ORIGINAL ARTICLE

Screening patients with sensorineural hearing loss for vestibular schwannoma using a Bayesian classifier

Nouraei, S.A.R.,* Huys, Q.J.M.,† Chatrath, P.,* Powles, J.,* & Harcourt, J.P.*

*Department of Otolaryngology, Charing Cross Hospital, London, and †Gatsby Computational Neuroscience Unit, Queen Square, London, UK

Accepted 15 March 2007
Clin. Otolaryngol. 2007, **32**, 248–254

Objectives: Selecting patients with asymmetrical sensorineural hearing loss for further investigation continues to pose clinical and medicolegal challenges, given the disparity between the number of symptomatic patients, and the low incidence of vestibular schwannoma as the underlying cause. We developed and validated a diagnostic model using a generalisation of neural networks, for detecting vestibular schwannomas from clinical and audiological data, and compared its performance with six previously published clinical and audiological decision-support screening protocols.

Design: Probabilistic complex data classification using a neural network generalization.

Settings: Tertiary referral lateral skull base and a computational neuroscience unit.

Participants: Clinical and audiometric details of 129 patients with, and as many age and sex-matched patients without vestibular schwannomas, as determined with magnetic resonance imaging.

Main outcome measures: The ability to diagnose a patient as having or not having vestibular schwannoma.

Results: A Gaussian Process Ordinal Regression Classifier was trained and cross-validated to classify cases as 'with' or 'without' vestibular schwannoma, and its diagnostic performance was assessed using receiver operator characteristic plots. It proved possible to pre-select sensitivity and specificity, with an area under the curve of 0.8025. At 95% sensitivity, the trained system had a specificity of 56%, 30% better than audiological protocols with closest sensitivities. The sensitivities of previously-published audiological protocols ranged between 82–97%, and their specificities ranged between 15–61%.

Discussion: The Gaussian Process ORdinal Regression Classifier increased the flexibility and specificity of the screening process for vestibular schwannoma when applied to a sample of matched patients with and without this condition. If applied prospectively, it could reduce the number of 'normal' magnetic resonance (MR) scans by as much as 30% without reducing detection sensitivity. Performance can be further improed through incorporating additional data domains. Current findings need to be reproduced using a larger dataset.

A vestibular schwannoma is a benign nerve sheath tumour which most commonly arises from the Schwann cells of the vestibular division of the eighth cranial nerve.[1] It has a reported incidence of 1 in 100 000 and grows at a slow mean rate of approximately 1.2 mm/year. Many lesions reach a static size without surgical intervention and a small proportion of tumours may spontaneously regress.[1] Conversely, tumour growth at the cerebello-pontine angle can lead to potentially life-threatening neurological complications (Fig. S1),[2] and furthermore, when surgery is indicated, excision of a smaller tumour is associated with less postoperative morbidity.[2] This is therefore a diagnosis that once suspected, should be secured or satisfactorily discounted.

The majority of patients with a vestibular schwannoma present with asymmetrical sensorineural hearing loss,[3] but in terms of the overall number of otolaryngology consultations for the evaluation of audiovestibular symptoms, this diagnosis remains an uncommon cause of a very common presentation. Indeed, as many as one in five of all patients presenting to general ENT clinics have symptoms which could be considered compatible with the diagnosis of vestibular schwannoma.[4] This presents the otolaryngologist with the difficult diagnostic and medicolegal conundrum of deciding which of the many patients evaluated for audiovestibular symptoms are at higher risk of harbouring a vestibular schwannoma, and should be

Correspondence: Dr Reza Nouraei MA (Cantab) DO-NHS, Department of Otolaryngology, Charing Cross Hospital, London W6 8RF, UK, Tel.: 0044 7841 124610; fax: 0044 870 4580775; e-mail: RN@cantab.net.
Presented at the Royal Society of Medicine, London, UK (March 2006)
Poster presented at the Triological Society, Marco Island, Florida, USA (February 2007)
Dr Nouraei and Dr Huys contributed equally to the manuscript.

