Proceedings of the 2007 IEEE Symposium on Computational Intelligence in Image and Signal Processing (CIISP 2007)

Evolutionary Algorithms in the Classification of Mammograms

D. C. Hope E. Munday S. L. Smith
Department of Electronics
The University of York
Heslington, York YO10 5DD, UK
{dch104, sls5, em24}@ohm.york.ac.uk

Abstract-**The application of pattern recognition techniques to radiology has the potential to detect cancer earlier and save lives, and consequently much research has been devoted to this problem. This worked tackled a subset of the problem, investigating a novel method of classifying mammograms using an evolutionary approach known as Cartesian Genetic Programming (CGP). Microcalcifications, one of two major indicators of cancer on mammograms, were used for the classification. A large software framework was written in order to investigate this, which allows the viewing of images, manual segmentation of lesions and then automatic classification. Two classification approaches were pursued, the first classifying on texture features and the second, a new approach, classifying by using the lesion's raw pixel array. Early results using the system showed some potential. It was found that during training, networks could obtain correct classification rates of between 80 and 100%. The best results were approaching those in the contemporary literature and suggest the technique warrants further investigation.**

I. Introduction

Breast cancer accounts for one third of all the cancers in women [1] and in 2003 accounted for the deaths of 10,500 women in the UK alone. Since breast cancer screening was introduced in 1988, incidence of breast cancer has risen but the number of deaths has declined. The five year survival rate for people diagnosed as having breast cancer is predicted by Cancer Research UK at 76% and the earlier the cancer is identified the greater than success of treatment. Screening is therefore essential in reducing deaths from breast cancer and it is important to make this as accurate as possible.

The most innovative approach over the past ten years has been the use of Computer Aided Diagnosis (CAD), employing specially developed image processing and pattern recognition systems. A number of commercial systems are now available [2] and improved performance has been claimed with an increase in sensitivity from 74% to 87%. Importantly, it is estimated that for every 100,000 cancers detected using traditional approaches, an additional 20,500 could be detected using CAD [3]. However, experience of CAD systems in the real world has been disappointing and performance obtained below that previously claimed [4]. It is therefore essential to continue development of CAD systems, not only to ensure that false negatives are minimized, but false positives as well, which impact on the workload of the radiologist.

The aim of the work reported in this paper was to assess the potential benefit of using evolutionary algorithms in the classification of mammograms as part of a CAD system and determine whether further development of such algorithms will lead to a more confident diagnosis.

The implementation of a full CAD system is a huge undertaking and not viable or necessary for the evaluation of the algorithms proposed. Therefore, rather than develop a complete CAD system that acquires, preprocesses and segments appropriate sections of the mammogram, this investigation will rely on prior knowledge by using previously acquired and processed images of known pathology. Thus, only small sub-images taken from previously diagnosed mammograms are used where the nature and location of the suspicious regions are known and have been documented as such by clinical personnel.

The problem presented to our algorithms reduces to one of deciding if the suspicious area is an indication of cancer (malignant) or harmless (benign). Two powerful indicators of cancer that are commonly used in evaluating mammograms are known as masses and microcalcifications. Masses are the larger of the two indicators and can be either benign or malignant. An example of this type of potential growth can be seen in Fig. 1. Characteristics such as the border and density of the mass, which is greater for malignant examples, can be used for classification. Traditionally, masses are more difficult to classify than microcalcifications. Microcalcifications are essentially small calcium deposits which occur as the result of secretions from ductal structures that have thickened and dried. They tend to occur in clusters (as can be seen in Fig. 2.) and it is reported that 40-50% represent cancer [5]. Features that have previously been used to distinguish benign and malignant microcalcifications include their shape, density, distribution and definition. Not only are these characteristics useful for a radiologist attempting to classify a mammogram, but they have been used extensively in feature extraction for established image processing techniques.

Although it is believed that evolutionary algorithms can be used effectively to analyze masses it was decided, initially, to work exclusively with microcalcifications as more work has already been done in this area, providing a greater source of literature to which comparisons can be made. Additionally, microcalcifications are easier to identify than masses and so are more expedient for this work.

Previous work undertaken in the classification of microcalcifications using both traditional image analysis techniques and evolutionary algorithms is considered in Section II. The evolutionary algorithm used in the current work will then be described in Section III and results applying this technique to a number of digitized mammograms will be considered in Section IV. Finally, the potential of the proposed algorithm will be evaluated in Section V.

