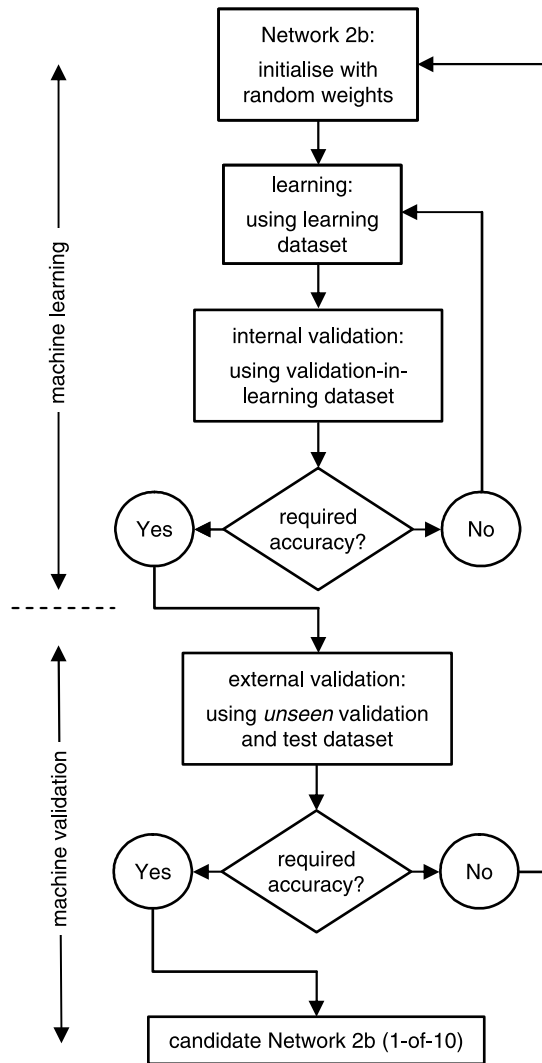


Figure 3. Learning strategy for hypertropic classification using Network 2. Learning proceeds until a relatively modest level of accuracy is *just achieved* for the validation-in-training data-set: this favours the *generalisability* of the classification. Initially, 10 network instances (candidates) are found, the best is then identified by its robustness to PCT validation-and-test data artificially contaminated by random noise. In principle, this selects the most generalized solution. Network 3 is derived and optimized for hypotropic classification by the same method.

### 2.1. Clinical validation of the 6-position StrabNet model

The output space of the expert system was constructed as one of two sets (corresponding to hypertropic and hypotropic deviations) each of 1-of-4 of the 8 classes described in the StrabNet Graphical Strabismus representation. This rigid classification is relaxed in the clinical reduced 6-position model to allow for the *not classifiable* case. This is achieved by modifying the winner-take-all strategy by setting the constraint that a winning margin must exceed a threshold. Any potential classification failing to reach this threshold is assumed equivocal and returned as ‘no diagnosis found’. This threshold was set at the 4th

percentile *viz* 2-from-50 estimated over a bootstrap (with replacement) series of learning trials. Informally, this implies a false-alarm rate of 0.04 interpreted on the basis of the training data being statistically entirely representative: this assumption must be treated with caution but is pragmatically useful.

Two clinical data-sets were assembled: Clinical Data-set 1 ( $n = 41$ ) from historical patient records and Clinical Data-set 2 ( $n = 63$ ) prospectively. These data-sets included a number of atypical complex vertical deviations which do not correspond to StrabNet Graphical Strabismus classes and had not been represented in the artificial training data-sets. The concurrence of StrabNet with clinical experts was determined. Clinical Data-set 1 was analysed by co-authors AC and IPC, whereas Clinical Data-set 2 was analysed by an independent consultant orthoptist who had had no previous involvement with the StrabNet project.

