Automated identification of Fos expression

D. YOUNG

Laboratory for Statistical Neuroimaging, Mailman Research Center, McLean Hospital, 115 Mill St. Belmont, MA 02478, USA; Department of Psychiatry, Harvard Medical School, Boston, MA 02114, USA; Robertson Centre for Biostatistics, University of Glasgow, Glasgow G12 8QQ, Scotland, UK

J. MA

Molecular Pharmacology Laboratory, Mailman Research Center, McLean Hospital, 115 Mill St. Belmont, MA 02478, USA; Department of Psychiatry, Harvard Medical School, Boston, MA 02114, USA

S. CHERKERZIAN, M. P. FROIMOWITZ

Molecular Pharmacology Laboratory, Mailman Research Center, McLean Hospital, 115 Mill St. Belmont, MA 02478, USA

D. J. ENNULAT, B. M. COHEN

Molecular Pharmacology Laboratory, Mailman Research Center, McLean Hospital, 115 Mill St. Belmont, MA 02478, USA; Department of Psychiatry, Harvard Medical School, Boston, MA 02114, USA

M. L. EVANS

Neuroscience Research GlaxoSmithKline Pharmaceuticals, Harlow, UK

N. LANGE*

Laboratory for Statistical Neuroimaging and Molecular Pharmacology Laboratory, Mailman Research Center, McLean Hospital, 115 Mill St. Belmont, MA 02478, USA; Department of Psychiatry and Biostatistics, Harvard Medical School and School of Public Health, Boston, MA 02114, USA
nick@mclean.harvard.edu

SUMMARY

The concentration of Fos, a protein encoded by the immediate-early gene *c-fos*, provides a measure of synaptic activity that may not parallel the electrical activity of neurons. Such a measure is important for the difficult problem of identifying dynamic properties of neuronal circuitries activated by a variety of stimuli and behaviours. We employ two-stage statistical pattern recognition to identify cellular nuclei that express Fos in two-dimensional sections of rat forebrain after administration of antipsychotic drugs. In stage one, we distinguish dark-stained candidate nuclei from image background by a thresholding algorithm and record size and shape measurements of these objects. In stage two, we compare performance of linear and quadratic discriminants, nearest-neighbour and artificial neural network classifiers that employ functions of these measurements to label candidate objects as either Fos nuclei, two touching Fos nuclei or irrelevant background material. New images of neighbouring brain tissue serve as test sets to assess generalizability of the best derived classification rule, as determined by lowest cross-validation misclassification rate. Three experts, two internal and one external, compare manual and automated results

*To whom correspondence should be addressed.

for accuracy assessment. Analyses of a subset of images on two separate occasions provide quantitative measures of inter- and intra-expert consistency. We conclude that our automated procedure yields results that compare favourably with those of the experts and thus has potential to remove much of the tedium, subjectivity and irreproducibility of current Fos identification methods in digital microscopy.

Keywords: Amygdala; Artificial neural networks; Classification; Digital image analysis; Immediate-early gene protein; Immunohistochemistry; Microscopy.

1. INTRODUCTION

The immediate-early gene *c-fos* encodes a particular nuclear phosphoprotein and transcription factor, Fos. Immunohistochemists can stain sections of animal brain tissue to reveal Fos expression in cellular nuclei involved in the production of this protein. It is generally accepted that Fos expression in brain tissue is correlated with neuronal activity regionally and at the cellular level. A variety of stimuli, including psychoactive drugs, have been shown to up- and down-regulate Fos in unique regional patterns (Morgan *et al.*, 1987; Hunt *et al.*, 1987; Dragunow *et al.*, 1990; Deutch *et al.*, 1995). In particular, differential patterns of Fos expression in rat striatum, a subcortical part of the brain, occur post administration of antipsychotics (Nguyen *et al.*, 1992). Understanding these differential patterns of Fos expression is critical in drug development. Antipsychotic drugs are grouped into two classes: 'typical' and 'atypical'. Typical antipsychotic drugs are older medications that are efficacious yet cause deleterious side effects. Atypical antipsychotics are at least as efficacious as typicals, yet with reduced side effects. For example, reduced alteration in striatal Fos activation by atypical antipsychotics, such as clozapine, provides a direct link to observed reductions of extrapyramidal side effects, including Parkinsonian symptoms, common to typical antipsychotics such as haloperidol (Nguyen *et al.*, 1992; Robertson and Fibiger, 1992). In addition to the striatum, neurons in other brain structures such as the thalamus, prefrontal cortex and nucleus accumbens also express Fos in response to some or all antipsychotic agents and are involved in the actions of these and other drugs (Wan *et al.*, 1995; Cohen *et al.*, 1998, and in preparation).

More accurate and reliable identification of Fos-positive neurons through increased application of modern statistical techniques can aid our understanding of the complex effects of antipsychotic drugs and experimental compounds. Staining for Fos in brain slices from animals treated with a control compound or a drug requires timely, generalizable, accurate and consistent identification of participatory nuclei. Any reliable quantitative index of local and regional neuronal involvement in specific drug effects derived from these efforts would possess these desirable statistical quality control properties. Yet, somewhat surprisingly, little attention has been paid to this important issue. Typically, Fos nuclei are identified and counted by hand, a time-consuming and laborious task whose overall quality is difficult to assess statistically. Due to the subjective nature of human identification, such practice also makes comparison of data gathered by different researchers problematic. Small subsets of cells are often analysed and results attributed to an entire brain region (Robertson and Fibiger, 1992; Wan *et al.*, 1995; Ozaki *et al.*, 1998). While this approach would be valid if stained nuclei are distributed throughout the region according to a Poisson assumption of complete spatial randomness (Diggle, 1983; Diggle *et al.*, 1991), this is clearly not the case, since neurons are known to be organized according to complex yet systematic neural pathways and associations. Not only would automated identification enable neuroscientists to identify Fos expression in regions of larger relative volume, automation also offers the possibility to perform the task in a fraction of the time with significantly less human intervention and subjectivity and hence greater reliability and reproducibility.