II. Previous Work

Over recent years there has been much research into the application of computer aided diagnosis (CAD) to breast cancer with numerous different approaches being exploited. Many of these involve image analysis of the digitized mammogram – a low dose x-ray of the breast. A typical approach is to use a pattern recognition scheme comprises (i) sensing, (ii) segmentation, (ii) feature

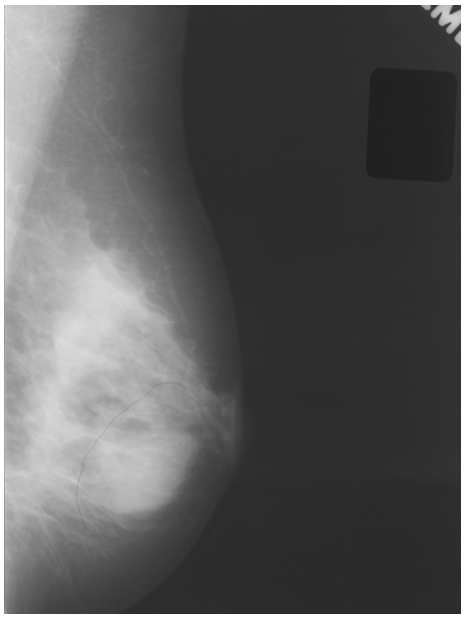

Fig. 1. Highlighted mass lesion. Taken from the DDSM database [6]

(a)

(b)

Fig. 2. (a) Region of breast tissue with microcalcifications and (b) binarised image showing their possible locations. Taken from the DDSM database [6]

extraction, (iv) feature selection and (v) classification, to isolate and then characterize the microcalcifications. Each stage of this processing is a potentially complex operation requiring much investigation.

The work presented in this paper is concerned specifically with the characterization and classification of the microcalcifications - the feature extraction, feature selection and classification stages of the pattern recognition scheme. Consequently, the sensing and segmentation stages of the scheme, while relevant and important in a fully implemented system [7] are not considered here and for the purpose of the experiments described in Section IV will be undertaken manually.

A. Feature extraction

Once segmentation is completed any microcalcifications located need describing in terms of features, these features are collected in the feature extraction stage.

Features, as described here, are real numbers obtained by applying some mathematical expression to image data, e.g. spatial domain pixel values or transformed spectral data. By examining these features one can come to a conclusion as to the nature of the calcification

The feature extraction process regularly exploits morphological features such as area and perimeter, texture features such as spatial grey level dependence matrices and features taken from the wavelet transform of the image. Morphological features are often referred to as shape features and are useful in classification of microcalcifications. Reference [5] provides information for radiologists about the varying features of benign and malignant calcifications. For example it advises that benign ones have a round ring like shape with well defined borders. Malignant on the other hand have varying shape and poorly defined borders. Such characteristics can be described using morphological feature extraction. Reference [8] used a number of morphological features and these included: area, mean density (calculated as average of pixels gray values above background level in the signal region), eccentricity, axis ratio, ratio of x direction to y direction moments.

In terms of texture features the spatial gray-level dependence (SGLD) matrix was used for many features derived including correlations, entropy, variance, angular second moment and others. Another neural network based paper [9] relied purely on texture features concentrating on ones from the SGLD matrix.

An alternative method is suggested in [10] whereby the discrete cosine transform is taken of the image and then they derived "block activity and spectral entropy from the DCT coefficients". Reference [11] also gives brief mention of Fourier methods and a wavelet method whereby standard features (energy and entropy) were extracted from each scale in the transform. A wavelet transform allows the splitting of an image into different scales for various positions in the image, hence why it is often referred to as a multiscale method.

B. Feature selection

At the end of the feature extraction stage there may be a very large number of features, and whether a statistical classifier is being used, a neural network or a genetic algorithm (as will be the case in this investigation) it is not helpful to have too many features. It may make the running time on a computer higher but on a more fundamental level the likelihood is that some of the features extracted may be of no relevance in discriminating benign and malignant lesions. Thus, it is advantageous to select those features which will be most effective in the following classification stage.