further investigated with MRI. The need for careful patient selection arises from the fact that MR does not at this time lend itself to indiscriminate screening of large numbers of patients, and furthermore, vestibular schwannoma as a condition does not fulfil the World Health Organization criteria for population level screening.[5] As a result, there has been much interest in developing sensitive and robust protocols to aid in the identification of the at-risk patient group from amidst the large population of patients with audiovestibular symptoms, so that they may be selectively referred for MRI. Pure tone audiometry, and the degree of hearing threshold asymmetry has been used to achieve this objective, and a number of audiometric decision-support protocols, with different levels of sensitivity and specificity have been described and implemented.[6–10]

The aim of the present study was to evaluate whether machine-learning technology can further improve the sensitivity and specificity of identifying patients at risk of harbouring a vestibular schwannoma based on clinical and audiological data at presentation. Machine-learning, of which Bayesian classifiers are an example, is a powerful method for classifying and separating complex data. At the beginning of the process, the 'untrained' system is provided with large numbers of cases, where both the input, in this case age and sex of the patient, presence of unilateral tinnitus or episodic vertigo and pure tone audiogram thresholds of each ear, and the outcome, the presence or otherwise of a vestibular schwannoma are known. In this way the machine 'learns' and discerns complex patterns of input–outcome association and adapts its internal algorithms, a process that is referred to as training. The 'trained' system thus acquires the ability to 'predict' whether a particular pattern of input variables, such as the clinical and audiological data from a patient with audiovestibular symptoms, is likely to be associated with an underlying vestibular schwannoma. At this point, the sensitivity and specificity of the system as a tool of diagnosis can be assessed in the same way as for any diagnostic test with receiver operator characteristic plots.

This powerful approach to complex data classification has been used in many areas of economic, engineering and biomedical disciplines, and with increasing frequency in clinical medicine for making diagnoses and predicting outcome. It has for instance been used to predict survival following treatment of laryngeal cancer more accurately than the standard prognostic models.[11]

A secondary objective of the present study was to evaluate the sensitivity and specificity of a number of previously-published decision-support protocols, comparing them against the performance of the Bayesian classifier, but also against known and accepted standards of care for the sensitivity of screening tests in comparable areas of medical practice.

Methods

Patient selection

Records of 129 patients with a proven diagnosis of vestibular schwannoma (VS^+) based on MR scans, and an equal number of patients in whom this diagnosis was suspected, but excluded on MRI (VS^-) were reviewed. Variables used for analysis included pure tone thresholds at 250 Hz, 500 Hz, 1 kHz, 2 kHz, 4 kHz and 8 kHz at the time of presentation, as well as patient age and sex, and the presence or absence of unilateral tinnitus or episodic vertigo.

Data analysis

A Gaussian Process Ordinal Regression Classifier was used to predict the likelihood of a patient having or not having a vestibular schwannoma on the basis of clinical and audiological data. This technology is a powerful Bayesian classification method, which is based on a Gaussian process. It is a generalisation of the neural network technology, operating at the limit of infinitely many hidden units.[12,13] This means that unlike finite neural networks, a Bayesian classifier makes no assumptions about, and is not therefore restricted by the inherent complexities and interactions within the dataset, and can therefore classify any dataset. In this study, following standard machine learning procedure, we randomly spliced the data into ten subsets. The train-test model was used to ensure that the performance of the system was tested on data which had not been used for training purposes, specifically to avoid the situation where a model based on the whole dataset has excellent diagnostic performance within the dataset, but bad generalisation to new data with unknown outcomes. All permutations of the subsets were used to train and test the classifier, with nine subsets being used for training. After training, the model was used to predict the probability of the presence of a vestibular schwannoma on the remaining subset. To produce the receiver operator characteristic, a threshold was chosen. If the probability was greater than threshold, a vestibular schwannoma was predicted. These predictions were compared against the true outcome, and varying the threshold from zero to unity produced the plot.

Evaluation of existing protocols

We tested the sensitivity and specificity of six published vestibular schwannoma audiological screening protocols.

We were unable to test the protocol proposed by the American Academy of Otolaryngology – Head and Neck Surgery, as it requires the hearing thresholds to be tested at 3 kHz, which we did not routinely do. The audiological, and when required, clinical data were tested against each protocol, with the outcome being whether or not the protocol recommended MRI. A recommendation of MRI in a patient subsequently proven as not having a vestibular schwannoma was a false positive, while a 'no scan' recommendation in a patient with a vestibular schwannoma was a false negative, and so on. The sensitivity and specificity of the different protocols were plotted on the receiver operator characteristic plot. The specificity of the each protocol was compared against the specificity of the Bayesian model at the protocol's corresponding sensitivity level using chi-square statistic.