Other researchers working in cellular imaging areas have developed automated identification procedures and assessed their results by comparison with human experts. We provide performance quality

measures reported by some of these researchers in order to place expectations for automated procedures in current context and technology through the following brief summaries. Carothers *et al*. (1983) proposed a statistical method that employed multiple-cell measurements to discriminate efficiently and without operator interaction between chromosomally normal and aneuploid ('wrong number') human cell lines. In these authors' opinion it was feasible to process 16–32 cells fairly rapidly, without user interaction and with average chromosome misclassification rates of 5–17% for all types of chromosomes; the proposed method was a potential basis for a realistic, fully automatic screening system. Vossepoel and de Kanter (1987) developed an automated chromosome analysis system to assist the expert in analysis of metaphase and prophase chromosome spreads, often a time-consuming and tedious job. To test the performance and robustness of the system, these authors used two test sets consisting of 21 absorption images and 79 fluorescence microscope images. For automatic segmentation, i.e. separating object from background, a success rate of 94% was achieved for detecting single chromosomes in the fluorescence images and 81% for the absorption images without user intervention. Garcia-Sagredo *et al.* (1994) considered automation of the scoring of sister chromatid exchanges by image analysis in a dose response experiment. Overall, about three-quarters of all exchanges were detected correctly, with a false positive rate of about 1.5% per cell. The correlation between visual and automatic scoring on a cell by cell basis was greater than 0.7 in all of the experiments conducted. Fantes *et al.* (1995) considered a chromosome painting approach to highlight damaging effects of low to moderate doses of ionizing radiation. A detailed account of the damage was assessed from visual analysis and compared with the results of an automatic image processing approach. Correlation of automatic and manual scores showed that, on average, the automatic analysis classified 60, 59 and 73% of the normal, abnormal and rejected metaphases correctly. Zhou and Pycock (1995) presented a robust and adaptable model-based scheme for cell identification that accommodated wide natural variation in cell appearance. To evaluate this automated system, outlines of 107 cells in 20 images, containing both touching and overlapping cells, were each segmented automatically and manually. The area within automatically detected and manually selected cell boundaries not common to both was less than 5% in 96% of the cases tested, indicating that the overall performance is comparable with corresponding manual segmentation.

1.1 *Purpose and organization of the paper*

This paper proposes a fully automated statistical method to identify and classify Fos stained nuclei directly from digitized micrographic images. Section 2 describes data collection methods. Section 3 outlines image analysis techniques, the first stage of our approach; Section 4 describes multivariate statistical methods used in our second-stage classification analysis. Section 5 employs the classification method with the lowest misclassifica error to identify Fos-positive nuclei in new neighbouring images. In Section 6 we assess the predictive performance of the method and compare automated findings with manual results provided by experts, and in Section 7 we consider application of the automation in a broader context.

2. Materials and methods

This paper deals with the automation of the most labour-intense aspect of our overall experimental design, and concerns a portion of a single rat brain. Our overall design involves eight different drugs, with 3–5 rats per drug. After sacrifice, each rat's brain is divided into a lengthy series of hundreds of ultrathin slices that are mounted on glass slides and read under a microscope for levels of Fos expression. Since Fos expression is affected by a variety of external factors, the rats are handled frequently in the days prior to the study, in particular to reduce the effects of stress associated with drug treatment. Animals were treated in accordance with the provisions of the Guide for the Care and Use of Laboratory Animals by the US

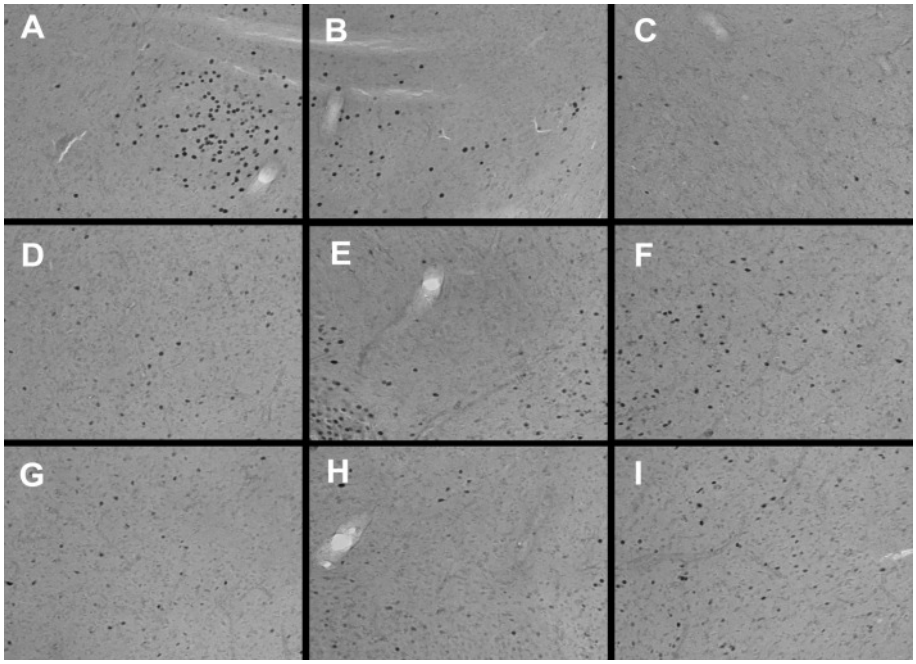

Fig. 1. Brain tissue sections from the central amygdaloid nucleus stained for Fos expression. Panel A is the training set and panels B–I are the test sets.

