

A hybrid evolutionary algorithm for multi category feature selection in Breast Cancer recognition

Ranadhir Ghosh, Moumita Ghosh, John Yearwood

School of Information Technology and Mathematical Sciences, University of Ballarat

PO Box 663, Victoria

Australia

{r.ghosh, m.ghosh, j.yearwood} @ballarat.edu.au

Abstract- Many previous efforts have utilized many different approaches for recognition in breast cancer detection using various ANN classifier-modeling techniques. Different types of feature extraction techniques have also been used. It has been observed that, beyond a certain point, the inclusion of additional features leads to a worse rather than better performance. Moreover, the choice of features to represent the patterns affects several aspects of pattern recognition problems such as accuracy, required learning time and a necessary number of samples. A common problem with the multi-category feature classification is the conflict between the categories. None of the feasible solutions allow simultaneous optimal solution for all categories. In order to find an optimal solution the search space can be divided based on an individual category in each sub region and finally merging them through decision support system. Combining the feature selection with the classifier has been a major challenge for the researchers. A similar technique employed in both the levels often worsens their performance. Some preliminary studies has revealed that while using traditional canonical GA has been a good choice for feature selection modules, however under perform for the classifier level module. An evolutionary based algorithm for the classifier level provides a much better solution for this purpose. In this paper we propose a hybrid canonical based feature extraction technique with a combination of evolutionary algorithm based classifier using a feed forward MLP model.

I. INTRODUCTION

Breast cancer is a primary cause of death in women. Early detection and diagnosis of breast cancer improves the chance of survival. However late detection and diagnosis often leads patient to unrecoverable stage of cancer ending in casualty. X-ray mammography is currently the most popular, cost-effective, low radiation dose and relatively accurate method of early detection of the disease [1]. The radiographs are searched for signs of abnormality by expert radiologists but mammograms are complex in appearance and signs of early disease are often small or subtle. That is the main reason of many missed diagnoses that can be mainly attributed to human factors [2,3]. Since the consequences of errors are costly, there has been a considerable interest in developing methods for automatically classifying mammography abnormalities, as a means of aiding radiologists by improving the efficacy of screening programs and avoiding unnecessary biopsies. Neural network computer-aided diagnosis for

detecting cases in mammograms, such as microcalcifications, has already been used [4-7].

In general, feature selection algorithms have two components: an evaluation function that scores candidate feature sets, and a search engine for finding those sets. Given a set of features the selection algorithm will examine a series of sets of features, and choose the one that maximizes the evaluation function. Recent comparative studies of feature selection algorithms can be found in [8], [9].

In practical pattern recognition problems, a classification function learns through an inductive learning algorithm that maps a given input pattern to one of the existing classes of the systems. However the classifier can work well when a meaningful set of input feature is provided to it. A single feature such as statistical or structural alone may not be the best possible choice. Hence a combination of different categories of features from the raw data set might provide very useful information for the classifier. This type of combination however leads to the formulation of multi-category features as input set. In addition the length of the feature vector thus increases. It has been observed that, beyond a certain point, the inclusion of additional features leads to a decrease rather than increase in performance. Moreover, the choice of features to represent the patterns affects several aspects of pattern recognition problems such as accuracy, required learning time and necessary number of samples. Therefore the main goal of feature subset selection is to reduce the number of features used in the classification while maintaining acceptable classification accuracy. While a standard canonical GA perform very well for selection of features, the same strategy under perform at classification level. Though a traditional MLP provides a good solution model for a classifier, but training it using the same strategy does not perform well when evaluating the fitness function of the chromosome iteratively. The traditional GA based algorithm fails because of problems such as competing convention where as traditional local gradient based learning fails because of its high sensitiveness for initial condition. Hence to maintain the proper convergence for the fitness of the population pool it is not only vital for the classifier-learning algorithm to be least sensitive as possible and also a faster training time.

In this section we explain the multi-category framework for the feature selection level. A common problem with multi-category feature classification is the conflict between the categories. None of the feasible solutions allow simultaneous optimal solution for all categories. Whether an optimal solution for all categories leads to an optimal solution for one combined set of mixed multicategory features can be another research question.

In this paper we propose a canonical GA based modular feature selection approach combined with standard MLP, which is capable of handling multi category features for the classifier. In order to find an optimal solution, the search space is divided based on an individual category in each sub region and finally merging them through a decision support system. We argue that the modular selection works much better than general selection in several aspects as follows:

Independence: The selection modules work independently. Hence each category of feature can be trained and tested independently and in parallel,

Recombination: Crossover combines two parent chromosomes to produce a new offspring. The idea behind crossover is that the new chromosome may be better than both of the parents if it takes the best characteristics from each of the parents. In general selection, each category of features will be treated uniformly. For single characteristics or category this will not cause any problem. But for multiple characteristics of feature, different characteristics will be combined together to produce the offspring. There could be a chance have mix offspring in next generation that can mislead the results.

