

A VERY RELIABLE METHOD FOR DETECTING BACTERIAL GROWTHS USING NEURAL NETWORKS¹

Victor Ciesielski, Jihan Zhu
Department of Computer Science
Royal Melbourne Institute of Technology
GPO Box 2476V
Melbourne Victoria 3001
AUSTRALIA
vc@goanna.cs.rmit.oz.au

Abstract

We describe a neural network based system for detecting bacterial growths on microbiology plates. A major goal of the system is to develop a classifier that is robust and reliable enough to be used in a microbiology laboratory. We have experimented with three layer feed forward networks of different architectures. The best performance (8 errors, 99.32% correct classification on a test set of 1174 growths) was achieved with a 400-10-1 network where the inputs were the intensity values of the pixels in a 20x20 square around a growth position. In order to reduce the number of errors at the expense of passing some decisions to an expert microbiologist we implemented a voting scheme involving 4 of the best performing networks. This classifier was able to detect “hard” cases which should be referred to a human technologist and gave only 4 errors in 1174 test cases at the expense of 31 cases referred to the microbiologist.

1 Introduction

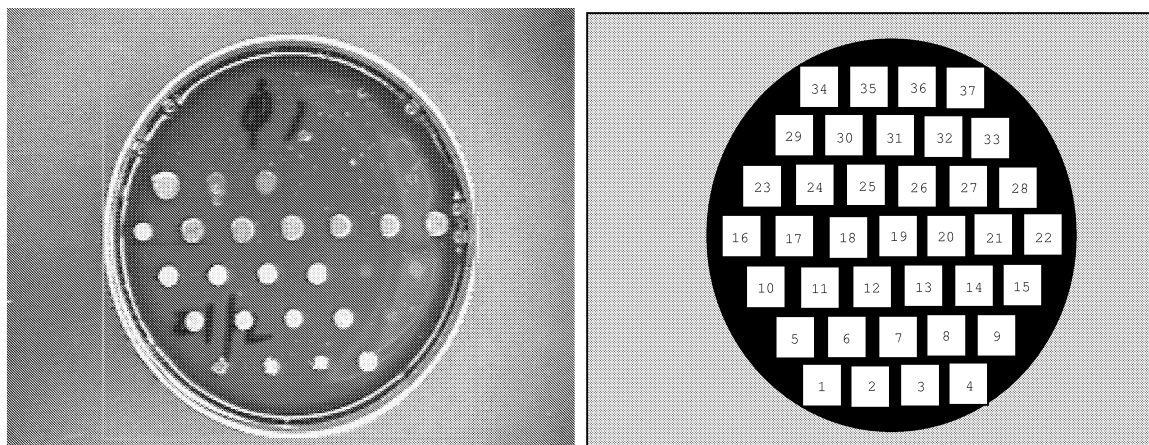


Figure 1: A Bacterial Growth Plate and Well Numbering System

There have been a number of reports of good performance by neural network classifiers for vision and pattern recognition problems, [1, 2, 5, 6, 7], for example. The performance reported in these investigations varies from 70% to 97% correct classification. These investigations appear to be mostly feasibility studies which have not attempted to achieve the best possible performance. We are interested in obtaining the best possible performance from a neural network classifier

¹ *Proceedings of the International Joint Conference on Neural Networks*, Pages 62–67, Beijing, November, 1992

on a problem where accurate classification is critical and on which human experts perform virtually without error.

There are a number of issues in using neural networks for computer vision which are currently somewhat controversial. These include whether inputs to the network should be grey scale or colour and whether the inputs should be raw pixel data or the results of preprocessing operations such as smoothing, Fourier transforms and feature extraction.

While our primary goal is not to investigate these issues we have had to take certain decisions in the development of our classifiers. We use raw pixel data with no preprocessing and we have chosen to compare decision making based on color and grey scale data.

2 Problem Definition

Detecting whether bacteria are growing on a bacteriology plate is a key step in many procedures in the clinical microbiology laboratory. A common procedure involves preparing plate media with different characteristics and nutrients, placing “seed” bacteria from patient specimens on each plate, determining the plates on which a particular specimen grows and then using this growth pattern to draw conclusions about various characteristics of the specimen, such as its identity for example.