Results

General

Clinical, audiological and MRI details of 258 patients were reviewed. There were 130 males and 128 females and the average age at presentation was 53 ± 15 years (± SD; range 16–97). On direct questioning, 121 patients (46%) reported unilateral tinnitus and 79 (30%) episodic vertigo. There were significant overlaps in the degree of hearing threshold asymmetry between patients with and without a vestibular schwannoma across all frequencies (Fig. S2). Averaging the degree of hearing threshold asymmetry across all frequencies reduced but did not eliminate overlap between the two populations (Fig. 1).

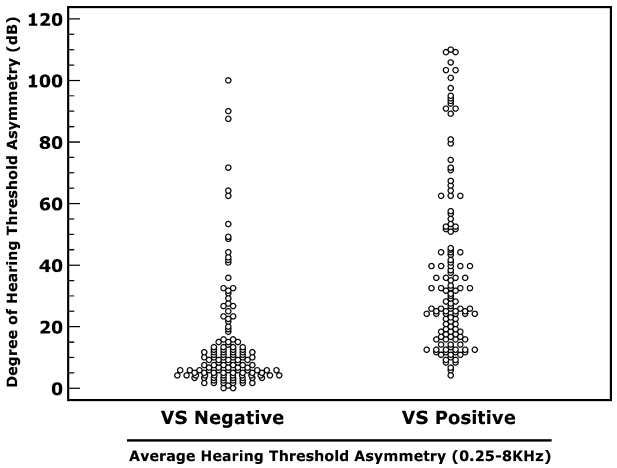

Fig. 1. Distribution of the average hearing threshold asymmetry across all measured frequencies between patients with and without vestibular schwannoma (VS).

The overall performance of the Gaussian Process Ordinal Regression Classifier

This is shown in Fig. 2. Area under the curve and d' were 0.8025 and 1.2029, respectively. The model significantly outperformed three of the protocols, and its specificity was comparable with a further three at their corresponding sensitivity levels. Table 1 shows the clinical and audiological criteria of the different protocols.

Discussion

Background to the present study

In this preliminary study a Gaussian Process Ordinal Regression Classifier[12,13] was used to differentiate between patients with and without vestibular schwannoma based on clinical and audiological data collected during a routine otolaryngological consultation. Vestibular schwannomas produce patterns of audiovestibular symptoms and in particular asymmetrical sensorineural hearing-loss which tend to be different between patients with and without a vestibular schwannoma (Fig. 1 and Fig. S2). This has been used to develop a number of different audiological screening and decision-support protocols to assist the otolaryngologist in risk-stratifying patients with audiovestibular symptoms, to selectively refer those at higher risk of harbouring a vestibular schwannoma for MRI (Table 1).[6–10] We hypothesised that machine learn-

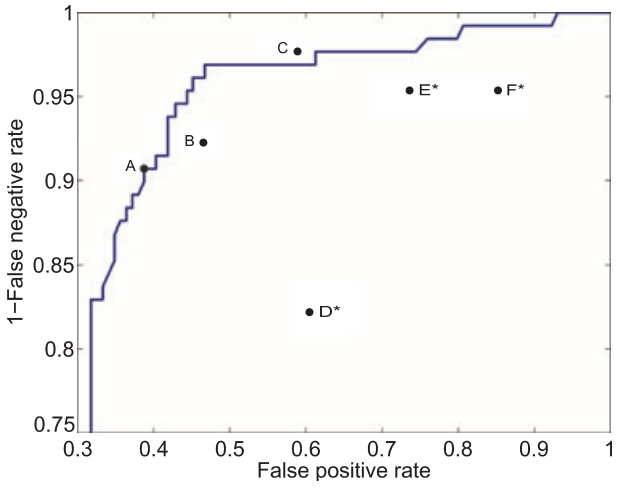

Fig. 2. Receiver operator characteristic plot of the Gaussian Process Ordinal Regression Classifier. Area under the curve and d' were 0.8025 and 1.2029, respectively. A, Seattle Protocol;[7] B, Charing Cross Protocol;[8] C, Nashville Protocol;[10] D, Oxford Protocol;[9] E, UK Department of Health; F, Sunderland Protocol.[6] *($P < 0.05$; chi-square).