A useful comparison of feature selection techniques is presented in [8]. This compares two methods of feature selection, Linear Discriminant Analysis (LDA) and a genetic

algorithm. In LDA features are added to and removed from the system used to decide which class (benign or malignant) a mammogram belongs to. All the features collected in the previous feature extraction stage are available to use. In stepwise LDA, the version described in the CAD literature, features are added one at a time. To decide if a feature is useful in discriminating between two classes the outputs of the system must be considered. There are two groups, outputs for when the input was malignant, and outputs for when it was benign. The analysis is done by comparing the within group sum of squares (i.e. variance, between the groups) and this is done in the case where the feature is included and when it is not. It is equivalent to saying that, if the means of the outputs between malignant and benign are similar without a feature and different with a feature then that feature is useful at discriminating. A threshold is used to determine if a feature is powerful enough. There is also a removal step where features are removed one at a time and excluded based on a threshold, i.e. if taking it out makes little difference it is excluded. Termination happens when the calculated power of all the features not chosen is less than that needed to enter and all those in are greater than the threshold for leaving. This is the more traditional selection technique but it is found in the comparison that "the GA could select a feature set comparable to or slightly better than that selected by stepwise LDA" [8].

C. Classification

In the case of breast cancer, the classifier decides if a given mass or microcalcification is malignant or benign. It is the central part of any computer aided diagnosis scheme and ultimately decides whether a breast is deemed potentially cancerous, and in need of further investigation, or benign. If a scheme is overly cautious then it will have financial and resource implications, in that there might be too many check ups, or it might unnecessarily use up valuable time for a radiologist if it presents too many potential lesions for them to examine. On the other hand if it only selects the very obvious cases then it may pick up less than a radiologist and leave many potential cancers unnoticed. Therefore it requires careful design. A number of popular classifiers are identified by [11] and listed here:

• Neural networks: a parallel information processing network based on the structure of neurons. It is noted in [11] that they are advantageous in the situation where "only a few decisions are required from a massive amount of data and for the applications where a complex non-linear relation needs to be learned".

• K-nearest neighbors: This starts with a set of patterns for a known sample, for example a set of simple statistics for a set of microcalcifications that are known in advance to be cancerous. Then new unknown patterns can be compared to the known ones. The K nearest samples will be classified as having cancer as well.

• Bayesian classifier: This considers the probability $p(w_i|\boldsymbol{x})$ that a given pattern $\boldsymbol{x}$ belongs to a class w_j indicating, for example, malignancy. By Bayes' theorem this can be seen proportional to $p(w_i)p(\boldsymbol{x}|w_i)$. This type of classifier minimizes the total loss - the probability of assigning the pattern to a given class when it actually belongs to another class. We estimate the $p(\boldsymbol{x}|w_i)$ probability density functions (often as Gaussian) and use in $p(w_i)p(\boldsymbol{x}|w_i)$ in order to calculate the average loss in deciding that a pattern belongs to each possible class. The pattern is classified according to the class that yields the smallest loss. See [12] for more information.

D. Use of evolutionary algorithms

Evolutionary algorithms are a family of population based algorithms that use facets of biological evolution such as natural selection, reproduction, mutation and recombination to evolve solutions to problems. Examples of evolutionary algorithms including Genetic Algorithms and Genetic Programs are considered below.

Genetic algorithms (GAs) have previously been used in CAD schemes and they have proved successful. One of the keys papers that influenced this project is a GA based paper [8] in which a genetic algorithm was used for feature selection and it proved successful in this area. Performance was found to be a match for the well established LDA method and even better sometimes. The review paper [8] also reported the only use of genetic algorithms as being in feature selection as in the aforementioned paper. Neural networks are another biologically inspired technique that has been widely adopted and successfully but uses of GAs are limited and this raises the question of whether genetic algorithms could be further used. Genetic Programs (GPs) have previously been used in image processing by Cai, Smith and Tyrrell for noise removal from images [13]. In this case a form of genetic program called Cartesian Genetic Programming (CGP) was used (this will be explained shortly). Clearly, the removal of noise is a very different to pattern recognition but it suggests that application of genetic programs to this type of problem could be an interesting avenue to explore.

An example of the use of genetic algorithms as an alternative feature selection method starts with a data structure termed a *chromosome* which is the length of the total number of features available. Each gene in the chromosome is a bit which is 1 or 0 where 1 indicates that a particular feature is included. For example bit 5 might be chosen to represent image entropy. There is a population of random chromosomes and for each one classification is performed. A new population is generated using parent selection, crossover and mutation. When the parents are selected it is designed so that ones deemed fitter are more likely to be chosen. By fitter it is meant the ones that resulted in a more accurate classification. This is continued for either a certain number of population generations or until a certain level of classification is obtained. It should be noted that there might be bias in the classifier, such that a certain set of input values might favor a particular set of features; to avoid this, the broadest range of data sets should be used.