In many laboratories each plate is divided into a fixed number of positions or “wells” and bacteria are placed in these positions by a stamping machine. The general appearance of a plate as used in our experiments is a round plastic dish containing a reddish jelly on which are rows of small whitish/yellowish/greyish circles (Figure 1). A trained microbiologist then determines whether or not there is a growth in each position. There can be 20-50 plates in a batch, depending on the microbiological procedure being performed. The method of seeding the wells ensures that well i on each plate is seeded from the same specimen from the same patient. There may also be one or more ‘control’ plates. These plates contain growth media on which all specimens in all wells should grow. The results of the analysis of growth patterns on all plates are collated into final conclusions about the specimen.

2.1 Performance Objectives

The goal of the project is to determine whether it is possible to develop a neural network classifier which is robust and reliable enough to perform this inspection task in a microbiology laboratory and which runs in real time on inexpensive personal computers. Since results from the system will be used in patient care it is critical that an automated system should make the correct decisions. Also, in this domain, it is acceptable for the automated system not to make a decision in some cases and pass the well on to a human microbiologist.

2.2 Domain Characteristics

There are a number of characteristics of the domain which make it difficult for an automatic system to make accurate decisions.

- There is a large variety in the colour characteristics of the growths. Since each plate contains specimens from many different patients there are many different organisms on each plate. Some of these have very bright, distinctive colours with strong contrast to the background (figure 1, well 5). Other growths have very similar colours to the background (figure 1, well 25).
- Some bacteria are strong growers giving large bright growths (figure 1, well 4) some are weak growers giving only small faint growths (figure 1, well 15).

- Most growths are circular in shape, however a significant number are deformed (figure 1, well 24).
- The stamping machine which puts the bacteria on the plates sometimes leaves marks which look very similar to growths.
- The stamping machine is not fully rigid. There is some variability in the position of each well between plates. Thus it cannot be assumed that the centre of each well will be in the same place on each plate, nor can it be assumed that the distances between centres of adjacent growths will be the same number of pixels apart on consecutive plates. This distance can differ by up to 5 pixels.
- Fresh growth media is prepared every few days, giving a slightly different background colour.
- Lighting conditions result in some shadows and reflections on the plates (in figure 1, there is a crescent of reflection and shadow from well 36 through wells 21 and 22 to well 9).
- Black plate identification characters or paper labels (not shown in figure 1) appear on the bottoms of the plates.

2.3 Hardware

There is a requirement that the classification system run in real time on relatively inexpensive personal computers. In practice this means that the system should be capable of classifying wells at the rate of about one per second. This requirement determines our stand on the “preprocessing vs raw pixels” issue. We use raw pixel data because the hardware is not capable of executing any significant preprocessing operations within this time constraint.

Although the classifier needs to run within the above constraints the training of the classifier need not be done on the same equipment. We have used an Encore multimax for training the networks.

The vision hardware includes a TV camera and image capture board which deliver a 256x200 colour image with 5 bits each for red, blue and green. Most growths are about 10 pixels in diameter.

3 Neural Network Experiments

We performed a variety of experiments with different three layer feed forward networks which were trained with back-propagation using the software available with [3] and [4]. The network architectures and the performance of each network are given in table I. In experiment 9 the raw red, blue and green values from the TV camera were used as inputs to the network. The pixels of a 10x10 square centred on each growth position were used giving three input neurons per pixel and 300 total inputs. The middle layer in experiment 9 had one neuron and there was one output neuron. The data from 263 (153 growths, 110 empty) wells taken from 3 different batches was used to train the network and this took 1.35 cpu minutes on an encore multimax with 8 processors. The final two columns of the table summarize the performance, 15 errors on the test set of 1174 wells giving a 98.72% correct classification rate on the test set.