Time Complexity: As we are dividing the search space into different sub-regions the time for modular selection to reach an optimal solution will be much faster than general selection. A good parallel implementation of the algorithm can have a much better time complexity than the general selection method.

II. METHODOLOGY

The research methodology can broadly be classified into four modules, such as Preprocessing, Feature extraction, Feature subset selection, and Neural network based classification.

A. Pre processing

We use the Digital Database for Screening Mammography (DDSM) dataset from the University of South Florida. Each volume is a collection of cases of the corresponding type. Each case contains four mammograms from a screening exam. Once the digital mammogram is decompressed, suspicious areas are extracted from the mammogram. Suspicious areas are marked in all digital mammograms of DDSM by three expert radiologists.

B. Feature Extraction

All together 40 features have been extracted that are based on texture. These features can be sub divided into three categories such as statistical, structural and grey level dependency. Statistical descriptors include the mean, standard deviation, skewness etc. These describe the distribution of the gray level on that area. Structural descriptors contain energy, entropy, histogram, contrasts etc. These give an indication of how the grey levels are distributed. Grey level dependency is related to the spatial grey level dependence matrix [1]. The matrix is constructed by counting the number of occurrences of pixel pairs at a given displacement.

C. Feature Subset Selection

The general framework is described in Figure 1. Each of the modules works independently on its own domain. They are built and trained for its specific task. Each of them is responsible to find out the best combination of features from each category. The final decision is made on the results of the individual networks, often called expert networks or agents. The decision system is a Neural Network that is responsible for classifying the input.

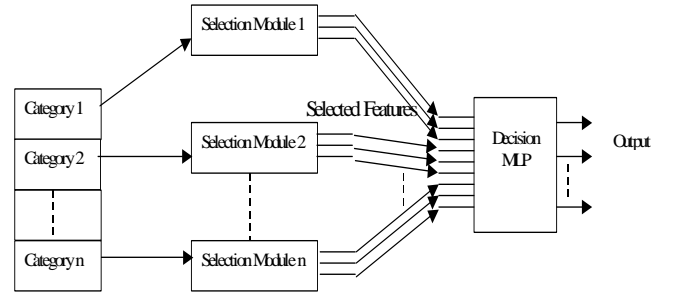


Figure 1 Architecture Framework

C. Selection Module

As described earlier, each selection module is responsible to select the best combination from a given set of features as input. Feature selection algorithms have two components: an evaluation function that scores candidate feature sets, and a search engine for finding those sets. The training phase and the evaluation phase work together (Figure 2). In the evaluation phase the population is initialized randomly. For each member in the population, if the bit position holds a zero value the feature is assigned to zero and a new data set is created. With that dataset the neural network is trained. So, for individual members in the population, there are individual neural networks that have to be trained with the separate dataset. The traditional EBP algorithm is used to train the neural network, then that trained neural network is used to calculate the fitness. To calculate the fitness of the individual population, the feature vector is multiplied by the individual population.

If a particular feature is not selected, that place is neutralized. So the feature is multiplied by zero and deactivates its effect on fitness. The stopping condition for training the neural network is to be the same for all the members in the population and it is taken as the classification error. The stopping criterion of the genetic algorithm is the number of generations.

D. Decision Module

A decision module is responsible for classifying the results on the basis of output of each selection module. We use a Neural Network as a decision system. Output from each selection module is fed to the decision NN. Depending on the feature selected from the different selection module, the decision neural network classifies the input pattern in three classes (Malignant, Benign, Normal).