III. IMPLEMENTATION OF THE EVOLUTIONARY ALGORITHM

A graph based evolutionary algorithm system called Cartesian Genetic Programming (CGP) has been chosen for this work. A more recent form of Genetic Programming developed by Julian Miller [14], it differs from conventional genetic programming in structure that is evolved. Rather than using trees representing computer programs it uses a two dimensional array or network of functions, which can be visualized as being closer to a digital circuit than a program.

More formally, it can be viewed as genetic programming generalized from tress to acyclic graphs where edges mark connections and nodes are functions. The functions here are not in the form of any programming language but can be simple blocks such as an AND gate and some of the earliest uses of CGP have been in evolving digital circuits. A CGP system involves the evolution of these networks to find the optimal one for a problem. Fig. 3 shows an example of the structure but without connections which go from inputs to functions to either more functions or outputs. Note that the

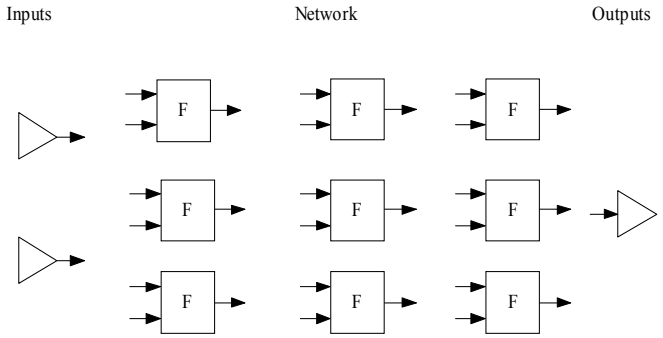

Fig. 3. CGP network structure

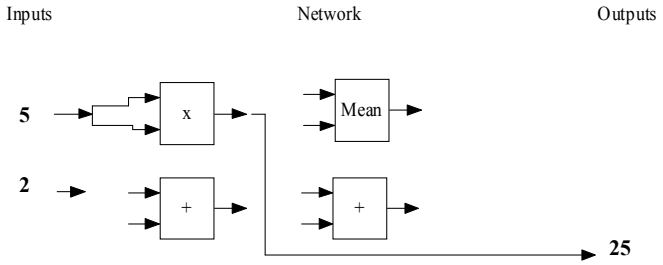

Fig. 4. An example CGP network

signal path can only go towards the output, i.e. from one column to one on its right, not backwards. As an example consider a very simple network with two inputs, one output and size 2x2 (4 functions) as shown in Fig. 4. Assuming each function has two inputs, the network inputs could be the numbers 2 and 5. One function could be multiplication, the second addition, the third a mean function and the fourth another addition. If the output had connected to the first function and both the function's inputs connected to the first network input then the output would be 25.

A CGP network is initially specified by the following parameters:

- Number of rows
- Number of columns
- Number of inputs to the network
- Number of inputs to a node
- Number of network outputs
- Available functions
- Number of columns back a node can connect

Connections are made randomly, and functions are randomly chosen out of the listed available functions. For each network the inputs are applied to the network input and an output is produced. A fitness function is applied to the output to calculate fitness. Once all fitnesses are calculated a new generation is made and this is all repeated as explained earlier for genetic algorithms.

A. Strengths and weaknesses of CGP

It was stated previously that GAs are a search technique finding effective solutions in the solution space. The key observation to make about many engineering solutions is that they are based on set procedures and algorithms.

Design of a digital filter would be based around some key mathematics such as discrete Fourier and Z-transforms. Design of a digital adder is very methodical, building up a circuit from some set block units. The use of these methodologies is essential in engineering in making the problem simple so that people can work with it. However they also cut out many potential solutions.

CGP adopts an evolutionary strategy that provides a randomly guided search in which no assumptions are made regarding the search space, allowing solutions to occur that would not normally result with conventional techniques. One of the most powerful techniques at the moment in image processing is the use of wavelets, effectively a transform similar to Fourier but in this offers both scale and frequency information instead of just spatial domain information or just frequency. However whilst being sophisticated it is still quite conventional with implementation involving blocks of filter banks. For the basis or mother wavelet only one of 3 or 4 families (for example Haar or Daubechies) tends to be used because a lot is known about them, not because it is best for the problem. CGP is far less conventional and solutions could involve what appear to be random arrangements of adders, multipliers, filters, comparators, functions to calculate means or anything else.