Experiments 1-4 were motivated by the observation that the network needs to discriminate between two categories - growth and empty. Since each growth is about 10 pixels in diameter, there should be two kinds of squares - mostly growth and totally empty. Experiments 1-4 involve attempts to train 4 networks to make this distinction based on intensity values ($Intensity = r + b + g$). Experiments 5-8, with an input of a 20x20 square of pixels, were essentially an

Expt No	Input Type	Square Size	Network Architecture	Training Time(mins)	Number of Errors	Performance
1	Intensity	10x10	100-1-1	4.48	23	98.04%
2	Intensity	10x10	100-2-1	4.48	23	98.04%
3	Intensity	10x10	100-5-1	4.48	26	97.79%
4	Intensity	10x10	100-10-1	1.75	25	97.87%
5	Intensity	20x20	400-1-1	2.28	10	99.14%
6	Intensity	20x20	400-2-1	2.28	9	99.23%
7	Intensity	20x20	400-5-1	4.48	21	98.21%
8	Intensity	20x20	400-10-1	8.54	8	99.32%
9	Raw RGB	10x10	300-1-1	1.35	15	98.72%
10	Raw RGB	10x10	300-2-1	1.35	11	99.06%
11	Raw RGB	10x10	300-5-1	3.36	18	98.46%
12	Raw RGB	10x10	300-10-1	5.22	18	98.46%
13	Raw RGB	16x16 ¹	768-1-1			N.C.
14	Raw RGB	16x16	768-2-1	31.35	13	98.89%
15	Raw RGB	16x16	768-10-1	36.33	34	97.10%
16	Raw RGB	16x16	768-5-1	48.19	17	98.55%
17	Norm RGB	16x16	768-10-1	51.37	28	97.62%
18	Norm RGB	10x10	300-5-1	2.27	31	97.36%
19	Control Plate	20x20	800-2-1	3.23 ²	4	99.36%
20	Voting	10x10 20x20 10x10 16x16	INTEN 300-10-1 INTEN 400-10-1 Raw RGB 100-2-1 Raw RGB 768-2-1		4	99.66%

Table 1: Summary of neural network experiments, training set 263/1437 patterns, test set 1174/1437 patterns, N.C. - training did not converge

attempt to make a network learn the difference between a square with a circle in the middle and an empty square. Experiments 9-16, in conjunction with experiments 1-8 were designed to investigate the effect of colour on the decision making. In experiments 17 and 18 we explored the effect of normalizing the data, that is $NormR = r/(r + b + g)$, with a similar calculation for the blue and green components. In experiment 19 the control plate was used. Recall that the control plate contains media on which all bacteria are expected to grow. The expectation here was that the network would learn the difference between growths on control plates and growths on test plates.

The output of each network is a real number in the range (0.0,1.0). In all of the above experiments an output of less than 0.5 was treated as an empty well and an output of 0.5 or greater was treated as a growth well.

3.1 Analysis of Results

Since the observed error rate on a test set of 1000 cases is within 1% of the true error rate ([8, p28-30]), we conclude that most networks reached about the same level of performance. Analysis of the errors revealed that there were a handful of wells misclassified by all networks and other wells which were correctly classified by some networks but misclassified by others. Visual inspection of the wells shows that many of these “rogue” wells are similar to wells that

¹The largest colour network that would fit in memory.

²On a reduced training and test set.

have been classified correctly. Some of these wells can be seen in figure 3. This observation and the fact that not all networks with one hidden layer could be successfully trained suggest that the problem is “almost” linearly separable.

Interestingly enough the black writing on the bottoms of the plates did not affect classification performance, however wells over the paper labels could not be classified correctly and were removed from the data set.

The performance of the normalized networks is significantly lower. This is probably due to the fact that normalization effectively removes intensity information, leaving only colour, which apparently is not sufficient for decision making.

The preliminary work with the use of the control plates (experiment 19) showed slightly improved performance on a reduced test set. This suggests that using this additional information will lead to improved performance.

Increasing the size of the middle layer significantly increased the training times. The experiments with intensity and colour are somewhat inconclusive, they tend to indicate that colour is not a big factor in the decision making.

Due to the nature of the wells and the stamping machine mentioned above we considered that explicit implementation of shift invariance might be necessary. However we worked on the assumptions that there would be enough variation in the centres of the growths to have shift invariance built into the training set. The results indicate that we have been successful in achieving shift independence.

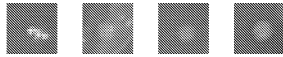


Figure 2: The Four Wells Mis-Classified by the Voting Scheme.