## 2.2. Implementation

A web-based version [3] (Figure 4) of StrabNet running under the JavaScript-enabled MS IE 5+ browser on an MS Windows PC platform is maintained at [www.strabnet.com](http://www.strabnet.com) [4] and is publicly available. This implementation uses a novel client/server system referred to as MatSOAP (see [www.matsoap.org.uk](http://www.matsoap.org.uk)). This enables the User at a client browser to access the resources of a full MatLab instance in a maintained server-side environment

Figure 4. StrabNet input GUI at [www.strabnet.com](http://www.strabnet.com) [3] based on the standard graphical representation of strabismus (see Figure 1). Prism values from the PCT are either entered *manually* using drop-down boxes for left or right ‘involved’ eyes or *automatically* as complete exemplar examinations from an on-line library. The response is the differential diagnosis corresponding to StrabNet Classes 1–8. Other pages on the site provide a description of the standard PCT examination and details the eight differential diagnoses. (The status of the GUI in the above example is set to the 6-position subset PCT: note corner-most positions are disabled.)

using SOAP XML messaging and automation services. This approach is novel and details will be published elsewhere. The immediate advantage of the MatSOAP implementation for StrabNet is (over conventional cgi-based systems) is provision of a Java Script rich client (GUI: graphical user interface) and high speed execution. A deprecated cgi version of StrabNet ('Web\_StrabNet') is available to the interested Reader.

### 3. Results

StrabNet achieves diagnostic accuracies of 100 and >94% with the original artificial data of 400 PCT measurement sets for the 10-position and 6-position subset PCT's, respectively.

The results of the two clinical studies are shown in Figure 5. Here concurrence is defined as the one-to-one matching of a StrabNet classification with that of the clinical experts. A *disagreement* reported by StrabNet can be either a misclassification or a 'no diagnosis found' response. Figure 5 sub-classifies disagreements as arising from either typical PCT measurement sets unequivocally belonging to the set of StrabNet Graphical Strabismus classes [1...8] or complex atypical sets of a type not included in the training set. The misclassification rate for Clinical Data-set 1 was zero per 41 for an accuracy of ~82%, and for Clinical Data-set 2, 1 per 63 for an accuracy of ~84%. The mean accuracy

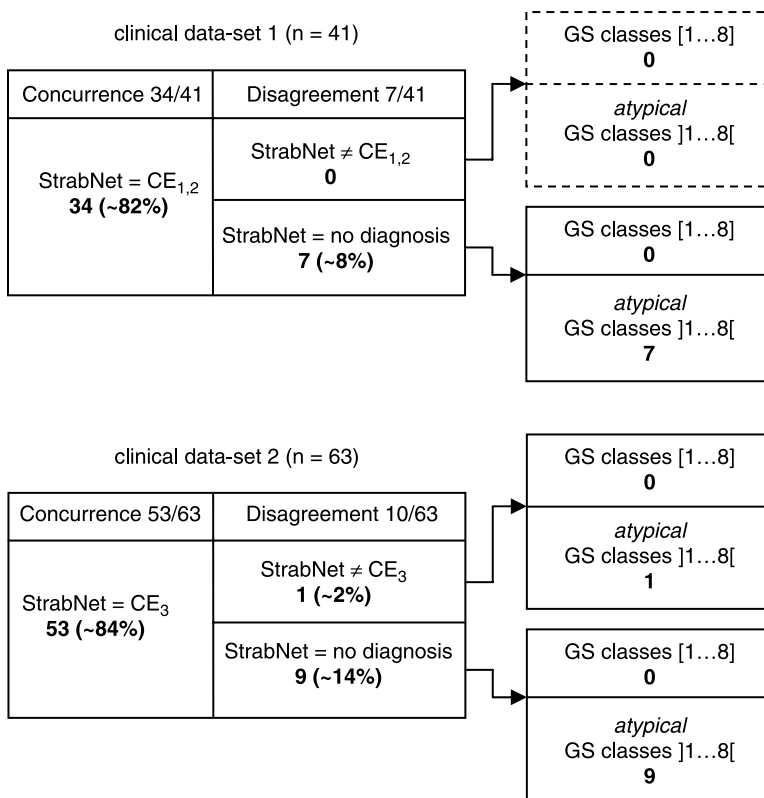


Figure 5. Concurrence of StrabNet diagnoses with clinical experts for two independent sets of clinical data. upper: clinical experts AC and IPC in consensus (C<sub>1,2</sub>). lower: independent clinical expert (consultant orthoptist, C<sub>3</sub>). Key: GS graphic strabismus representation (8 classes).

across both clinical data-sets was  $\sim 83\%$  and the false classification rate was 1%. Of 17 *atypical* measurement sets, 16 were correctly returned as ‘no diagnosis found’ which is consistent for the amorphous group of those not included in the training data-set. There was one misclassification of an atypical PCT measurement set.

