Multiple Network CGP for the Classification of Mammograms

Katharina Völk[1], Julian F. Miller[1] and Stephen L. Smith[1]

[1] The University of York, Department of Electronics, Heslington, York YO10 5DD, UK
{kv500, jfm7, sls5}@ohm.york.ac.uk

Abstract. This paper presents a novel representation of Cartesian genetic programming (CGP) in which multiple networks are used in the classification of high resolution X-rays of the breast, known as mammograms. CGP networks are used in a number of different recombination strategies and results are presented for mammograms taken from the Lawrence Livermore National Laboratory database.

Keywords: Evolutionary algorithms, Cartesian genetic programming, mammography

1 Introduction

Breast cancer is one of the leading causes of death in women in the western world. In 2005, some 40,000 new cases were detected in the UK alone and 10,300 women died in 2005 as a result of the disease [1] making breast cancer the most common cancer in women and the second most common cause of cancer death. The number of breast cancer related deaths has fallen since a screening programme was introduced in 1988. Nonetheless, it is predicted that one in seven woman will develop breast cancer at some point during their life [2].

The detection of breast cancer in the early stages of the disease significantly increases the survival rate of patients. The main method for screening patients is the mammogram, a high resolution x-ray of the breast. The process of identifying and evaluating signs of cancer from mammograms is a very difficult and time-consuming task that requires skilled and experienced radiologists. This assessment is also, by its nature, highly subjective and susceptible to error, leading to cancers being missed and the patients misdiagnosed. To achieve a more accurate and reliable diagnosis, Computer Aided Detection (CAD) systems have been investigated which provide an objective, quantitative evaluation. CAD systems have the potential to help in two main ways: (i) the detection of suspicious areas in the mammogram that require further investigation and (ii) the classification of such areas as cancerous (malignant) or non-cancerous (benign).

The aim of the work reported in this paper is to assess the potential benefit of a new representation of using evolutionary algorithm 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 algorithm 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. 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 can have a great variety of mostly benign causes, but might also be an indication for malignancy. They are fairly common on mammograms and their appearance increases with age, so that they can be found in 8% of mammograms of women in their late 20s and in 86% of mammograms of woman in their late 70s [2]. Microcalcifications that indicate malignancy are usually less than 0.5mm in size and often grouped into clusters of five or more. Any calcification larger than 1mm is almost always benign [2].

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 work by the authors indicate 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 provide a more reliable source of data for both training and testing the algorithms.

Due to the nature of the mammograms, traditionally there is a need for pre-processing, particularly for detection or classification where methods such as wavelets or morphological filtering are used. One of the advantages of using an evolutionary method such as Cartesian Genetic Programming (CGP) is that there is no need for prepossessing the images.

Previous work undertaken in the classification of microcalcifications using both traditional image analysis techniques and evolutionary algorithms is considered in Section 2. The evolutionary algorithm used in the current work will then be described in Section 3 and results applying this technique to a number of digitized mammograms will be considered in Section 4. Finally, the potential of the proposed algorithm will be evaluated in Section 5.

2 Previous work

Over recent years there has been much research into the application of computer aided diagnosis 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 that employs (i) sensing, (ii) segmentation, (ii) feature extraction, (iv) feature selection and (v) classification, to isolate and then characterize a feature of interest. 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 [3], are not considered here and for the purpose of the experiments described in Section 4 will be undertaken manually.

2.1 Feature Extraction

Once segmentation is completed any microcalcifications located needs 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 the 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 [4] provides information for radiologists about the varying features of benign and malignant microcalcifications. For example it advises that benign examples have a round ring like shape with well defined borders. Malignant microcalcifications on the other hand have varying shape and poorly defined borders. Such characteristics can be described using morphological feature extraction. Reference [5] 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 and 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 and angular second moment. Another neural network based paper [6] relied purely on texture features concentrating on ones from the SGLD matrix.

An alternative method suggested in [7] uses the discrete cosine transform of the image to derive “block activity and spectral entropy from the DCT coefficients”. Reference [8] also gives brief mention of Fourier methods and a wavelet method whereby standard features (such as 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, often referred to as a multiscale method.

2.2 Feature Selection

At the end of the feature extraction stage there may be a very large number of features and whether a statistical classifier, neural network or a genetic algorithm (as will be the case in this investigation) is to be used, it is helpful to reduce the number of features. 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 [5]. 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, to the between group one, 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" [5].