Table 1. A description of published audiological screening protocols for vestibular schwannoma

	Name	Clinical criteria	Asymmetry of thresholds
A	Seattle[7]	Nil	≥ 15 dB between the average of 1–8 KHz
B	Charing Cross[8]	Nil	≥ 15 dB in any two neighbouring frequencies if the average hearing threshold of the better ear ≤ 30 dB≥ 20 dB in any two neighbouring frequencies if the average hearing threshold of the better ear >30 dB
C	Nashville[10]	Nil	≥ 15 dB at any frequency between 0.5 and 4 KHz
D	Oxford[9]	Unilateral tinnitus, age < 70 years	≥ 15 dB between the average of 0.5 to 8 KHz
E	UK Department of Health	Vertigo	≥ 20 dB at any frequency between 0.5 and 4 KHz
F	Sunderland[6]	Unilateral tinnitus or Ménière's disease symptoms	≥ 20 dB between two neighbouring frequencies

ing could improve the sensitivity and specificity of this risk-stratification process.

A brief description of Bayesian statistics

A Gaussian Process Ordinal Regression Classifier is a Bayesian tool for classifying high-dimensional data and a generalisation of neural network technology.[12,13] In general terms, Bayesian inference involves a *likelihood function* and a *prior distribution*, which are combined to form the *posterior distribution*. The likelihood function is the observation model and specifies the relative likelihood of data, given a set of *statistical parameters*. The prior distribution on the other hand embodies pre-existing knowledge or assumptions about the statistical parameters. The posterior distribution over these parameters takes both the data and the prior assumptions into account, arriving at a fully probabilistic prediction over the distribution of parameters. Inferring a distribution over parameters means that they are treated as random variables and this allows for the propagation of uncertainty through all parts of the model to arrive at a consistent predictive distribution, rather than just a point prediction.[14] A general introduction to Bayesian statistics can be found in an article by Gurrin *et al.*[15] and the current Gaussian process ordinal regression model have been further explained mathematically in the appendix.

Synopsis of key findings

The results of the present study indicate that this methodology provides significant flexibility and favourable combinations of sensitivity and specificity for differentiating between patients with and without vestibular schwannoma based on data collected during a routine otolaryngological consultation. It significantly outperformed three of the previously-published audiological screening protocols at their corresponding levels of sensitivity and specificity. Moreover, setting the sensitivity level to 95%, it reduced the number of false positive MR scans by almost 30%, when compared with audiological protocols with closest sensitivities to 95% (Fig. 3).

Comparing vestibular schwannoma with carotid artery stenosis

The sensitivity level of 95% was chosen because it corresponds with that of MR angiography for diagnosing carotid artery stenosis, requiring surgical intervention

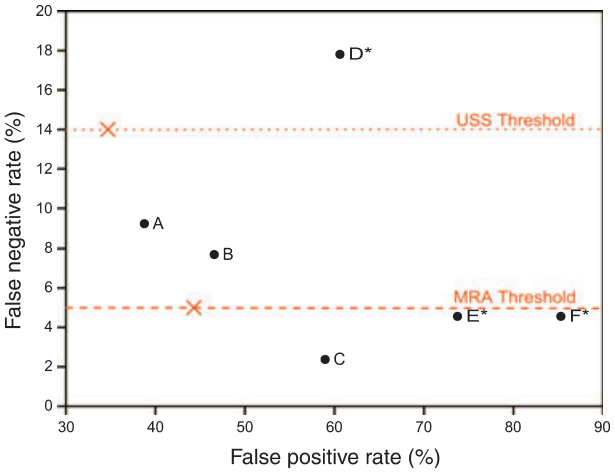

Fig. 3. A comparison of the Gaussian Process Ordinal Regression Classifier with existing audiological screening protocols. A, Seattle Protocol;[7] B, Charing Cross Protocol;[8] C, Nashville Protocol;[10] D, Oxford Protocol;[9] E, UK Department of Health; F, Sunderland Protocol.[6] The threshold levels (95% and 86%) correspond to the pooled sensitivity of MR angiography (MRA) and carotid ultrasonography (USS) for detecting significant carotid artery stenosis on meta-analysis.[16] The crosses on the threshold lines correspond to the specificity of the system at those levels.

