Clinical Endocrinology (2011) **75**, 226–231 doi: 10.1111/j.1365-2265.2011.04020.x

ORIGINAL ARTICLE

Early diagnosis of acromegaly: computers *vs* clinicians

Ralph E. Miller*, Erik G. Learned-Miller†, Peter Trainer‡, Angela Paisley‡ and Volker Blanz§

**Division of Endocrinology, Department of Medicine, University of Kentucky, Lexington, KY, †Department of Computer Science, University of Massachusetts, Amherst, MA, USA, ‡Manchester Academic Health Science Centre, The Christie NHS Foundation Trust, University of Manchester, Manchester, UK and §Fachgruppe Medieninformatik, Institute for Vision and Graphics, University of Seigen, Seigen, Germany*

Summary

Background Early diagnosis of a number of endocrine diseases is theoretically possible by the examination of facial photographs. One of these is acromegaly. If acromegaly were found, early in the course of the disease, morbidity would be lessened and cures more likely.

Objectives, design, patients, measurements Our objective was to develop a computer program which would separate 24 facial photographs, of patients with acromegaly, from those of 25 normal subjects. The key to doing this was to use a previously developed database that consisted of three-dimensional representations of 200 normal person's heads (SIGGRAPH '99 Conference Proceedings, 1999). We transformed our 49, two-dimensional photos into three-dimensional constructs and then, using the computer program, attempted to separate them into those with and without the features of acromegaly. We compared the accuracy of the computer to that of 10 generalist physicians. A second objective was to examine, by a subjective analysis, the features of acromegaly in the normal subjects of our photographic database.

Results The accuracy of the computer model was 86%; the average of the 10 physicians was 26%. The worst individual physician, 16%, the best, 90%. The faces of 200 normal subjects, the original faces in the database, could be divided into four groups, averaged by computer, from those with fewer to those with more features of acromegaly.

Conclusions The present computer model can sort photographs of patients with acromegaly from photographs of normal subjects and is much more accurate than the sorting by practicing generalists. Even normal subjects have some of the features of acromegaly. Screening with this approach can be improved with automation of the procedure, software development and the identification of target populations in which the prevalence of acromegaly may be increased over that in the general population.

Correspondence: Ralph E. Miller, Division of Endocrinology, Department of Medicine, University of Kentucky, 140 Cherokee Park, Lexington, KY 40503, USA. Tel.: +859 277 0930; Fax: +859 323 5707;
E-mail: remill0@uky.edu

(Received 10 January 2011; returned for revision 18 January 2011; finally revised 11 February 2011; accepted 11 February 2011)

Introduction

Acromegaly is a disease in which excessive growth hormone, usually from a pituitary tumour, causes excessive growth of tissues over a period of many years. Most noticeable in an untreated patient are enlargement of the hands, the feet and, in the original description by Pierre Marie, the face. Classically, the brow and nose are enlarged, the zygomatic arches and jaw are prominent and the abundant facial tissue contains deep nasolabial folds. These changes take place so slowly that they are not recognized, especially by persons who see the patient frequently.

Early diagnosis of acromegaly would facilitate cures. Presently, 7–10 years pass between the onset of symptoms and diagnosis,[1–3] and, in many, complete cures are not possible. Quite recently, three, noncomputer based, screening studies have revealed higher than expected prevalences of acromegaly in three separate populations (106, 124 and 1034 per million respectively).[2,4,5]

The ultimate goal of our research is to develop an automated, photographic, voluntary disease finding system for acromegaly. The present work is a preliminary attempt to use facial photos to find unsuspected disease in humans, while it is still possible to remove most or all of the offending pituitary tumour. Ideally, facial recognition of acromegaly could be applied to a collection of photographs as well as used for screening in a clinical setting. For example, when obtaining a photograph for a driver's licence, one could voluntarily choose to be screened automatically for various conditions, such as acromegaly. If the result of the computer analysis was positive, it would be recommended that the driver see a physician for further analysis presumably including tests of the patient's blood.

There have been a number of efforts to do automatic statistical analysis of face shape.[6,7] These methods require a specialized apparatus to acquire three-dimensional information about the subject. Stereo methods[6] and full three-dimensional scans[7] are the common methods for the acquisition of three-dimensional information.