CGP is not without its disadvantages, however. Firstly, since the design has not been engineered but rather evolved it can be hard to analyze and so unlike a conventional design it's difficult to know why it works. As a consequence it is difficult to guarantee it will work reliably. There is no guarantee that CGP will find a good solution, for example if it is not given the right functions or enough functions then it will fail. CGP can be very slow to evolve, taking minutes for example to evolve a 4 bit multiplier on a P4 2.66 MHz processor in tests. In complex problems it may prove too slow to be used practically even if it is possible for a solution to be found given the functions available to the network.

CGP has proven useful in image processing already. For example, it has been found to have some effectiveness in removing noise from an image and so it is quite likely that it can be of use in processing mammograms.

B. How CGP can be used for classification

So far a CGP network has been described and also GAs have been described, but how these fit together remains to be seen. The proposed flow for using CGP is therefore presented here:

1. Begin with the pixel locations of a microcalcification
2. Extract features from these pixels and place into an array of features
3. Repeat steps 1 and 2 for all the microcalcifications that are to be used.

4. At this point there are N arrays of features. Split these arrays into a training set and a testing set
5. Select a set of functions which can be used by the CGP networks (arrays of functions)
6. Initialize a population of random CGP networks built using random connections and random functions chosen from within the function set.
7. Take an array of features for a single microcalcification and apply it to a network. Use a threshold function on the output in order to choose between malignant or benign.
8. Repeat 7 for all microcalcifications
9. Compare the actual outputs against the desired ones, i.e. the true pathology and use this to calculate a fitness for each network
10. Repeat 7, 8 and 9 for all networks
11. Using some predetermined evolutionary strategy generate a new population
12. Repeat step 7 to 11 until a perfect fitness has been obtained, i.e. it gets the pathology right on all mammograms or till it has run for a certain amount of time
13. Now do steps 7 to 9 using the testing inputs, applying to the fittest network and analyze the performance

C. Choice of function set

The functions that are available to a CGP network will have an important effect on what fitness that network can ultimately achieve. For instance, if a network was of size 5 columns by 5 rows there would be a total of 25 functions. If all these functions were either a logical AND function or a logical OR function then it is very unlikely the network will be able to perform complex image processing unless a much larger network is utilized.

If instead those 25 functions include exponentials, multiplications, additions then there is a much better chance of complex processing being done. What is best to use is dependent on the problem, using an exponential function is inappropriate if evolving a digital circuit, for example.

In this instance the problem is an image processing one and so it is in fact more likely that a good selection of mathematical functions would be necessary (considering pattern recognition is normally done using involved mathematical techniques). For this project any type of function was considered as the project was being implemented in software and was about proof of principle. It might be that at a later stage that hardware implementation is desirable and this would place restrictions on the functions available. This is not the case here though so practical implementation constraints are not considered. However, run time of the training algorithm was an issue and it should be remembered that selection of a cosine function over an add function might increase run time significantly depending on the processor being used.

Up to a point, choice of function is arbitrary as the whole reason to use the CGP technique is that evolution is being allowed to decide how best to solve the problem and what is best is not known in advance. Some sensible choices though are as follows: *Add and multiply functions* - many filters and other functions can be made out of these; *Comparison functions* - these allow the filtering of one of two values dependent on a condition; *Divide function* - is useful, allowing a large number to be obtained here multiplying two numbers below 0 (as would happen with CGP features) would make even smaller numbers; *Complex functions such as sine and exponential* - making these available could allow easier processing of frequency content. There is no way to be certain what is ideal without experimentation but it is important to ensure the functions available allow a lot of potentially varied and powerful networks to be evolved.Those functions chosen were as follows:

- add – all inputs added
- subtract – inputs 1-N subtracted from input 0 for N inputs
- multiply – calculate the product of all the inputs
- divide – input 0 divided by input 1, returns 1 if second is 0
- greatest – outputs the largest input
- least – outputs the smallest input
- greater than or less than – compares input 0 to input 1, returns 1, -1, or 0 for greater I[0] > I[1], I[0] < I[1], I[0] = I[1] respectively where I is the input vector.
- mean – calculates the mean of the inputs

D. Choice of evolutionary strategy

In generating a new population there are a number of different strategies that can be employed. A tournament strategy is where two chromosomes are picked at a time and the fitter of the two is used in the new population. Crossover can then be applied to this fitter population. The advantage of this technique is that it ensures a lot of diversity in the population. The disadvantage is that it can be slow to converge and requires a large population which is computationally intensive. Therefore a different strategy will be employed known as λ+1. In this case a lambda of 5 will be used keeping the population small. It works as follows:

- Evaluate the fitness of each chromosome
- Select the fittest
- Replicate the fittest 5 times
- Mutate 4 of the 5

This technique can converge quickly and relies purely on mutation for diversity. Because it is only ever keeping the fittest there is a danger of local maxima being found and it must be ensured that there is sufficient mutation to limit this.