#### 4. Conclusions and discussion

There have been some previous attempts at models of computer-aided investigation of strabismus [6–8]. However, these are underpinned by a parametric biomechanical model and do not employ real clinical data in the form of measurements commonly obtained in clinical circumstances for strabismus and are not implemented in the form of an expert system. The approach adopted in the StrabNet expert is straightforward and achievable using readily available computing resources. The use of the simple MLP-based architecture and the Levenberg–Marquardt back-propagation method is justified by its ease of implementation and its diagnostic performance established in synthetic and real clinical data-sets, no assumption is required to be made *a priori* as to linearity or complexity. The use of the MatSOAP resource provides user accessibility without the requirement for local programming environments other than a basic web browser client.

For the clinical data-sets it is demonstrated that decreasing the number of input covariates from 10 to 6 has only a small deleterious effect on classification accuracy. This implies that the 10-position measurement set includes covariates which are non-orthogonal, and as such, the input space is over-determined. The intriguing prospect is that the input space dimensions could be further reduced with little effect on the clinical performance. This would be extremely valuable in the clinical situation where patient compliance is often an issue. In the present study, the design of the 6-position model had the benefit of *a priori* identification which of the six possible 10 positions are *cardinal*. Training an expert system based on an architecture of ANN-MLP’s, as in the present embodiment of StrabNet, for an arbitrary number and combination of gaze positions in-real-time is not practical. Future work will investigate solutions to this problem. One candidate approach is that of the generalized linear model with Bayesian preconditioning. Such a model will converge very quickly in training (*wrt* ANN-MLP’s) but will be limited by not accommodating non-linearities: however, this might not be relevant.

Finally, it is important to emphasize that the Graphic Strabismus representation of strabismus as employed in the present development of StrabNet considers only the most common and major vertical deviations. There are indeed many subtle and more complex presentations which cannot in general be represented by measurements made in the PCT. These appeared in the present study as *atypical* PCT measurements sets in the two clinical data-sets. However, there is no obvious reason to suggest that the input dimensions of StrabNet could not be increased to accommodate these.

#### References

- [1] A.M. Ansons and H. Davies, *Diagnosis and Management of Ocular Motility Disorders*, Blackwell Science, Oxford, 2001.
- [2] A.C. Fisher, A. Chandna, and I.P. Cunningham, *The differential diagnosis of vertical strabismus from prism cover test data using an artificially intelligent expert system*, Med. Biol. Eng. Comput. 45 (2007), pp. 689–693.
- [3] Internet resource, *Web\_StrabNet: Expert System of Strabismus Differential Diagnosis at URL* www.strabnet.com, A.C. Fisher and A. Chandna (October 2008).



- [4] Internet resource, *ibid* link About the PCT.
- [5] A. Jampolsky, *A simplified approach to strabismus diagnosis, in symposium on strabismus*, in *Transactions of the New Orleans Academy of Ophthalmology*, H.M. Burian, et al., eds., Mosby, St Louis, MO, 1971, pp. 34–912.
- [6] D. Kault, D. Stark, and K. Stark, *An automated system of strabismus management. A direct approach using a 'reverse' model*, *Invest. Ophthalmol. Vis. Sci.* 30 (1987), pp. 276–287.
- [7] J.M. Miller and J.L. Demer, *Clinical applications of computer models for strabismus*, in *Clinical Strabismus Management*, A. Rosenbaum and A.P. Santiago, eds., Saunders, Philadelphia, PA, 1999, ISBN: 0-7216-7673-1.
- [8] J.M. Miller and D.A. Robinson, *A model of the mechanics of binocular alignment*, *Comput. Biomed. Res.* 17 (1984), pp. 436–470.