A key feature of the present work is that while the original statistical morphable model (see Glossary) required the acquisition of three-dimensional laser scans, it can be applied to other

databases, like ours, consisting only of standard two-dimensional photographs.[8,9] Although there is an enormous literature on face recognition, it is less common to use face databases to group images into categories. An example of such an application is the classification of faces as man or woman.[10] We are not aware of any previous work in trying to identify patients with acromegaly by computer examination of photographs.

A general statistical model, based on laser scans of 200 normal faces, developed by one of us (VB) and collaborators, was applied to the diagnosis of acromegaly from an examination of 49 facial photographs. The current study addresses the facial changes of acromegaly and how well they can be detected by human observers and by our computer model.

Although in this study we used a binary classification of faces as 'normal' or consistent with acromegaly, we are very interested in being able to classify faces according to a continuum from no signs of acromegaly to the full blown disease. To find whether this might be possible, one of us (RM, an endocrinologist) attempted to classify, by subjective analysis, the 200 'normal', two-dimensional faces in our database into four groups from those with few signs of acromegaly to those with more.

Finally, we asked 10 general physicians to pick the patients with acromegaly from the same 49 photographs that had been examined with the computer and then compared the accuracy of the computer analysis with that of the 10 physicians.

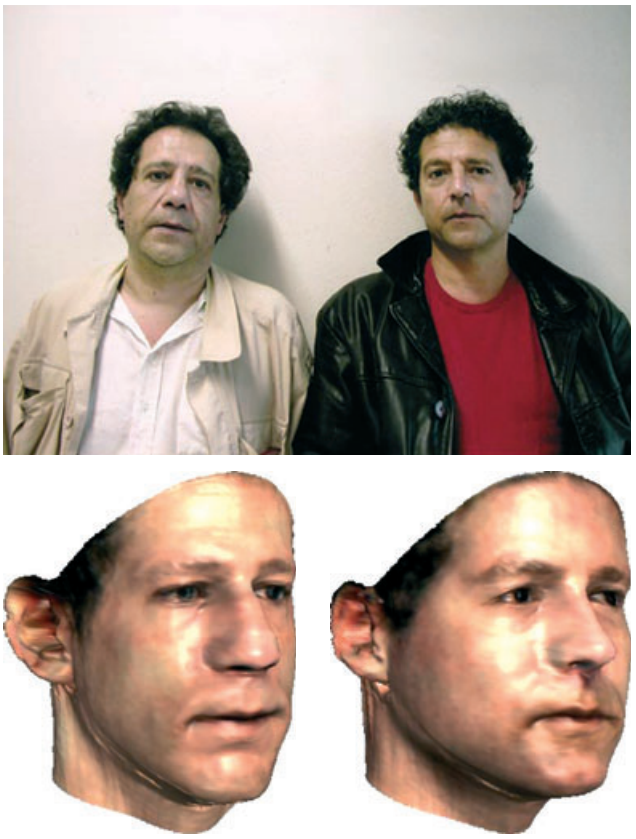

Fig. 1 These men are identical twins. Which one is sick, and what is his condition? By fitting the models of Blanz and Vetter to each of these faces, the system described in this paper correctly identified the man on the left as a patient with acromegaly and the man on the right as healthy. This 'Medical Mystery' first appeared in the New England Journal of Medicine.[12] *Lower left*: A rendering from the 3D model captured the swelling of the man's nose, a strong indicator of acromegaly which is difficult to detect as an image feature using traditional feature detectors operating on the original image. Although it is significantly more subtle, the 3D model also seems to have captured the coarseness of the man's lips, another indicator of possible acromegaly. Using our classifier on the parameters of the 3D model, we classified this man as having acromegaly. *Lower right*: A 3D morphable model was adapted in an analysis-by-synthesis loop (see the Glossary) to match the photograph of the man on the right above. After a manual initialization of the approximate pose and lighting in the photograph, the fitting on the 3D model was fully automatic. Although the model is not a perfect replica of the person in the photograph, it captures important properties of the face and skull which prove to be very effective in distinguishing between patients that do or do not have acromegaly.