2.3 Classification

In the case of breast cancer, the classifier determines 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 [8] and listed here:

• Neural networks: a parallel information processing network based on the structure of neurones. It is noted in [8] 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 neighbours: 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 that a given pattern x belongs to a class wj indicating, for example, malignancy. 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 [9].

2.4 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 include Genetic Algorithms, 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 [5] 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 [5] 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 [10]. 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 algorithms 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.

3. Implementation

Cartesian Genetic Programming (CGP) has been used in this work [11]. CGP is a graph-based genetic programming system which has been shown to perform well within a wide range of problem domains. A CGP solution consists of an n-dimensional grid (where n is typically 1 or 2) in which each grid location contains a function. Program inputs and outputs are delivered to and taken from specific grid cells. Interconnections between functions, inputs and outputs are expressed in terms of the grid's Cartesian co-ordinate system. The 2-D CGP general form is shown in

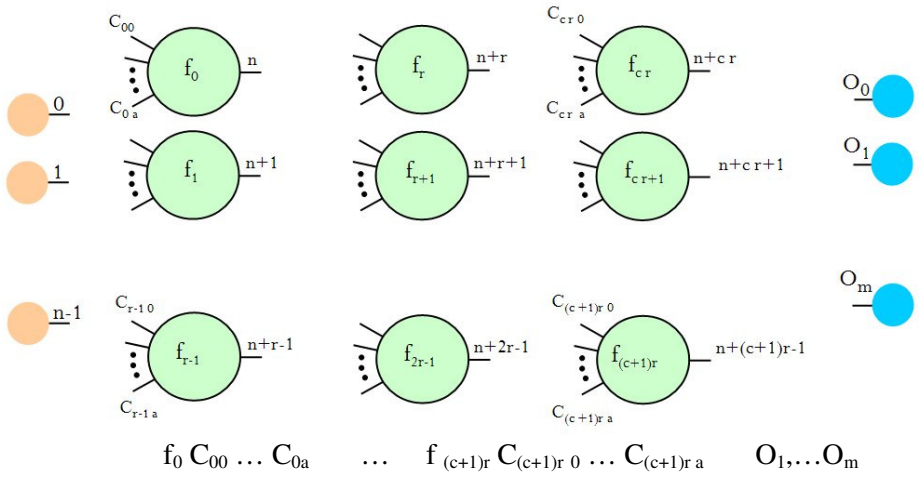

Fig 1. General form of two-dimensional CGP. It is a grid of nodes whose functions are chosen from a set of primitive functions. Each node is assumed to take as many inputs as the maximum function arity a. Every data input and node output are labelled consecutively (starting at 0) which gives it a unique data address which specifies where the input data or node output value can be accessed (shown in the figure on outputs of inputs and nodes). Nodes in columns cannot be connected to each other. In most cases the graph is directed (as in this paper) so that a node may only have its inputs connected to either input data or the output of a node in a previous column. In general there may be a number of output genes (O_i) which specify where the program outputs are taken from. The structure of the genotype is seen below the schematic. All node functions genes f_i are integer addresses in a look-up table of functions. All connection genes C_{ij} are integers taking values between 0 and the address of the node at the bottom of the previous column of nodes.

A mutation operator can alter both the function present within a grid cell and the connections between components. The efficacy of CGP has been attributed to both implicit reuse of sub-expressions (due to its graphical representation) and its use of functional redundancy [11][14]. In this paper multiple CGP networks are used to evolve a single mammogram. The image is divided into equal sized, non-overlapping parts and each one of these is assigned its own CGP network which is evaluated independently as shown. The division of the mammogram into parts should not, however, be confused with a conventional image processing windowing operation.

More specifically, in this work each genotype consists of 256 independent CGP chromosomes using a grid of 32 rows and 128 columns. Each chromosome has 64 inputs corresponding to an 8x8 grid of grey scale pixel values (0 to 255). Each chromosome has a single output gene. The mammogram images are divided into 256, 8x8 pixel areas. The 64 grey scale pixel values (0 to 255) for each of these areas form the inputs to an individual CGP network, encoded by a chromosome. A summary of these details is given in Table 1. The chromosome mutation rate defines the percentage of genes in each chromosome that are mutated when a genotype is mutated (i.e. for 32x128 nodes there are 3x32x128+1 genes. The one corresponds to the single output gene). The function set is shown in Table 1 (where x&y represents the bitwise AND function). The output from all node operations is kept within the range 0 to 255, by truncation.