on meta-analysis.[16] We believe that comparing vestibular schwannoma with carotid artery stenosis is clinically appropriate. Carotid stenosis, a potentially preventable cause of embolic strokes[17] is an uncommon cause of a constellation of common and often non-specific symptoms, occurring in less than 5% of patients with compelling symptoms and signs of transient ischaemic attacks.[18] Like a vestibular schwannoma therefore, it is a diagnosis that needs to be considered and excluded in a large number of patients. A cerebrovascular physician assessing patients with possible transient ischaemic attack employs carotid duplex ultrasonography, in particular, the Doppler blood velocity within the carotid arteries to risk-stratify patients for further investigation,[19] in the same way that an otolaryngologist may use hearing threshold asymmetry to risk-stratify patients for vestibular schwannoma. The sensitivity and specificity of carotid duplex ultrasonography varies depending on what Doppler velocity threshold is chosen,[19] but the pooled sensitivity of duplex ultrasonography on a meta-analysis has been found to be 86%.[16] Figure 3 provides further information about sensitivities and specificities of different protocols.

It is very reassuring to note therefore, that the sensitivities of a number of current audiological screening protocols for risk-stratifying patients with suspected vestibular schwannoma already exceed those of carotid duplex ultrasonography for risk-stratifying patients as high-risk or low-risk for stroke due to significant carotid artery stenosis. The high sensitivities that can be achieved with protocols like the Nashville come however at the expense of high false positive rates (Fig. 3) which has significant resource implications.

Strengths of the present study

Although, given the significant overlap between patients with and without vestibular schwannoma (Fig. 1 and Fig. S2), the Gaussian Process Ordinal Regression Classier cannot also fully separate the two populations, it can significantly reduce the number of false positive MR scans, compared with current audiological protocols, while maintaining detection sensitivity. Furthermore, it allows units to pre-determine sensitivity and specificity levels, which are most suited to their local policy and resource availability, which might be particularly helpful in countries where MRI resources are not so widely available. An important consideration in the development of audiological protocols is ease of application in clinical settings. As such, by necessity, they discard information. Furthermore, they are applied to data that is non-linear in respect of the underlying noise as well as the superimposed signal. To address these, some protocols, like the Nashville[10] recommend scanning in the presence of small single-frequency asymmetry which produces good sensitivity but poor specificity. Other protocols, like the Oxford,[9] are based on averaging across a range of frequencies and are unable therefore to detect tumours whose compression of the auditory nerve leads to changes in a limited number of frequencies. Machine-learning application used in this study makes use of the available information in its entirety. We tested other technologies such as Support Vector Machines which do not look at the entire dataset, but the most discriminating features within it, and obtained inferior results to those presented. Furthermore, during the training process, the Bayesian classifier tests out different possible interactions within and between the data for their ability to differentiate between cases and controls.

Clinical applicability

The trained system can be exported as software code, and the resulting computer program can be used as a decision-support tool in clinical practice as a desktop or Internet-based application in a conceptually similar way to the ADJUVANT! SYSTEM (http://www.adjuvantonline.com). The latter is an Internet-based program which was developed from the San Antonio oncology database. It allows the clinician to obtain individualised information about the prognosis and the likely efficacy of postoperative chemotherapy in an individual patient with cancer, based on the clinical and pathological variables entered through the web site or onto a PERSONAL DIGITAL ASSISTANT. It is our eventual intention to develop a non-commercial Internet-based application, where clinicians could enter the clinical details and hearing thresholds of their patients via a web page, and obtain a 'scan' or 'no scan' recommendation, to aid in the evaluation of patients with audiovestibular symptoms.

Limitations of the study

For this objective to be achieved however the findings of the present preliminary study need to be replicated and further tested for generalisability using much larger datasets. We are currently in the process of collecting more cases and would warmly welcome a collaborative approach. Another important limitation of this and other studies using machine-learning methodology is their cultural acceptability to medicine and by clinicians. The majority of machine learning applications are, almost by definition, 'black box' solutions whose inner workings remain inaccessible to intuition even at a mathematical

level, and yet from which clinicians are invited to obtain assistance in making important decisions about the care of their patients. This is a proposition that some clinicians regard with some scepticism. Machine-learning technology has however been extensively used in many fields including medicine and Otolaryngology[11] and their more widespread acceptability and application is something that can only be achieved as more clinicians become familiar with their scope, abilities and limitation.