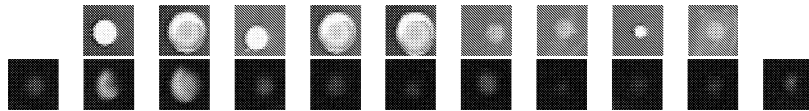


Figure 3: Twenty No-Decision Wells from the Voting Scheme

Although the performance of the networks is very good it is still below that of the experts which is virtually 100%. However in making decisions on difficult growths the microbiologists hold the plate in their hands and look at the suspicious growth from several directions and in several different lighting conditions. Using this technique they are able to observe the vertical dimension of the bacterial growth. This is not possible in the automated system with only one fixed camera and only one light source. In some difficult cases consultation with other microbiologists is necessary.

On the assumption that outputs close to 1.0 were definite growths, those near 0.0 were definitely empty wells and those around 0.5 were uncertain we tried the following decision criteria: 0.00-0.19 – No growth, 0.20-0.79 – No decision, refer to microbiologist and 0.80-1.00 – growth. Contrary to our expectations this resulted in a large number of cases referred to the microbiologist without significant increase in accuracy.

The need for a reliable way of finding the hard cases and the presence of the rogue wells described above suggested the voting scheme of experiment 20. If all networks voted for growth then the output was growth and similarly for empty wells. If there was disagreement between

any of the networks then the well was referred to the microbiologist. This gave only 4 errors (figure 2) and resulted in 31 cases being referred (figure 3). This voting method appears to capture most of the “hard” cases in the domain. Further details of this work can be found in [9]

4 Conclusions

We have shown a successful application of neural network technology in a difficult real world domain. We have developed a classifier that is reliable and robust enough to be used in a microbiology laboratory on AT class hardware. We have also shown that it is possible to use a voting scheme based on networks of different architectures to reliably detect the “hard” cases in this domain. These cases should be referred to a human expert.

We have shown that, for this domain, very accurate classification can be achieved using raw pixel data and that there is no difference between the use of intensity and colour data.

The systems we have developed currently do not give an indication of the size of the growth. Further work is needed for applications where size is important.

Acknowledgements We would like to thank our colleagues, Dr. John Spicer and Claire Franklin, at the Alfred Hospital in Melbourne for their time, expert advice and many discussions and Robert Novella who performed the experiments with the control plates.

References

- [1] S. Haykin and C. Dong. Classification of radar clutter using neural networks. *IEEE Transactions on Neural Networks*, 2(6):589–595, 1991.
- [2] D. Y. Kim, S. L. Chien, and H. Son. Multiclass 3-d aircraft identification and orientation estimation using multilayer feedforward neural network. In *Proceedings of the International Joint Conference on Neural Networks*, pages 759–761, Singapore, Nov. 1991.
- [3] J. L. McClelland and D. E. Rumelhart. *Explorations in Parallel Distributed Processing*. MIT Press, Cambridge, Massachusetts, 1989.
- [4] Y. Miyata. *A Tool For Constructing, Running, and Looking into a PDP Network*. Boulder, USA.
- [5] T. F. Pawlicki, D. Lee, J. J. Hull, and S. N. Srihari. Neural network models and their application to handwritten digit recognition. In *Proceedings of the IEEE First International Conference on Neural Networks*, pages 63–70, San Diego, Ca, June 1987.
- [6] D. A. Pomerleau. Rapidly adapting artificial neural networks for autonomous navigation. In D. S. Touretzky, editor, *Advances in Neural Information Processing*, pages 429–435. Morgan Kaufmann, 1991.
- [7] A. Rajavelu, M. T. Musavi, and M. V. Shirvaikar. A neural network approach to character recognition. *Neural Networks*, 2(3):378–393, 1989.
- [8] S. M. Weiss and C. A. Kulikowski. *Computer Systems That Learn-Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning, and Expert Systems*. Morgan Kaufmann, NY. USA, 1991.
- [9] J. Zhu. An analysis of computer vision and neural network techniques for detecting bacterial growths in microbiology images. *Masters Thesis — Royal Melbourne Institute of Technology (in press)*, 1992.