Methods

Methods I

Several years ago, Blanz and Vetter developed a statistical model that directly uses the densely sampled geometry of a set of exemplar faces obtained from a cyberware™ laser scanning device (See Glossary and Ref. 8). Laser scans of the entire heads of 100 adult young women and 100 adult young men were obtained. With each subject standing upright, 512 horizontal levels of the head were defined. Five hundred and twelve vertical lines around each head were also determined. Each point of intersection was recorded as a vertex ($512 \times 512 = 262\ 144$, see Glossary). The surface of each head was thus defined by these points. A model of the three-dimensional structure of heads was developed in which the parameters (i.e. measurements) represent typical modes of shape variation in normal adults. In a process called analysis-by-synthesis, a three-dimensional estimate of a subject's head can be developed by varying the parameters of the statistical model so that a rendered version of the model approximates as closely as possible, a two-dimensional photo of the subject (see Glossary, Refs 8,9 and Fig. 1).

We were generously given direct access to the software of Blanz and Vetter and used it to develop three-dimensional models from two-dimensional photographs. The majority of the 24 patients with acromegaly were previously treated and had controlled disease. Photographs of them were taken in Manchester, UK, and photographs of 25 control subjects were taken in Lexington, KY, according to the following specified protocol:

- The camera make and model was a Nikon Cool-Pix 5000 digital camera in both locations.
- The background for the photograph: identical pieces of coloured fabric.
- The expression of the subject: relaxed, neutral expression, including a closed mouth and open eyes.
- The orientation of the patient: front facing.
- Similar general lighting conditions.

In addition to global models of shape, a secondary procedure was used in which smaller parts of the face were extracted and modelled separately.[8] This procedure gives greater detail and accuracy for the following facial regions or groups of regions: (i) the

nose, (ii) the eyes, eyebrows and brows, (iii) the mouth, and (iv) all other features such as cheeks, chin, forehead and neck.

The review boards of the University of Kentucky and Manchester, England, approved the protocol, and patients whose photographs appeared in the manuscript gave signed permission for this.

The 49 subjects were classified using a leave-one-out cross-validation paradigm (see Glossary). This involved picking a single face of a control subject or of a patient with acromegaly, modelling all of the remaining control faces, as a whole as well as the four parts, and secondly modelling all of the faces of the patients with acromegaly, in parts and as a whole. The parameters of the 'left out' single subject's face were then compared to see into which group it best fit. At the present time, part of this process requires human intervention. It is not completely automatic. We used the publicly available support vector machine (SVM) package called SVM Lite (Fig. 2).[11]

Methods II

One of us (RM, an endocrinologist), rapidly and in a global, subjective fashion, classified 200 normal, two-dimensional faces (100 men, 100 women) into one of four categories: from least like a patient with acromegaly to most like a patient with the disease. These were the same photos that were the basis for the three-dimensional computer models. The photographs in each group were then averaged by a computer.

Methods III

We showed the same set of 49, 8 × 10 colour photographs, 24 of patients with acromegaly and 25 of normal subjects, which were examined with the computer, to 10 general physicians. These physicians were internists or family practitioners who practice medicine at the University of Kentucky. They range in age from approximately 30 to 60. We asked them to pick out the photographs of the patients with acromegaly. We then compared their choices to those of the computer.