Conclusions

We have shown that a Gaussian Process Ordinal Regression Classifier, and, accepting higher false positive rates, a number of current audiological screening protocols have sensitivity levels that are higher than those currently accepted to be the standard of care for excluding carotid artery stenosis.[16] What constitutes an acceptable false negative rate is a question whose answer requires active input from healthcare policymakers and funding bodies. We believe that a 'no scan' recommendation from a sensitive protocol, be that a machine-learning application or a standard audiological screening protocol, as well as absence of other sings and symptoms of cerebello-pontine angle pathology following a careful history and clinical examination satisfactorily discounts the diagnosis of vestibular schwannoma as the underlying cause of audio-vestibular symptoms. The only way to identify the small minority of patients whose lesion may elude diagnosis is by performing a blanket MR scan in all patients by the mere virtue of their presentation to an otolaryngologist. This would be a *de facto* population screening program which cannot be undertaken without an explicit policy decision on the part of healthcare funding bodies, and an attending allocation of additional radiological resources to achieve it.

Acknowledgements

The authors wish to thank Drs A. Azran, P. Dayan, Z. Gahahramani and F. Perez-Cruz for their helpful suggestions during data analysis.

Conflict of Interest

None to declare.

References

1 Yoshimoto Y. (2005) Systematic review of the natural history of vestibular schwannoma. *J. Neurosurg.* **103,** 59–63

2 Sanna M., Taibah A., Russo A. et al. (2004) complications in acoustic neuroma (vestibular schwannoma) surgery. *Otol. Neurotol.* **25,** 379–386

3 Matthies C. & Samii M. (1997) Management of 1000 vestibular schwannomas (acoustic neuromas): clinical presentation. *Neurosurgery* **40,** 1–9

4 Harcourt J.P., Vijaya-Sekaran S., Loney E. et al. (1999) The incidence of symptoms consistent with cerebellopontine angle lesions in a general ENT out-patient clinic. *J. Laryngol. Otol.* **113,** 518–522

5 Wilson J.M.G. & Jungner G. (1968) Principles and practice of mass screening for disease. In *Public Health Paper Number 34*. World Health Organization, Geneva

6 Dawes P.J.D. & Jeannon J.P. (1998) Audit of regional screening guidelines for vestibular schwannoma. *J. Laryngol. Otol.* **112,** 860–864

7 Mangham C.A. (1991) Hearing threshold difference between ears and the risk of acoustic tumor. *Otolaryngol. Head Neck Surg.* **105,** 814–817

8 Obholzer R., Rea P.A. & Harcourt J.P. (2004) Magnetic resonance imaging screening for vestibular schwannoma: analysis of published protocols. *J. Laryngol. Otol.* **118,** 329–332

9 Sheppard I.J., Milford C.A. & Anslow P. (1996) MRI in the detection of acoustic neuromas – a suggested protocol for screening. *Clin. Otolaryngol. Allied Sci.* **21,** 301–304

10 Welling D.B., Glassock M.E., Woods C.I. *et al.* (1990) Acoustic neuroma: a cost-effective approach. *Otolaryngol. Head Neck Surg.* **103,** 364–370

11 Jones A.S., Taktak A.G., Helliwell T.R. *et al.* (2006) An artificial neural network improves prediction of observed survival in patients with laryngeal squamous carcinoma. *Eur. Arch. Otorhinolaryngol.* **263,** 541–547

12 Wei C. & Ghahramani Z. (2005) Gaussian processes for ordinal regression. *J. Mach. Learn. Res.* **6,** 1019–1041

13 Rasmussen C.E. & Williams C.K.I. (2006) *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. MIT Press, Cambridge, MA

14 Jaynes E.T. (2003) *Probability Theory: The Logic of Science*. Cambridge University Press, Cambridge, MA.

15 Gurrin L.C., Kurinczuk J.J. & Burton P.R. (2000) Bayesian statistics in medical research: an intuitive alternative to conventional data analysis. *J. Eval. Clin. Pract.* **6,** 193–204

16 Nederkoorn P.J., van der Graaf Y. & Hunink M.G.M. (2003) Duplex ultrasound and magnetic resonance angiography compared with digital subtraction angiography in carotid artery stenosis: a systematic review. *Stroke* **34,** 1324–1332