E. Fitness function

In order for genetic algorithms to work it is very important to select the correct fitness function. If the designer misinterprets what result they are wanting to achieve then the network may train to a high fitness but will not do what is expected. Sometimes certain features the fitness function may cause unwanted side effect in what is evolved. Fortunately it is possible to define a relatively straightforward fitness function. Since the aim is to correctly classify mammograms then the total fitness can be defined as:

$$\sum_{i=0}^{N} F_i \tag{1}$$

where N is the number of input sets (i.e. microcalcifications or sub images) and F_i is the fitness for a particular microcalcification. F_i is defined as

$$F_i = \begin{cases} 1 & if \quad e_i = a_i \\ 0 & if \quad e_i \neq a_i \end{cases} \tag{2}$$

where e_i is the expected output provided by the threshold and a_i is the actual output for input set i. The aim of the project was to investigate if there was any diagnostic merit in using CGP and so no conditions were added such as making a false negative more serious than a false positive or vice versa.

F. Features

As described in part B of this section, image features are the inputs to the CGP network and it is only by these features that the network is able to distinguish a benign mammogram from a malignant one. From the pixels the following features were selected:

- Mean
- Second Moment
- Third Moment

The spatial grey level dependence (SGLD - sometimes called grey level co-occurrence matrix) was also extracted and from this the following features were extracted:

- Mean
- Second Moment
- Third Moment
- Entropy
- Element difference moment
- Uniformity

More information and definitions of these features can be found in [12]. All the features selected are texture based features which were found in [8] to be more effective in classification than morphological ones.

IV. RESULTS

For the first set of results, 17 images were used for training the CGP network and 12 images for testing in one data set and 27 images used for training and 19 images used for testing the second data set. The CGP parameters used are shown in Table 1 and were determined through trial and error for the purpose of these preliminary experiments, but future work will investigate evolving optimal parameters. Run time for the algorithm was in the region of one and a half hours on a Pentium 4, 2.66GHz processor (pre HyperThreading generation) with 1 GB of RAM, using the parameters listed in Table 1. The results of the experiment are given in Table 2 and are averaged over 3 runs for each of the data sets.

TABLE1
CGP PARAMETERS – EXPERIMENT 1

Parameter	Setting
Number of rows	5
Number of columns	16
Number of inputs per node	2
Number of generations	14000
Mutation rate	0.1

TABLE 2
RESULTS - EXPERIMENT 1

Data set	Training Fitness	Testing Fitness
1	71% (12/17)	67% (8/12)
2	81% (22/27)	63% (12/19)

A round robin system was used whereby all inputs sets but one are used for training and the remaining one is used for testing. The testing set is then returned to the training set and another is used for testing. This continues until all items have been used for testing or for as long as is deemed necessary (e.g. due to time constraints). This is a standard technique often used in the literature (for example in [9]). In total, 20 combinations were used here and the results are given in Table 3.

TABLE 3
ROUND ROBIN PERFORMANCE – EXPERIMENT 2

Data set	Sensitivity	Specificity	FPR	FNR	Training Fitness	Test Fitness
1	0.9	0.6	0.4	0.1	80% (41/51)	70% (14/20)