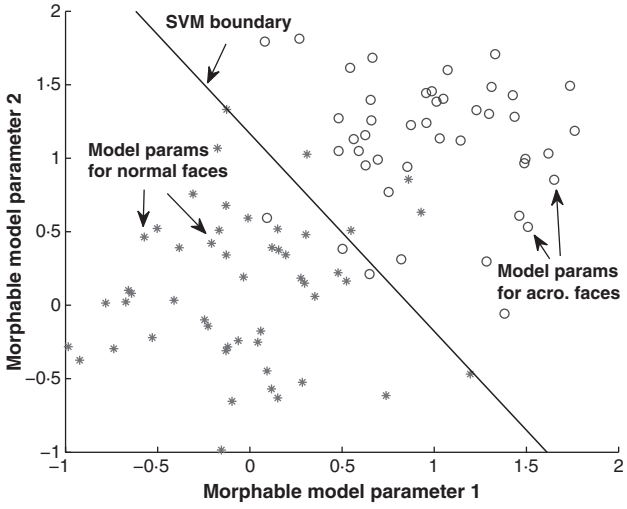

Fig. 2 This figure demonstrates how a 'support vector machine' (SVM, see Glossary) works. It shows schematically the data used in our classification experiments. Each x or round point represents the parameters (or coefficients) of the morphable model fit to each subjects face. In this illustrative example, only two dimensions were used represented by parameter 1 and parameter 2 not, the mathematically possible, 495 as in the examination of our subjects. After holding out one point (i.e. one subject), an SVM was trained using the remaining points. This resulted in a classification boundary used to classify the held out point.

Results

I The generation of three-dimensional views from two-dimensional photographs is shown in Fig. 1. Using the method described earlier, the man on the left of the photograph was diagnosed with acromegaly by the computer.[12] The separation of the faces of patients with acromegaly from those of normal subjects, using only two parameters, is shown in Fig. 2 as a very simplified example of how a SVM works (see Glossary). Using 99 parameters from a global model of each head, as well as 99 parameters from each of the four 'parts', 495 parameters in all, correct classification was achieved 85·7% of the time (Fig. 3). None of the normal faces were misclassified but seven of those of the patients with acromegaly were classified as normal; four of these are shown in Fig. 4. The sensitivity of the classification was 70·4%, the specificity 100% (See Table 1). As more parameters were used, accuracy was improved (Fig. 3).

II The examination of the four groups of computer averaged, two-dimensional faces of normal subjects, sorted subjectively according to how many subtle features of acromegaly they might have, shows a steady progression in the appearance of typical features of acromegaly (Fig. 5). There is progressive enlargement of the orbital ridges, nose, lips and jaw as one examines the categories from 1 (the least features) to category 4 (most). In general, faces went from delicate to coarse, from woman to man.

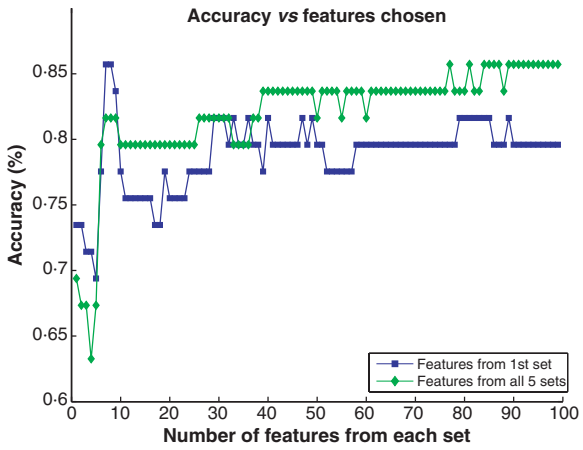

Fig. 3 This shows the leave-one-out classification accuracy (percentage of correct classifications) of a support vector machine (SVM) *vs* the number of features (parameters) of the morphable model that were used. The blue curve shows the performance using only the global model, while the green curve shows the performance using components from each of the five models (one global and four parts-based). See Glossary and Ref. 7 for details about the use of the SVM.

Fig. 4 Four patients, all with acromegaly, that the computer incorrectly classified as normal. Number 2 was diagnosed incorrectly as normal by nine physicians and number 39 by seven. In contrast, number 5 was diagnosed correctly as having acromegaly by all 10 physicians and number 38 by nine, all but one (see text for Discussion).

Table 1. Physician *vs* computer model

	Sensitivity	Specificity	Positive predictive value	Accuracy
Physicians				
Average	46	96	85	26
Worst	33	100	100	16
Best	83	96	95	90
Computer	71	100	100	86

Ten physicians, internists or family practitioners, examined 49, 8 × 10 colour photographs and decided which represented patients with acromegaly or normal subjects. The same photographs were analysed by the computer model. The computer was more accurate (86%) than all but one of the physicians (90%).