Department of Health and Human Services, and all procedures were approved by the McLean Hospital Institutional Animal Care and Use Committee.

Haloperidol (Research Biochemicals Inc., Natick, MA) was administered by an intraperitoneal injection of 1 mg kg^{-1} dissolved in 2% lactic acid. These rats were studied concurrently with vehicle-treated control animals. The chosen dose was consistent with previous pharmacologic studies in rats; see Fink-Jensen and Kristensen (1994); MacGibbon *et al.* (1994); Nguyen *et al.* (1992); Robertson and Fibiger (1992); Wan *et al.* (1995).

Two hours after the injection, rats were deeply anaesthetized with sodium pentobarbital and transcordially perfuse with 4% paraformaldehyde. Brains were removed and post-fixed in paraformaldehyde and serial coronal sections 40 μm thick were then cut from each brain on a frozen microtome. Immunohistochemical staining was performed and the sections mounted with permount.

3. IMAGE ANALYSIS

Coronal sections of one Fos-stained rat forebrain after administration of the typical antipsychotic drug haloperidol are first examined at 25$\times$ magnification on a Leica DMRB microscope. Computerized images are then captured using a Sony XC-77 video camera. Figure 1 displays the nine 465 $\times$ 630 pixel images analysed in the present work, taken from sections of the central amygdaloid nucleus.

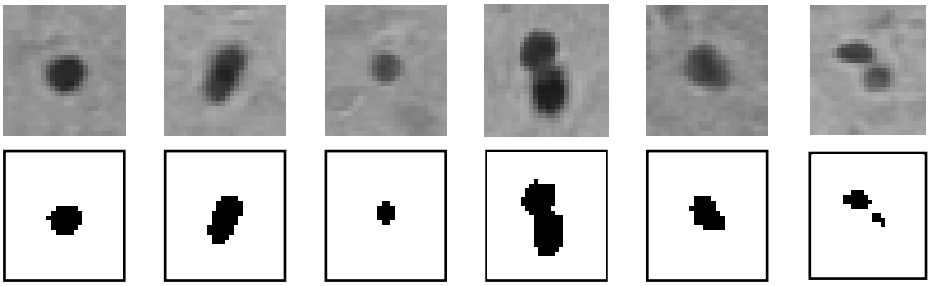

Fig. 2. Examples of Fos-positive nuclei before and after thresholding.

3.1 *Software*

The approach of Kittler and Illingworth (1986), encoded in the interactive image analysis environment Z-Image (CSIRO Mathematical and Information Sciences, Sydney, Australia) is used to threshold panel A of Figure 1 in an effort to segment Fos-stained nuclei from image background. Thresholding methods assume that object and background pixels in an image can be distinguished by their grey-level values. By choosing a grey-level threshold between the dominant values of object and background intensities, the original grey-level image can be transformed into a binary form so that the image points associated with background and objects will assume values 0 and 1 respectively. Kittler and Illingworth (1986) considered thresholding as a classification problem and optimize a criterion function relating to the average pixel classification error rate.

Measurements are made on the objects in the thresholded binary image produced using Z-Image. Each object is a connected pixel region defined by 8-pixel connectivity, as follows: for a given pixel, the eight neighbouring pixels are examined and any that match the original pixel are chosen as part of an object; this process continues until no more neighbours match, and the object is complete. Area (number of pixels), perimeter and lengths of the major and minor axis of a fitted ellipse are recorded for each object.

We condense the four recorded measurements of area, perimeter and major and minor ellipse axis lengths into two measurements per object, namely an estimate of size (area/perimeter) and shape (major axis/minor axis). More details about extracting quantitative information from images can be found in Stoyan and Stoyan (1994) and Glasbey and Horgan (1995), where size and shape are defined differently, more generally in the former work.

3.2 *Fos-positive nuclei*

Examples of cell nuclei determined by our experts to be Fos-positive are shown in Figure 2. These show variations in size and shape and also instances in which two nuclei are touching. The results of applying the thresholding algorithm described in Section 3.1 are also shown. This fails to split one set of touching cells and thus the classification rule will have to recognize touching cells using the size and shape measurements. Figure 3 displays examples of other image objects that, according to our experts, should not be classified as Fos-positive nuclei. Some of these objects are nuclei whose centroids lie in a different focal plane either above or below those that are more darkly stained, which by our current criteria we do not identify as Fos. Other artefacts are from brain tissue not of present interest and background debris from the image capture process. After thresholding, one of these artefacts is considered a possible cell candidate.

Table 1. *Linear and quadratic discrimination*

Expert	Linear discrimination			Quadratic discrimination		
classification	Not-Fos	One-Fos	Two-Fos	Not-Fos	One-Fos	Two-Fos
Not-Fos	56	8	0	51	1	0
One-Fos	0	103	1	5	109	2
Two-Fos	0	0	10	0	1	9

4. STATISTICAL ANALYSIS

We consider several competing multivariate statistical rules to classify identified objects as Fos-positive nuclei or not given our shape and size measurements. The problem here is to use a training set of objects labelled with their correct classifications, and build from this set a classification rule to allocate future cases to one of three classes: (1) not-Fos, (2) one-Fos and (3) two-Fos. The training set was determinded by our experts who manually identified Fos-postitive nuclei from Figure 1, panel A; no thresholding was used for this task. After creating a classification rule using panel A of Figure 1, the training set, we apply the rule to the images in B–I, the test sets. The primary measure of success is the rate of disagreement between the experts' classification and those predicted by the statistical rules.