17 European Carotid Trialists' Collaborative Group (1991) MRC European carotid surgery trial: interim results for symptomatic patients with severe (70–90%) or with mild (0–29%) carotid stenosis. *Lancet* **337,** 1235–1243

18 Bhatti T.S., Harradine K.L., Davies B. *et al.* (1999) First year of a fast track carotid duplex service. *J. R. Coll. Surg. Edinb.* **44,** 307–309

19 Walker J. & Naylor A.R. (2006) Ultrasound based measurement of 'carotid stenosis >70%': an audit of UK practice. *Eur. J. Vasc. Endovasc. Surg.* **31,** 487–490

Supplementary material

The following supplementary material is available as part of the online article from http://blackwell-synergy.com:

Figure S1. Radiology of a case.

Figure S2. Distribution of the degree of hearing threshold asymmetry in patients with and without a vestibular schwannoma.

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Appendix 1

A mathematical description of the machine learning process. Let $\{x^i\}_{i=1}^N$ be training data from N patients. For each patient, indexed by $i = 1...N$, x^i is a D-dimensional vector containing the data on the basis of which we would like to predict the presence of vestibular schwannoma. In this study $D = 16$ when all the data was used: auditory thresholds for each ear at six frequencies and age, sex and presence of tinnitus and vertigo. For each patient output is a binary variable $y_i \in \{0,1\}$ indicating the presence (y^i=1) or absence (y^i=0) of vestibular schwannoma.

The aim is to use a new patient's input vector x and the training data $\{y_i, x^i\}_{i=1}^N$ to probabilistically predict whether the patient suffers from a vestibular schwannoma, i.e. the probability $P(\hat{y}|x, D)$. A Gaussian Process Ordinal Regression Classifier (GPORC)[12,13] achieves this in several steps. First, GPORC partitions the real line into two parts using a logit function, i.e. for each value $f \in R$ on the real line, there is a $p(\hat{y} = 1|f)$. x is mapped onto the real line, writing $p(\hat{y}|f(x))$, i.e. rather than using the simple input x, some function $f(x)$ is used for prediction. This is related to what hidden layers achieve in neural networks and is a powerful approach.[12,13]

Rather than assuming a particular $f(x)$ however, a GPORC averages over all possible mappings, weighted by some prior distribution $p(f(x))$ and the predictive distribution of interest then becomes $p(\hat{y}|x) = \int df(x)p(y|\widehat{f(x)})$.

This prior is chosen to incorporate the evidence from the data D, writing the predictive prior given the data.

$$p(f(x)|D) = \int p(f(x)|\mathbf{f})p(\mathbf{f}|D)\mathrm{d}f$$
$$= \int p(f(x)|\mathbf{f})\frac{p(D|\mathbf{f})p(\mathbf{f})}{\int \mathrm{d}f' p(D|\mathbf{f}')p(f')}\mathrm{d}\mathbf{f}$$

where $\mathbf{f}$ is a vector, with its ith component the mapping $f(x^i)$ for each input data point x^i and y is similarly a vector with $y_i = y^i$. The second equality holds by Bayes' theorem. Let $p(D|\mathbf{f}) = {}_i p(y^i|f(x^i))$ be the likelihood of all the outcomes $\mathbf{y} = \{y^i\}_{i=1}^N$ given the input x^i given by the logit function described above. Let finally the joint distribution $p(\mathbf{f})$ of all hidden functions $f(x^i)$ for all i be a normal distribution (this is a Gaussian Process prior):[12,13]

$$p(\mathbf{f}) = N(0, \Sigma)\sum_{ij} = \exp(-\frac{k}{2}\sum_{d=1}^{D}(x_d^i - x_d^i)^2)$$

with parameter $\kappa > 0$ which is chosen during training. Then, using approximations to some of the hard integrals, we can evaluate the distribution over outcomes given the data: $p(\hat{y}|x, D) = \int \mathrm{d}f(x)p(\hat{y}|f(x)p(f(x)|D)$.

Thus, assuming a joint prior $p(\mathbf{f})$ over *hidden functions* $f(x^i)$ of the data, together with a mapping from these hidden functions onto probabilities of binary events y allows a full probabilistic data classification.