III When presented with 49 choices, general physicians missed, on average, 12·4 diagnoses. They missed 1·2 of the 25 control patients and 10·6 of the 24 patients with acromegaly. In contrast, the computer missed none of the control patients but 7 of the 24 patients with acromegaly (Table 1).

Of particular interest are a few patients with acromegaly, all of whom the computer classified, incorrectly, as normal (Fig. 4). Number 2 was not diagnosed by 9 of 10 physicians as a patient with acromegaly; #39 was not diagnosed by 7. On the other hand, #5 was diagnosed by all 10 physicians and #38 by all but one as having acromegaly.

Discussion

In the present example of screening a group of persons for acromegaly, the computer model had an accuracy of 85·7% and was a good deal better than the diagnostic accuracy of general physicians. A possible reason the physicians did not do better was because many of the patients had only subtle signs of the disease (see Fig. 4).

That a patient with acromegaly has subtle but definite characteristics was suggested first by the subjective separation of normals into four progressively coarser appearing groups in which the typical signs of acromegaly became ever more prominent and secondly, the high accuracy of one of the 10 physicians (90%).

Computers have been used in screening nonmedical phenomena for a number of years. For instance, a video camera can survey a parking lot 24 h/day to guard against theft or damage. Review of the raw film would be very time-consuming. By giving a computer rules to follow, unusual events can be sorted and placed at the start of the video, common events at the end. If a door is pried open or a window broken these events would be at the beginning, all episodes of one or a group of people simply walking to a car and getting in, at the end.[13]

Ideally medical screening is performed in a population with: (i) a high incidence of an unrecognized disease, (ii) an inexpensive, and easily applied, test for the disease, (iii) with a test which has a low false-positive rate, and (iv) for a disease for which there is effective treatment. At times, it is necessary to have a second, perhaps more expensive but more precise test, to rule out false-positive results found with the first screening test. Many programs, even though valuable, do not satisfy all of these criteria in an ideal fashion.

Screening of newborns, by blood samples, for phenylketonuria was begun in the 1960s by Robert Guthrie and others. After this test was introduced, the recorded incidence, of the disease, quadrupled.[14]

Congenital hypothyroidism was widely added to screening programs in the 1970s.[15] Later screening was performed for metabolic

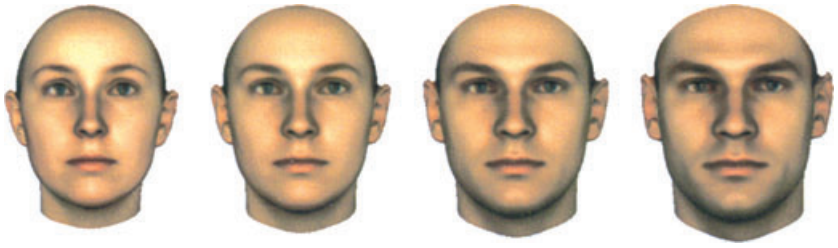

Fig. 5 Examination of the average face from each of the four groups shows a steady progression in appearance of typical features of acromegaly. Progressive enlargement of orbital ridges, nose, lips and jaw between category 1, the least, to four the most features of the disease. In general from delicate to coarse and, it appears, from woman to man.

diseases.[16,17] and congenital adrenal hyperplasia.[18] Screening adults for thyroid and adrenal disease has also been successful.[19,20] Related to visually observed physical findings, like the present work, an ultrasound screening examination of foetuses has been performed in Europe.[21]

As the prevalence of a disease in a population rises, the outcome of screening becomes more successful. In the last few years, the presumed prevalence of acromegaly has risen from about 50 per million subjects to 106, 124 and 1034 per million in three separate investigations.[2–4] To make the computer-assisted approach to screening for acromegaly efficient, specific populations with a suspected higher than average incidence of the disease need to be examined. For instance, since it is known that many patients with acromegaly have sleep apnoea, photographing of patients reporting to a sleep apnoea clinic may be more productive than examining a large number of photographs taken for automobile licences. Another factor which would make approaching larger populations more practical would be making our program fully automatic.