Table 1. Parameters for multiple CGP network

Parameter	Value
No. parts per image	256 (16x16)
Part size	8x8 pixels
Chromosome rearrangement rate	3%
Chromosome mutation rate	1%
No. runs	10
No. generations	1000
No. columns in each CGP network	128
No. rows in each CGP network	32
Function set	x, x+y, abs(x-y), abs(2x-y), x&y, largest(x,y), smallest(x,y)

One of the motivations for developing this representation was to implement recombination in CGP but at a whole chromosome level as opposed to recombining genes within chromosomes. This multi-chromosome approach has been shown to have a number of advantages [12]. So whole CGP chromosomes could be exchanged rather than components (genes) within one network. This approach not only aims to improve the system's performance on one given part of an mammogram but also allows for improvement of the system's performance on the whole mammogram by allowing individual CGP chromosomes to be swapped and reused for other parts of the image depending on the success or fitness they achieve.

The evolutionary algorithm is a 1+2-ES, in which there are three genotypes (each consisting of 256 chromosomes). There will be one parent selected whose genotype will consist of 256 chromosomes, each of which is the best chromosome chosen from

the three population members. However, we have investigated another mutational step after this best genotype is assembled. According to another mutation rate, which we call a re-arrangement mutation rate (see table 1), chromosomes may undergo either a swap or replacement with another of the 256 chromosomes, chosen at random. Specifically:

(i) a random swap in which any chromosome might be swapped with another (random-swap);

(ii) a neighbouring swap in which a part might only be swapped at random with its four direct spatial neighbours. The neighbouring swap has been implemented to target structures that continue from one part of the image to the next. Neighbouring parts might therefore have similar image properties and are therefore likely to respond equally well to the same chromosome. (neighbourhood swap);

(iii) a copying operation where a random chromosome is chosen to overwrite a different chromosome (re-use).

If after a rearrangement a chromosomes fitness declines from the operations then the rearrangement is disallowed. If, however, the rearranged part makes an improvement to the resulting fitness, then the exchange is preserved. Although there is a risk that the diversity of chromosomes might be reduced by deleting ones that do not perform well and substituting them with a copies of a fitter one, this approach gives the genotypes a higher opportunity for individual mutation which in itself has the potential of restoring diversity to some extent. We can see this because if every chromosome is unique (no copies) then mutations can only be beneficial independently. If there are duplicated chromosomes, any mutation occurring in those would have to be, on average, beneficial to all of them. This means one chromosomes fitness might be reduced if all other copies gained a higher fitness, through the rearrangement.

Images used in this study are constructed from mammograms in the LLNL database that feature microcalcifications. As described in the introduction the images have been manually edited to avoid the need to automatically locate microcalcifications. In each case a 128x128 pixel image is constructed containing at least one microcalcification from a particular mammogram. Each image is then logically divided into 256 parts and the status of each part labelled as either being benign or malignant according to the radiologist.

When an image is processed by the system the output value generated for each chromosome by its respective CGP network is compared to a predetermined threshold. An output value above the threshold is interpreted as an indication of malignancy and an output value below the threshold an indication of benignity. In this study output values ranged from 0 to 255 and the threshold adopted was 4. This bias toward benign results reflects the relative scarcity malignant areas within the image. A fitness value can be then be calculated on this basis of this predicted value and the predetermined status of that part of the image as identified by the radiologist.

4 Results

As previously stated, the mammograms used in this study were taken from the Lawrence Livermore National Laboratory database [13] that specifically featured microcalcifications. The mammograms were cropped to images of 128x128 pixels containing at least one microcalcification. In total 31 images were created, of which 13 contained malignant microcalcifications and 18 benign microcalcifications. Some 67% of these images were used for training the CGP network and the remaining 33% for the testing stage.

In the training phase, the single parts reached different fitness values with an average of 81.2% to 90.6% depending on the method used. Full results are given in Table 2 which also details the performance of a number of chromosome rearrangement strategies. Graphs for average and best fitness are also given in Figures 2 and 3 respectively.