We consider the following classification methods: classical linear and quadratic discrimination methods (Section 4.1), a non-parametric method that classifies cases on the basis of those in their surrounding neighbourhoods (Section 4.2), and an artificial neural network that is a flexible, nonlinear extension of multiple logistic regression (Section 4.3). We employ the linear and quadratic discriminant, nearest-neighbour, and artificial neural network functions provided by the Splus (MathSoft Inc., Seattle, WA) MASS library of Venables and Ripley (1997) which we tailor to our specific needs.

4.1 *Linear and quadratic discrimination*

Linear discriminant analysis (see, for instance, Mardia *et al.* (1979), Venables and Ripley (1997)) seeks a linear combination of object measurement variables that yields a maximal separation of class means relative to the within-group variance. The resulting transformed variables are linear discriminants used as boundaries to separate objects into the three groups. The result of applying linear discrimination to the test data set is shown in Figure 4(a). Using these boundaries, each new observation from the test sets can be classified into one of the groups depending on where it lies in relation to the boundaries. In this way, the discriminant function serves as a predictive rule for new cases. In quadratic discriminant analysis (see, for instance Krzanowski (1988), pp. 339–340 and McLachlan (1992), pp. 52–59), Mahalanobis distance from the data points in our size–shape space to the class centre define the boundaries, as shown in Figure 4(b). The agreement between the experts' classification and those from linear and quadratic discrimination are shown in Table 1. The linear method tends to overestimate the number of Fos-positive nuclei, while the quadratic method underestimates this number relative to the experts' opinion.

4.2 *Nearest-neighbour classification*

A non-parametic alternative to classical discriminant analysis forms the decision boundary in a model-free way by using k-nearest-neighbour classification with a Euclidean metric (Dasarathy, 1991; Devijver and Kittler, 1987; Ripley, 1996). The k-nearest-neighbour method, employed as a discriminant rule, involves placing a neighbourhood containing k classified points around each as-yet-unclassified point in our size–shape space. When $k = 1$, the rule assigns the unclassified point to the class of its nearest neighbour. For $k > 1$, a simple majority vote of the classified points in the neighbourhood decides the group of the

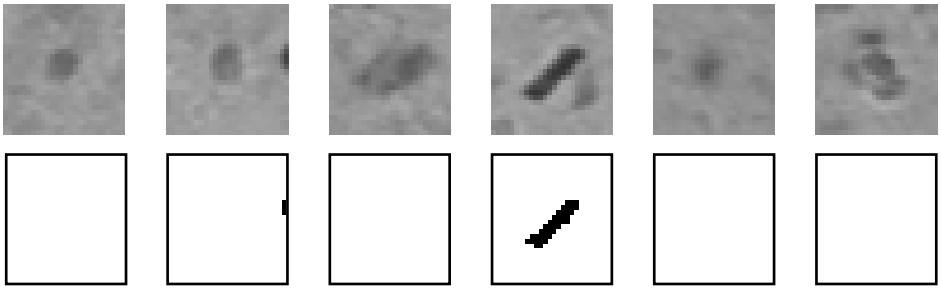

Fig. 3. Examples of non-Fos objects before and after thresholding.

unlabelled point. A discriminant boundary is thus generated as one moves through the size–shape space measuring Euclidean distances between points. The boundary generated for $k = 1$, shown in Figure 4(c), is highly curvilinear as it adapts to the data much more than in the previous parametric discriminant analyses, giving an improved classification thereby. However, being so highly data dependent, the nearest-neighbour boundary would look quite different for a new set of data with different point positions. Perhaps one way to improve the generalizability of the nearest-neighbour rule would be to increase the number of nearest neighbours considered in computing the boundary. Figure 4(d) displays the resulting boundaries using $k = 3$ nearest neighbours. Table 2 shows the agreement between these methods and the manual results of the experts.

4.3 *Artificial neural networks*

Artificial neural networks (ANNs) provide a third classification procedure more general than linear and quadratic discriminant analysis, yet less data dependent and perhaps more generalizable than nearest-neighbour methods: see Moody (1991), Cheng and Titterington (1994), Cherkassky *et al.* (1994), Ripley (1996), Cherkassky and Mulier (1998). For biostatisticians, ANNs 'can be regarded as a graphical notation for specific regression and classification techniques' (Poggio and Girosi, 1993). Modern neuroscience now reveals that actual biological neural networks, across many species from the marine snail *Aplysia* to non-human primates and humans, are much more complex than the McCulloch–Pitts neuron that inspired ANNs over half a century ago (McCulloch and Pitts, 1948). Our artificial neural network is equivalent to that described by Venables and Ripley (1997, p. 340), being a non-linear generalization of logistic regression with more than two classes: a 'linked logistic regressions' model (Ripley, 1996, p. 145). The idea here is that one may be able to do better than the 'single layer' network of inputs (the data) and outputs (the estimated classifications) represented by the linear and quadratic discriminant techniques, or by classical logistic regression with more than two classes, by adding a 'hidden layer' of adaptive weights between input and output to provide some added flexibility to the parametric models already employed. Binary logistic regressions (one class versus the rest) can thus 'talk' with one another through the added layer and perhaps improve overall predictive performance. This added flexibility involves some subjective choice for 'tuning parameters' for ANNs, such as the number of 'nodes' which determine the size of the hidden layer, the so-called 'weight-decay' of the network, being a regularization of inputs to the hidden layer controlled by a parameter λ, and a random seed to begin network training. Figure 4(e) shows the result for a network of size 2, weight decay $\lambda = 0$ (no smoothing) and seed point 1; Figure 4(f) shows the result for a network of the same size with decay $\lambda = 0.001$ and seed point 3. Agreement with the experts' classifications are shown in Table 3.