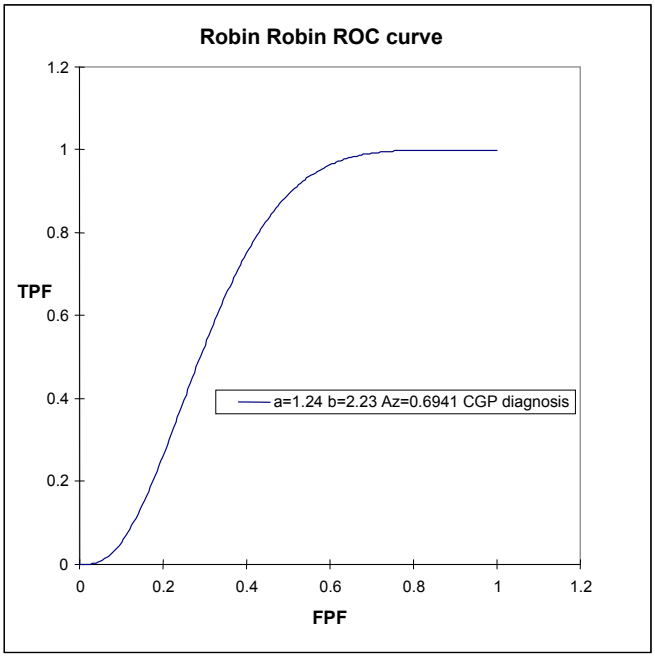

Fig. 5. Receiver Operating Characteristic (ROC) curve for round robin performance (Experiment 2). TPF – True Positive Fraction, FPF – False Positive Fraction

Significant performance can be seen throughout these results with high sensitivity and specificity. Training results are also strong here even though a much larger data set of 51 images has been used for training each time. Overall classification accuracy is also good at 70%. The Receiver Operating Characteristic (ROC) curve analysis given in Fig. 5 shows an area under the ROC A_z of 0.69 which is comparable to values found in the literature. For example, review paper [11] found two neural network classifiers to have A_zs of 0.74 and 0.6 and a K-nearest neighbor method to produce an area of 0.82 for grey level features and 0.72 for SGLD based features as used in the CGP method described in this paper.

Results up to this point were based around feeding an array of features into a CGP network's input. This technique is a very conventional way of doing the classification in that it is very similar to how a neural network is used, although clearly

the system itself will behave differently. The experiment evaluates performance of CGP by using the network in a very different way. In this case no features are extracted from the mammogram, but a pixel region of size 8x8 is fed directly to the network (after having first linearised the array).

TABLE 4
CGP PARAMETERS – EXPERIMENT 2

Parameter	Settings 1	Settings 2
Number of rows	16	16
Number of columns	16	16
Number of inputs per node	2	2
Number of generations	8000	20000
Mutation rate	0.1	0.1

The experimentation for this approach was carried out with 2 sets of CGP parameters which are presented in Table 4. Note that again the network size has been significantly increased. This is because there are now 64 inputs as opposed to the 15 used before. The number of generations was further increased for the latter 3 runs since the first suggested they were insufficient. Four runs were carried out in total.

The results for the 4 runs are presented in Table 5. The training scores are a lot lower than before, high sensitivities are countered by low specificities and so the overall classification is always higher than chance but not significantly.

TABLE 5

RESULTS - EXPERIMENT 2

Data set	Sensitivity	Specificity	FPR	FNR	Training Fitness	Test Fitness
1	0.9	0.22	0.77	0.1	55%	58%
2	0.9	0.44	0.55	0.1	57%	65%
2	0.9	0.3	0.7	0.1	59%	60%
2	0.8	0.3	0,7	0.2	57%	58%

A ROC analysis indicated the technique is more effective than the sensitivities and specificities would suggest. The ROC is plotted in Fig. 6 for the second row in Table 5 which is the run with the most effective classification. A_z is close to 0.78 which is comparable to, if not better than, performance in commercial classifiers, although it is lower in the other runs, being closer to 0.7 and below. Thus, some potential is shown but with a lot more work needed.

V. CONCLUSIONS

In this paper a novel application of an evolutionary algorithm, Cartesian genetic programming (CGP) has been applied to the classification of microcalcifications segmented from mammograms. CGP was used not only to optimize a number extracted features from the image, but also classify the microcalcifications based on the raw pixel values. This is effectively evolving an algorithm to extract and select features and classify the image accordingly. Initial results in both cases were variable in performance but reached levels comparable to those in the literature. The data set was relatively small and so the statistical significance of the results is limited. However, it does indicate that CGP has potential in the classification of mammograms and further work should be carried out. In addition, it indicates that the novel technique of classifying purely on raw pixels as opposed to features may be effective and should also be investigated further. With more sophisticated segmentation, extraction of further features (e.g. morphological), a larger data set and refinement of the CGP algorithm, it is anticipated a significant increase in performance can be achieved.