In the present study, patients #2 and #39 had very subtle changes, and their diagnoses were missed by most of the physicians as well as by the computer model. More troubling, why did the computer miss patient #5 and #38? The three-dimensional model of #5 was fairly accurate. It is not clear why the patient was not classified correctly. One possibility is that relative jaw size in this database was the best indicator of acromegaly, and this particular individual's signs are best described as frontal bossing and prominent cheekbones, not large jaw size. It is possible that increasing our database would allow us to classify such examples correctly. The three-dimensional representation of patient #38 was poor (data not shown). This may be because the face differs so much from a statistically common face that it is difficult for the principal component model to represent it well. That this patient with obvious acromegaly was missed by the computer requires careful follow-up and possible changes in the way in which the computer program deals with very uncommon faces.

Several aspects of the computer analysis of patients were not ideal. The patients were photographed in Manchester and control subjects in Lexington. In general, the distance from the camera to the subject was not identical, and while the patients with acromegaly were men and women of more than one race, the control subjects were all Caucasian men. Further, the control subjects were all middle aged (30–50). Older subjects may be more difficult for the computer to diagnose correctly. On the other hand, if we could use newly diagnosed untreated patients, with acromegaly, their abnormal features might be more easily recognized by the computer. We plan to improve the size and quality of our data sets and thus the accuracy of the classifier. At this point, we have found that the parameters of the computer model capture many three-dimensional features of the subjects head from just a single, two-dimensional photograph, and these features are very good clues with which to classify many members of a group as persons with or without acromegaly.

Although there were no false-positive identifications in this study, the pretest probability of finding them in screening is high. This is true because the assumed incidence (1–4/million) and prevalence (50–1000/million) is relatively low. A fairly high false-positive rate is acceptable, however, and better than a high false-negative rate, because false positives can be readily identified with an IGF-1 blood test, whereas false-negative subjects are lost after screening. Unfortunately in this study, the sensitivity of the computer analysis was 70%, meaning that the computer diagnosed 30% of the patients with acromegaly, incorrectly, as normal.

In conclusion, some subtle features of acromegaly can be found in varying degrees even in two-dimensional photographs of normal faces. Although patients with acromegaly are not numerous in the general population, their numbers may be greater in selected groups. If they are found early in the course of the disease, they can be completely cured. Computer screening of two-dimensional photographs would be an inexpensive first step towards diagnosis. If fully automated, photographs would be sent, via the internet, to a central sorting location and positive findings quickly returned to the point of origin. 'Positive' patients would be tested for elevated blood concentrations of IGF-1. It is suspected that a fully automated computer screening program would be easier to perform, and less expensive, than drawing blood for IGF-1 analysis on a large group of patients, but at this point, it is difficult to know.

The present computer model can sort photographs of patients with acromegaly from normal control subjects and was more successful than 9 of 10 practicing generalists. If perfected, the approach used in this study could be used to screen for other diseases such as Graves disease and Cushing's disease in which facial features may be suggestive of a diagnosis. The computer approach may be improved with a larger database of photographs and a fully automatic computer program. Without the latter, it might be more expensive than chemical testing and thus not practical for broad use in a general medical clinic.

Acknowledgement

No grants or fellowships supporting the writing of this paper.

Disclosure

Nothing to declare.

References

1 Melmed, S. (2006) Acromegaly. *New England Journal of Medicine*, **355**, 2558–2573.
2 Fernandez, A., Karavitaki, N. & Wass, J.A. (2009) Prevalence of pituitary adenomas: community based, cross-sectional study in Banbury (Oxfordshire, UK). *Clinical Endocrinology (Oxford)*, **72**, 377–382.
3 Freda, P.U. (2000) Advances in the diagnosis of acromegaly. *Endocrinologist*, **10**, 237–244.
4 Daly, A.F., Rixhon, M., Adam, C. *et al.* (2006) High prevalence of pituitary adenomas: a cross-sectional study in the province of Liege, Belgium. *Journal of Clinical Endocrinology and Metabolism*, **91**, 4769–4775.
5 Schneider, H.J., Sievers, C., Saller, B. *et al.* (2008) High prevalence of biochemical acromegaly in primary care patients with elevated IGF-1 levels. *Clinical Endocrinology*, **69**, 432–435.