Table 2. Fitness values for training of CGP networks

Recombination	Best part's fitness (%)	Worst part's fitness (%)	Average fitness (%)
No swap, no reuse	94.9	54.5	81.2
No swap, reuse	95.7	67.8	85.5
Neighbouring swap, no reuse	96.9	62.0	87.8
Neighbouring swap, reuse	96.9	63.9	89.0
Random swap, no reuse	96.5	70.6	89.4
Random swap, reuse	96.9	71.4	90.6

One of the problems that might occur when applying the evolved programme to test images is that some of the CGP networks may not have been trained with image parts containing a microcalcification and therefore will only recognise background breast tissue. To overcome this problem each part of the test image is evaluated with every evolved CGP chromosome. The highest fitness score generated is then used to classify that respective part (without any knowledge of its true class). The fitness values of the test set of images are shown in Table 3.

Table 3. Fitness values for test images

	Best part's fitness (%)	Worst part's fitness (%)	Average fitness (%)
No swap, no reuse	97.3	84.7	93.3
No swap, reuse	98.0	81.6	94.1
Neighbouring swap, no reuse	98.4	82.0	94.5
Neighbouring swap, reuse	98.4	81.2	94.5
Random swap, no reuse	98.0	82.7	94.5
Random swap, reuse	98.4	81.2	94.9

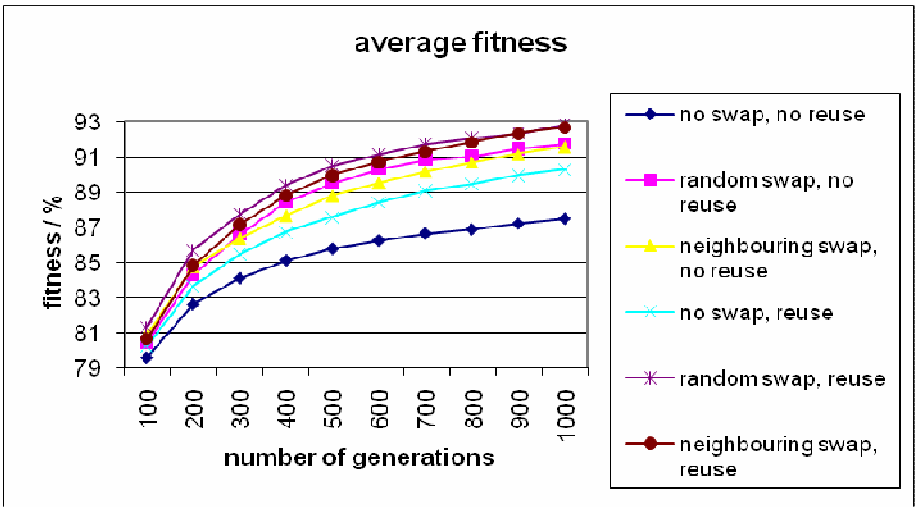

Fig 2. Average fitness for training phase of CGP networks with different recombination strategies.

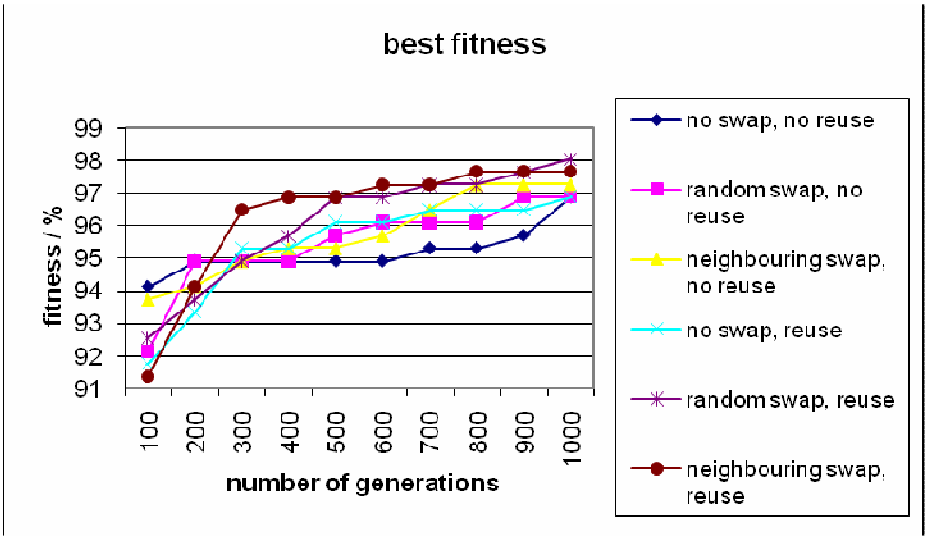

Fig 3. Best fitness for training phase of CGP networks with different recombination strategies.