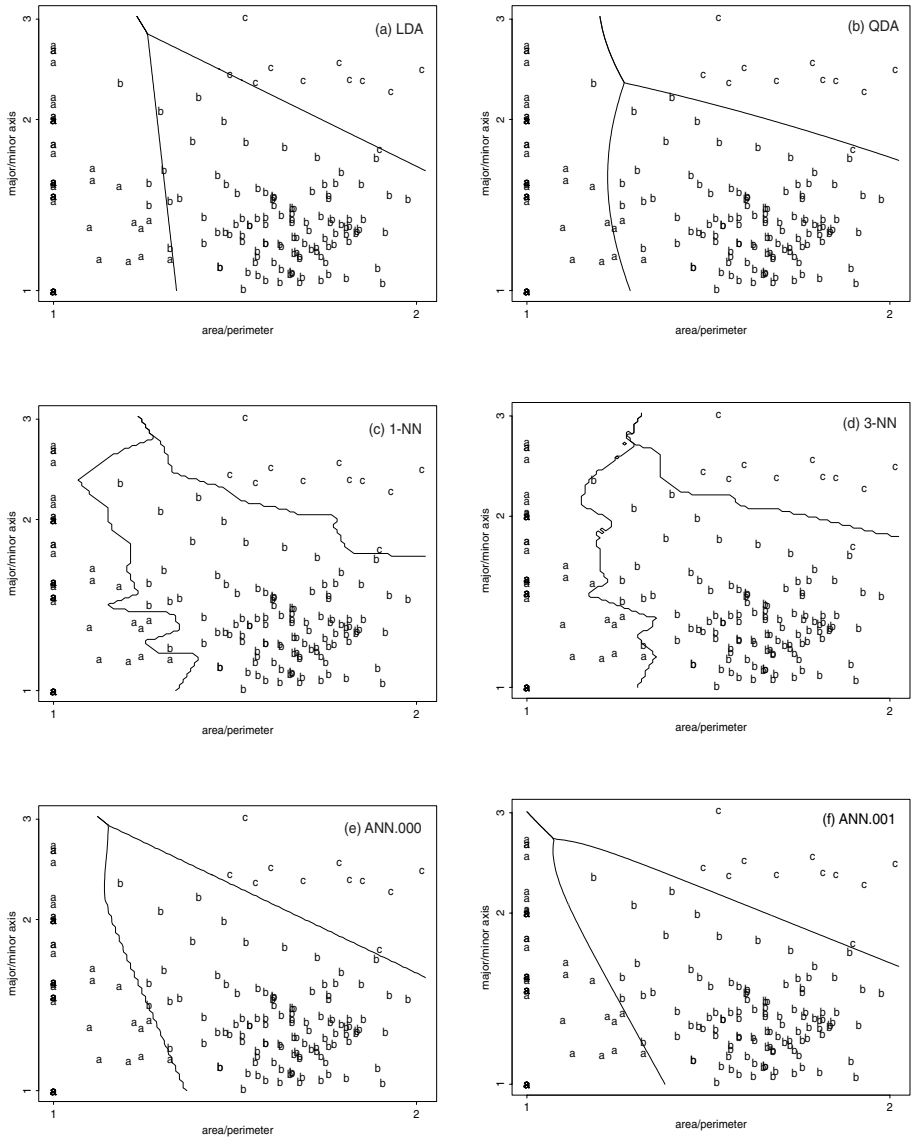

Fig. 4. Results of six classifiers of candidate Fos nuclei in panel A of Figure 1: a, not-Fos; b, one-Fos; c, two-Fos; (a) LDA, linear discriminat analysis; (b) QDA, quadratic discriminant analysis; (c) 1-NN, one nearest neighbour; (d) 3-NN, three nearest neighbours; (e) ANN.000, artificial neural network with weight decay $\lambda = 0.0$; (f) ANN.001, artificial neural network with weight decay $\lambda = 0.001$.

5. RESULTS

Predictive performance of the classification rules is assessed using tenfold cross-validation. The training set is split into ten equally sized parts and the discrimination function is then trained using nine parts and tested on the tenth. This is done ten times and the results averaged. The misclassification rates are shown in Table 4. The lowest cross-validation misclassification rates are for 3-nearest neighbours and

Table 2. *Nearest neighbours*

Expert	1 nearest neighbour			3 nearest neighbours		
classification	Not-Fos	One-Fos	Two-Fos	Not-Fos	One-Fos	Two-Fos
Not-Fos	55	3	1	56	3	0
One-Fos	1	107	1	0	108	1
Two-Fos	0	1	9	0	0	10

Table 3. *Artificial neural networks (ANNs)*

Expert	ANN, $\lambda = 0$			ANN, $\lambda = 0.001$		
classification	Not-Fos	One-Fos	Two-Fos	Not-Fos	One-Fos	Two-Fos
Not-Fos	55	1	0	55	1	0
One-Fos	1	108	2	1	109	1
Two-Fos	0	2	9	0	1	10

Table 4. *Summary results for the six classification methods*

Statistical classifier	Misclassification rate (%)
Linear discrimination	5.06
Quadratic discrimination	5.06
1 nearest neighbour	3.93
3 nearest neighbours	2.25
ANN, $\lambda = 0$	3.37
ANN, $\lambda = 0.001$	2.25

the artificial neural network size 2, decay $\lambda = 0.001$. Since the decision boundaries produced by the artifical neural network have more easily interpretable parameters (being much smoother), the results of this classifier shown in Figure 4(f) are used to identify Fos expression in further images. Examples of the nuclei identified using this network are shown in Figure 5 where black circles are superimposed to indicate the Fos-positive nuclei identified by the ANN in panels B and D of Figure 1. A physical interpretation of the classification rule in Figure 4(f) is as follows: very small objects are classified as not-Fos (label a) regardless of their shape. Medium-to-large objects are classified as one-Fos (label b) if they are roughly circular or as two-Fos (label c) if thinly shaped. An intermediate zone classifies small objects (greater than 1 yet less than $4/3$ on the area/perimeter scale) as not-Fos if they are quasi-ellipical or one-Fos if they are clearly elliptical in shape.