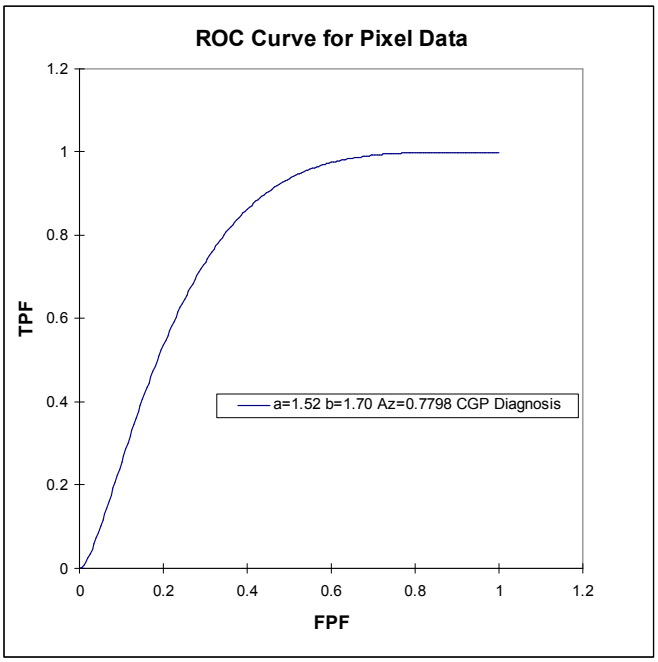

Fig. 6. Receiver Operating Characteristic (ROC) curve for pixel data set 2. TPF – True Positive Fraction, FPF – False Positive Fraction.

REFERENCES

[1] "Breast Cancer Incidence rises while deaths continue to fall" National Statistics website: http://www.statistics.gov.uk/cci/nugget.asp?id=575

[2] Y. Jiang, C. Metz, R. Nishikawa and R. Schmidt, "Comparison of Independent Double Readings and Computer-Aided Diagnosis (CAD) for the Diagnosis of Breast Calcifications", *Academic Radiology,* vol 13, no. 1, 2006.

[3] "New technologies to help improve mammography," imaginis website: http://imaginis.com/breasthealth/cad.asp

[4] D. Gur, J. Stalder, L. Hardesty, B. Zheng, J. Sumkin, D. Chough, B. Shindel, and H. Rockette, "Computer-aided Detection Performance in Mammographic Examination of Masses: Assessment," *Radiology,* vol 233, 2004, pp. 418-423.

[5] V. Andolina, S. Lille and K. Willison, *Mammographic Imaging, A Practical Guide,* Lippincott-Raven, 1993.

[6] Heath M., Bowyer K.W., Kopans D. et al., "Current status of the Digital Database for Screening Mammography," in *Digital Mammography*, Kluwer Academic Publishers, 1998.

[7] M. Gavrielides, J. Lo, R. Vargas-Voracek and C. Floyd, "Segmentation of suspicious clustered microcalcifications in mammograms," *Medical Physics*, vol 27, 2000.

[8] H. Chan, B. Sahiner, K. Lam, N. Petrick, M. Helvie, M. Goodsitt, and D. Adler, "Computerized analysis of mammographic microcalcifications in morphological and texture feature space," *Medical Physics*, vol. 25, no.10, 1998.

[9] J. Kim and H. Park, "Statistical Textural Features for Detection of Microcalcifications in Digitized Mammograms," *IEEE Transactions Medical Imaging,* vol. 18, no.3, 1999, pp 231-238.

[10] J.C. Fu, S.K. Lee, S.T.C. Wong, J.Y. Yeh, A.H. Wang and H.K. Wu, "Image segmentation feature selection and pattern classification for mammographic microcalcifications," *Computerized Medical Imaging and Graphics,* vol. 29, 2005, pp. 419-429.

[11] H.D. Cheng, X. Cai, X. Chen and L. Hu, X. Lou, "Computer-aided detection and classifcation of microcalcifications in mammograms: a survey," *Pattern Recognition*, vol. 36, 2003, pp. 2967 – 2991.

[12] R. Gonzalez and R. Woods, *Digital Image Processing*, Prentice Hall, 2002.

[13] X. Cai, S. L. Smith and A. M. Tyrrell, "Benefits of Employing an Implicit Context Representation on Hardware Geometry of CGP," Lecture Notes in Computer Science, vol 3637, 2005, pp. 143-154.

[14] J. F. Miller and P. Thomson, "Cartesian Genetic Programming," Lecture Notes in Computer Science, vol 1802, 2000, pp. 121-132.

.