An example result is shown in Figure 4. Each figure represents an image part (64 pixels), the number "1" indicates a malignancy in that area and number "0" indicates a benignancy. The radiologist's classification of malignancy is indicated by grey shading. Figure 4 shows the respective mammogram.

0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0
0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0
0 1 0 1 1 0 0 0 0 0 0 0 0 0 0 0
0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 1 0 1 1 1 1 0 0 0 0 0 0 0 0
0 0 0 1 1 1 1 1 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

Fig. 4. System's classification of image parts 1 = malignancy, 0 = benignancy. Grey shading radiologist's classification of malignancy.

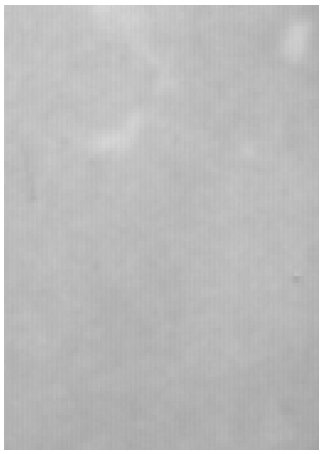

Fig. 5. Mammogram relating to results shown in Figs. 2 and 3.

5 Conclusion

This paper has described a novel multiple network CGP evolutionary algorithm applied to the classification of mammograms. The results presented have demonstrated that the method correctly classifies microcalcifications as being malignant or benign. Given the limitations of the training and the test sets available, and no pre-processing has needed to been applied, the results are very encouraging. The main limitation of the data available is the low number of usable images from a fairly old database. To overcome this problem, new databases of mammograms are being sought so further work to evaluate and improve the system can be conducted.

References

1. Office for National Statistics. http://www.statistics.gov.uk/cci/nugget.asp?id=575
2. Kopans, D.B., Lippincott Williams & Wilkins (2006)
3. Gavrielides, M. Lo, J., Vargas-Voracek, R. and Floyd, C. Segmentation of suspicious clustered microcalcifications in mammograms. Medical Physics, vol 27 (2000)

4. Andolina, V., Lille, S. and Willison, K. Mammographic Imaging: A Practical Guide. Lippincott-Raven (1993)
5. Chan, H., Sahiner, B., Lam, K., Petrick, N., Helvie, M., Goodsitt, M. and Adler, D. Computerized analysis of mammographic microcalcifications in morphological and texture feature space. Medical Physics, vol. 25, no.10 (1998)
6. Kim, J. and Park, H. Statistical Textural Features for Detection of Microcalcifications in Digitized Mammograms. IEEE Transactions Medical Imaging, vol. 18, no.3, pp 231-238 (1999)
7. Fu, J.C. , Lee, S.K., Wong, S.T.C., Yeh, J.Y., Wang, A.H. and Wu, H.K. Image segmentation feature selection and pattern classification for mammographic microcalcifications. Computerized Medical Imaging and Graphics, vol. 29, pp. 419-429 (2005)
8. Cheng, H.D., Cai, X., Chen, X. and Hu, L., Lou, X. Computer-aided detection and classification of microcalcifications in mammograms: a survey. Pattern Recognition, vol. 36, pp. 2967 – 2991 (2003)
9. Gonzalez, R. and Woods, R. Digital Image Processing, Prentice Hall (2002)
10. Cai, X., Smith, S.L. and Tyrrell, A.M. Benefits of Employing an Implicit Context Representation on Hardware Geometry of CGP. Lecture Notes in Computer Science, vol 3637, pp. 143-154 (2005)
11. Miller, J. F. and Thomson, P.: Cartesian Genetic Programming, Lecture Notes in Computer Science, Vol. 1802, pp.121-132 (2000)
12. Walker, J., Walker J. A., Miller J. F., Cavill R. A Multi-chromosome Approach to Standard and Embedded Cartesian Genetic Programming, Proceedings of the Genetic and Evolutionary Computation Conference (GECCO), ACM Press, (2006) 903-910.
13. Lawrence Livermore National Laboratory database. https://www.llnl.gov/
14. Miller J.F., Smith S.L. Redundancy and Computational Efficiency in Cartesian Genetic Programming. IEEE Transactions on Evolutionary Computation, **10** (2006) 167-174