6. Comparison of automatic and human experts

The performance of the ANN classifier is assessed by comparing its results with those provided by expert immunohistochemists through the following simple, pragmatic approach. There is actually no 'correct' classification for these nuclei without performing much more extensive immunohistochemistry and further detailed analyses whose complexity makes obtaining a 'gold standard' entirely void of subjective judgements prohibitive, if even possible. We seek instead to measure the rates of consistency between automated and manual methods in our experiment. The training sample, panel A of Figure 1, is chosen to reflect what in the expert's opinion the actual cell classifications would be. The ANN thereby 'learns' to classify in the same way a particular expert would classify and then classifies further images,

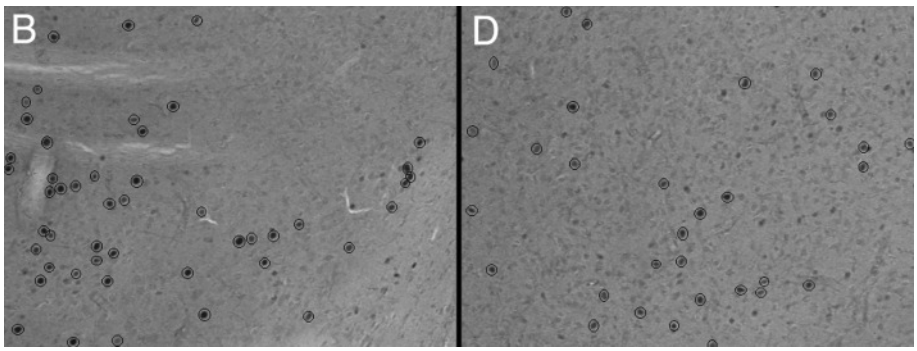

Fig. 5. Fos expression in panels B and D of Figure 1 identified automatically by the artificial neural network, $\lambda = 0.001$. A black circle indicates the object within is identified as a Fos-positive nucleus by the ANN.

Table 5. *Rates of disagreement % between the artificial neural network (ANN, $\lambda = 0.001$) and the first and second experimental results by the experts*

	ANN versus 1		ANN versus 2		1 versus 2		ANN versus 1 and 2	
Image	first	second	first	second	first	second	first	second
A	7.96	6.47	6.79	4.55	8.84	6.32	7.27	6.03
B	5.71	3.70	14.41	8.77	12.82	11.11	4.72	7.96
C	2.50	3.64	3.11	3.05	1.90	3.05	3.12	3.64
D	15.67	8.00	20.00	6.50	10.64	2.54	20.71	6.50
E	7.14	5.88	7.02	5.63	6.67	3.86	8.52	7.18
F	10.20	8.71	11.72	10.07	10.37	4.78	13.60	10.74
G	15.17	10.22	14.39	12.14	8.27	4.65	12.88	10.22
H	7.91	6.23	7.61	7.55	7.19	4.48	7.58	6.91
I	11.06	9.64	14.49	10.10	10.61	10.15	11.94	9.64

without retraining, in a consistent manner, eliminating inter- and intra-expert variability. Experts' manual results are compared with those from the automated method. The consistency of manual interpretation is assessed by comparing results produced by different experts (inter-expert agreement) and by considering the results of each of the experts separately at different times (intra-expert agreement).

Two experts from our laboratory (labelled '1' and '2' in Tables 5 and 6), identified Fos-positive nuclei in panels B–I of Figure 1 manually on two separate occasions about two weeks apart. The rates of disagreement for each expert on each occasion (labelled 'first' and 'second') are shown in Table 5. The rate of disagreement between two sets of results is defined as the percentage of cells in the images which are allocated to different groups (i.e. not-Fos, one-Fos or two-Fos). For instance, the first entry to Table 5 of 7.96 indicates that 7.96% of the cells in the image were not allocated to the same group by the expert as they were by the ANN. The disagreement between experts is given in the column labelled '1 versus 2'. To compensate for the differences between the experts, a set of common images where only Fos-positive cells are identified by both experts is constructed, and the disagreement between those and the automation are shown in the final column of Table 5 labelled 'ANN versus 1 and 2'.

To summarize our findings, experts had similar rates of disagreement when comparing the automatic ANN results to their subjective findings. Disagreement rates were less for both experts on their second assessments. The disagreement between experts is also slightly less on the second assessment, with the maximum rate of disagreement being approximately 12% (i.e. 12% of the candidate cells are not allocated

Table 6. *Rates of disagreement for consistency of internal experts over time to assess the intra-expert variability*

Image	A	B	C	D	E	F	G	H	I
1 versus 1	7.29	7.14	8.62	11.54	3.57	5.14	5.38	4.12	8.20
2 versus 2	9.64	14.75	1.24	18.44	4.38	10.49	8.21	17.58	8.76

Table 7. *Rates of disagreement between the ANN and the results produced by the independent external expert and the consistency for this expert over time*

Image	A	B	C	D	E	F	G	H	I
First assessment	8.37	10.34	3.64	11.54	5.61	10.07	8.89	7.25	11.82
Second assessment	8.78	12.61	2.45	21.23	7.12	9.13	12.14	5.54	12.75
First versus second	9.64	16.13	2.45	16.06	5.93	8.43	10.87	5.19	13.66

to the same group by both experts). The rate of disagreement between the automation and the results common to both experts also tends to be less on the second assessment. Consistency over the two week period for each expert is illustrated in Table 6 where row '1 versus 1' contains rates of disagreement between the first and second manual results for expert 1. Expert 1 has more experience in this type of work than expert 2. The inter- and intra-expert variability seen in Tables 5 and 6 can be eliminated through automation.