6 D'Appuzo, N. (2003) Measurement and modeling of human faces from multi-images. *International Archives of Photogrammetry and Remote Sensing*, **34**, 241–246.
7 Hammond, P., Huttorn, T.J., Patton, M.A. *et al.* (2001) Delineation and visualization of congenital abnormality using 3D facial images. Workshop on intelligent Data Analysis in Medicine and Pharmacology at MEDINFO 2001.
8 Blanz, V. & Vetter, T. (1999) A morphable model for the synthesis of 3-D faces. SIGGRAPH '99 Conference Proceedings.
9 Learned-Miller, E., Lu, Q., Paisley, A. *et al.* (2006) Detecting acromegaly: screening for disease with a morphable model. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, **2**, 495–503.
10 Moghaddam, B. & Yang, M.H. (2002) Learning gender with support faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, **24**, 707–711.
11 Joachims, T. (1998) Making large scale SVM learning practical. LS8-Report, 24, Universitat Dortmund, LS VIII Report.
12 Neiwlaat, P. (2004) A medical mystery – which twin is the patient? *New England Journal of Medicine*, **351**, 68.
13 Grimson, W.E.L., Stauffer, C., Lee, L. *et al.* (1998) Using adaptive tracking to classify and monitor activities in a site proceedings. *IEEE Conference on Computer Vision and Pattern Recognition*, **98cB36231**, 22–31. DOI: 10.1109/cvpr 1998.698583.
14 Guthrie, R. & Susi, A. (1963) A simple phenylalanine method for detecting phenylketonuria in large populations of newborn infants. *Pediatrics*, **32**, 338–343.
15 Klein, A.H., Agustin, A.V. & Foley, T.P. (1974) Successful laboratory screening for congenital hypothyroidism. *Lancet*, **2**, 77–79.
16 Chace, D.H., Kalas, T.A. & Naylor, E.W. (2003) Use of tandem mass spectrometry for multi-analyte screening of dried blood specimens from newborns. *Clinical Chemistry*, **49**, 1797–1817.
17 American Academy of Pediatrics Newborn Screening Authoring Committee. (2008) Newborn screening expands: recommendations for pediatricians and medical home, implications for the system. *Pediatrics*, **121**, 192–217.
18 Pass, K.A. & Neto, E.C. (2009) Update: newborn screening for endocrinopathies. *Endocrinology and Metabolism Clinics of North America*, **38**, 827–837.
19 Turnbridge, W.M.G. & Vanderpump, M.P.J. (2000) Population screening for autoimmune thyroid disease. *Endocrinology and Metabolism Clinics of North America*, **29**, 239–253.
20 Findling, J.W. & Raff, H. (2005) Screening and diagnosis of Cushing's syndrome. *Endocrinology and Metabolism Clinics of North America*, **34**, 385–402.
21 Garne, E., Loane, M., Dolk, H. *et al.* (2005) Prenatal diagnosis of severe structural congenital malformations in Europe. *Ultrasound in Obstetrics and Gynecology*, **25**, 6–11.

Glossary

Support vector machine: A computer technique for separating and classifying measurements into categories.

Vertices: In this work, the shape of the head is approximated by a polyhedron with a very large number of triangles. The 'vertices' in this work are the corners of the triangles in this large set, and the triangles, in turn, are generated from the large number of axial and vertical (cranial-caudal) scans of the head which is examined by laser scanning.

Morphable Model: A computer model (program) in which vertices (points) in a three-dimensional space can be morphed (moved) in three dimensions according to the program.

Analysis-by-synthesis: The process of developing a three-dimensional representation of a head from a two-dimensional photograph of a face. The measurements made on 200 normal individuals are used. In an analysis-by-synthesis, the quality of the three-dimensional model is evaluated by analysing the two-dimensional pictures it can produce or 'synthesize'. Thus, a three-dimensional model is improved by trying to improve the images it can synthesize.

Leave-one-out cross-validation program: The object to be classified (a face) is set aside. All the remaining faces are 'modelled' in three dimensions and average measurements made. This is done with the faces of the patients that have acromegaly as well as those of the normal subjects. The 'left-out-face' is then modelled, and its measurements compared with the measurements of both groups to see into which group it best fits.