To assess automation performance further, an independent external expert immunohistochemist with roughly the same experience as expert 1 provided additional manual results. Again, Fos-positive nuclei in the panels of Figure 1 are identified manually on two occasions, approximately the same interval apart as for the internal experts. Rates of disagreement between these and the ANN classifications are shown in Table 7. The third row of this table shows the consistency of this expert in terms of the rates of disagreement between the first and second assessments. For instance, the first entry in this table, 8.37, indicates that on the first assessment 8.37% of the cells in panel A of Figure 1 were allocated to different groups by the expert and the ANN. Consistency of the independent external expert is comparable with the consistency of experts in our laboratories. There is a notable increase in the rates of disagreement between the two assessments of panel D of Figure 1, from 11.54 to 21.23%. This increase can be attributed to the expert identifying objects as Fos-positive cells that the artificial neural network does not identify. Manual results for the external expert's second assessment did include several objects of the type shown in Figure 3, such as out-of-focus nuclei that we have chosen not to identify as Fos (see Section 3.2). In general, this expert identified more Fos-positive nuclei on the second assessment than on the first.

7. Discussion

The weight-regularized artificial neural network classifier proposed here for automated identification of Fos expression is fast, fully automatic and more accurate than several similar cellular imaging systems currently in use in other contexts. The automation removes variabilities of manual assessment without compromising overall accuracy. The rates of disagreement we report are comparable to the variability seen in manual identification by three different experts, including one expert external to our laboratories.

There are several issues that warrant further attention. First, we caution that our automation is intended at present for identification of Fos-positive nuclei only and not for cell counting or density estimation. Our convention to identify only those Fos nuclei whose centroids lie within the focal plane introduces a bias in observed number and density that can depend on section thickness as well as nuclear size,

shape and orientation. This bias can be eliminated if a physical or optical disector (Sterio, 1984) is used, or through an Abercrombie (1946) correction, with a well designed sampling plan (Beneš and Lange, 2001). However, since this bias is not drug-specific, differences in counts for identical neuroanatomic regions under different drugs are valid for comparisons of drug effects. With regard to classifier design, further applications of the automation will surely find 'three-Fos' and larger clusters of overlapping nuclei, especially where Fos intensity and neuronal density is high. Development of more sophisticated mathematical morphology techniques to separate these clusters into 'one-Fos' nuclei would perhaps be preferable to adding more groups. Also, inclusion of grey level at object centroids as an additional measurement (not shown) did not significantly improve classification ability in the present case. This was because thresholding took care of most of the light out-of-focus cells and only the darker cells were counted. The number of out-of-focus cells identified by the algorithm (but not by the expert) was actually very small and inclusion of the grey-level information made no significant impact on the agreement rates. Finally, we have demonstrated generalizability of the automation to neighbouring tissue slices in the amygdala. We will employ the validated automation as a practical alternative to manual methods for investigation of neuronal activity and circuitry for other brain regions of greater spatial extent.

ACKNOWLEDGEMENTS

This work was funded in part by the National Alliance for Research on Schizophrenia and Depression (BMC, DJE), the Engelhard Foundation (BMC) and NIH grants MH31154 (BMC) and NS37483 (NL).

REFERENCES

ABERCROMBIE, M. (1946). Estimation of nuclear population from microtome sections. *Anatomical Record* **94**, 239–247.

BENEŠ, F. M. AND LANGE, N. (2001). Two-dimensional versus three dimensional cell-counting; a practical perspective. *Trends in Neurosciences* **124**, 11–17.

CAROTHERS, A. D., RUTOVITZ, D. AND GRANUM, E. (1983). An efficient multiple-cell approach to automatic aneuploidy screening. *Analytical and Quantitative Cytology* **5**, 194–200.

CHENG, B. AND TITTERINGTON, D. M. (1994). Neural networks: a review from a statistical perspective (with comments). *Statistical Science* **9**, 2–54.

CHERKASSKY, V., FRIEDMAN, J. H. AND WESCHLER, H. (eds.) (1994). *From Statistics to Neural Networks, Theory and Pattern Recognition Applications*, Series F: Computer and Systems Sciences, **136** NATO-PCO Database. Berlin: Springer.

CHERKASSKY, V. AND MULIER, F. (1998). *Learning From Data*. New York: Wiley.

COHEN, B. M., WAN, W., FROIMOWITZ, M. P., ENNULAT, D. J., CHERKERZIAN, S. AND KONIECZNA, H. (1998). Activation of midline thalamic nuclei by antipsychotic drugs. *Psychopharmacology* **135**, 37–43.

DASARATHY, B. V. (ed.) (1991). *Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques*. Los Alamitos, CA: IEEE Computer Society Press.

DEUTCH, A. Y., ÖNGÜR, D. AND DUMAN, R. S. (1995). Antipsychotic drugs induce Fos protein in the thalamic paraventricular nucleus: a novel locus of antipsychotic drug action. *Neuroscience* **66**, 337–346.

DEVIJVER, P. A. AND KITTLER, J. V. (eds.) (1987). *Pattern Recognition Theory and Applications*. Berlin: Springer.

DIGGLE, P. J. (1983). *Statistical Analysis of Spatial Point Patterns*. New York: Academic.

DIGGLE, P. J., LANGE, N. AND BENEŠ, F. M. (1991). Analysis of variance for replicated spatial point patterns in clinical neuroanatomy. *Journal of the American Statistical Association* **86**, 618–625.

DRAGUNOW, M., ROBERTSON, G. S., FUALL, R. L.M., ROBERTSON, H. A. AND JANSEN, K. (1990). D2 dopamine receptor antagonist induce Fos and related proteins in rat striatal neurons. *Neuroscience* **37**, 287–294.

FANTES, J. A., GREEN, D. K., HILL, W., STARK, M. H., GORDON, J. M. AND PIPER, J. (1995). Application of automation to the detection of radiation damage using FISH technology. *International Journal of Radiation Biology* **68**, 263–276.

FINK-JENSEN, A. AND KRISTENSEN, P. (1994). Effects of typical and atypical neuroleptics on FOS protein expression in the rat forebrain. *Neuroscience Letters* **182**, 115–118.

GARCIA-SAGREDO, J. M., PIPER, J., RUTOVITZ, D., VAQUERO, J. J. AND VAZQUEZ-MAZARIEGO, Y. (1994). Automatic scoring of sister chromatid exchanges by image analysis in a dose response experiment. *Environmental and Molecular Mutagenesis* **23**, 194–199.

GLASBEY, C. A. AND HORGAN, G. W. (1995). *Image Analysis for the Biological Sciences*. Chichester: Wiley.

HUNT, S. P., PINI, A. AND EVANS, G. (1987). Induction of *c-fos*-like protein in spinal cord neurons following sensory stimulation. *Nature* **328**, 632–634.

KITTLER, J. AND ILLINGWORTH, J. (1986). Mimimum error thresholding. *Pattern Recognition* **19**, 41–47.

KRZANOWSKI, W. J. (1988). *Principles of Multivariate Analysis*. Oxford: Clarendon.

MARDIA, K. V., KENT, J. T. AND BIBBY, J. M. (1979). *Multivariate Analysis*. London: Academic.

MACGIBBON, G. A., LAWLOR, P. A., BRAVO, R. AND DRAGUNOW, M. (1994). Clozapine and haloperidol produce a differential pattern of immediate early gene expression in rat caudate-putamen, nucleus accumbens, lateral septum and islands of Calleja. *Molecular Brain Research* **23**, 21–32.

MCCULLOCH, W. S. AND PITTS, W. (1948). The statistical organization of nervous activity. *Biometrics* **4**, 91–99.

MCLACHLAN, G. J. (1992). *Discriminant Analysis and Statistical Pattern Recognition*. New York: Wiley.

MOODY, J. E. (1991). Note on generalization, regularization and architecture selection in nonlinear learning systems. *First IEEE-SP Workshop on Neural Networks in Signal Processing*, Los Alamitos, CA: IEEE Computer Society Press, pp. 1–10.

MORGAN, J. I., CURRAN, T., COHEN, D. R., HEMPSTEAD, J. L. AND CURRAN, T. (1987). Mapping patterns of of *c-fos* expression in the central nervous system after seizure. *Science* **237**, 192–197.

NGUYEN, T. V., KOSOFSKY, B. E., BIRNBAUM, R., COHEN, B. M. AND HYMAN, S. E. (1992). Differential expression of *c-fos* and zif 268 in rat straitum after haloperidol, clozapine and amphetamine. *Proceedings of the National Academy of Science, USA* **89**, 4270–4274.

OZAKI, T., KATSUMOTO, E., MUI, K., FURUTSUKA, D. AND YAMAGAMI, S. (1998). Distribution of Fos and Jun-related proteins and activator protein-1 composite factors in mouse brain induced by neuroleptics. *Neuroscience* **84**, 1187–1196.

POGGIO, T. AND GIROSI, F. (1993). Learning algorithms and network architectures. In Poggio, T. and Girosi, F. (eds), *Exploring Brain Function: Models in Neuroscience, Proceedings of the Dahlem Conference*, New York: Wiley.

RIPLEY, B. D. (1996). *Pattern Recognition and Neural Networks*. Cambridge: Cambridge University Press.

ROBERTSON, G. S. AND FIBIGER, H. C. (1992). Neuroleptics increase *c-fos* expression in the forebrain: contrasting effects of haloperidol and clozapine. *Neuroscience* **46**, 315–328.

STERIO, D. C. (1984). The unbiased estimation of number and sizes of arbitrary particles using the disector. *Journal of Microscopy* **163**, 127–136.

STOYAN, D. AND STOYAN, H. (1994). *Fractals, Random Shapes and Point Fields: Methods of Geometrical Statistics*. New York: Wiley.

VENABLES, W. N. AND RIPLEY, B. D. (1997). *Modern Applied Statistics with S-PLUS*. New York: Springer.

VOSSEPOEL, A. M. AND DE KANTER, M. A. (1987). A two-stage approach to automatic segmentation of metaphase images. *Proceedings of the* 10*th Scandinavian Conference on Image Analysis, Lappeenranta, Finland* 1, 365–372.

WAN, W., ENNULAT, D. J. AND COHEN, B. M. (1995). Acute administration of typical and atypical antipsychotic drugs induces distinctive patterns of Fos expression in the rat forebrain. *Brain Research* **688**, 95–104.

ZHOU, P. AND PYCOCK, D. (1995). Robust statistical model-based cell image interpretation. *Proceedings of the 6th British Machine Vision Conference (Birmingham)*. pp. 117–126.