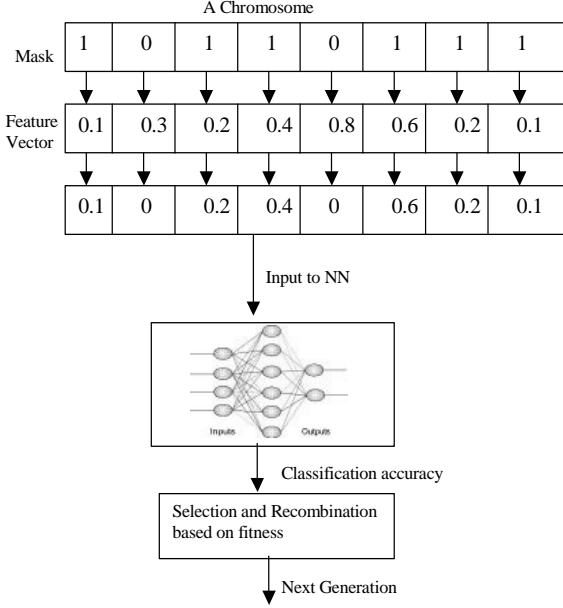


Figure 2 Selection Modules

E. Evolutionary strategy based classifier module

Evolutionary algorithms (EAs) are search methods that take their inspiration from natural selection and survival of the fittest in the biological world. EAs differ from more traditional optimization techniques in that they involve a search from a "population" of solutions, not from a single point. Each iteration of an EA involves a competitive selection that rejects the poor solutions. The solutions with high "fitness" are "recombined" with other solutions by swapping parts of a solution with another. Solutions are also "mutated" by making a small change to a single element of the solution. Recombination and mutation are used to generate new solutions that are biased towards regions of the space for which good solutions have already been seen.

Let $W = (W_h, W_o)$ be an n dimensional solution vector and σ be the corresponding step size. Let m be the number of population in a generation where each population is the pair (W_e, σ_e) .

In the first generation m populations are generated randomly. In the subsequent generations the population set is created by selection and mutation. The Evolutionary algorithm is as follows

- **Step1:** Randomly initialize m population vector.
- **Step2:** The parents are mutated as follows
for $j = 1, 2, \dots, n$

$$\sigma'_i(j) = \sigma_i(j) \exp(\tau' N(0,1) + \tau N_j(0,1))$$

$$W'_i(j) = W_i(j) + S \sigma'_i(j) N_j(0,1)$$

where the values of τ' and τ as follows

$$\tau' = \frac{1}{\sqrt{2n}}$$

$$\tau = \frac{1}{\sqrt{(2\sqrt{n})}}$$

$W_i(j)$, $W'_i(j)$, $\sigma_i(j)$, and $\sigma'_i(j)$ denote the j th component of the vectors W_i , W'_i , σ_i , σ'_i respectively. $N(0,1)$ denotes a normally distributed one-dimensional random number with mean and variance of 0 and 1 respectively. $N_j(0,1)$ denotes that the random number is generated a new for each value of j . S is a sign variance operator, the value of which is considered from a normal distribution $N(0,1)$.

- **Step3:** Calculate fitness of individual population in the generation.
- **Step4:** Create new generation by extracting members of the current population using roulette wheel selection scheme.
- **Step5:** If stopping criteria are satisfied stop else go to Step 2.

III. EXPERIMENTAL RESULTS

The proposed approach has been implemented in C++ and UNIX. We have used 100 cases of each Malignant, Benign and Normal for training. Hence the length of the training dataset was 300. Also we have used 20 cases of each Malignant, Benign and Normal for testing. Hence the length of the testing dataset was 60. The RMS error goal and the number of generations were fixed for all chromosomes to train the network.

The experimental results are shown in Table 1. The percentage classification accuracy given in table 1 is a 10-point cross validation result. The results are compared with the Modular network with the Decision Module trained by Back propagation algorithm [10].

Table1 Experimental Results

Model	Malignant	Benign	Normal	Type I Error	Type II Error	Total
Modular -EA	91	90	88	3.33	8.3	88.33
Modular -BP	90	85	85	3.33	10	86.66
General	80	70	70	6.66	20	73.33

The Table (Table 1) shows that the results have improved both for training and testing dataset.

The following table (Table 2) shows the time take for each algorithm to converge to a solution.

Table2 Time Complexity results

Model	Training Time (m) ¹
Modular-EA	110
Modular-BP	92
General	140

The above table (Table 2) shows that the time take for the Modular Network with Evolutionary approach as a decision module is much higher than that with Back-propagation algorithm. This is because the back propagation algorithm is a local search approach which converge much faster than the global search Evolutionary algorithm. But there is a clear trade off between quality of solution and the time taken to converge. For our case the quality of the solution is much more important that the converging time. Hence the proposed algorithm is much better that the earlier attempts in terms of quality of solution.

Figure 3 shows the improvement of classification accuracy of the Modular neural network with EA over the Modular neural network with BP, and the general model.

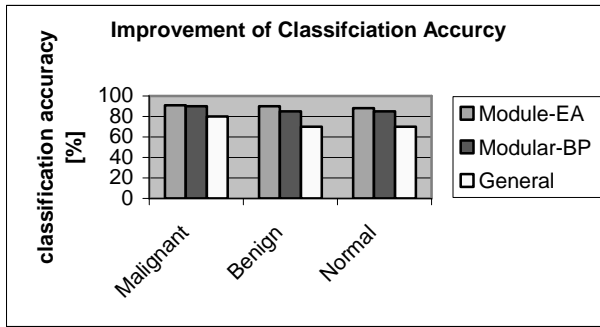
Figure3: Improvement of Classification accuracy

Figure 4 shows the comparison of Type I and Type II error in both the cases.

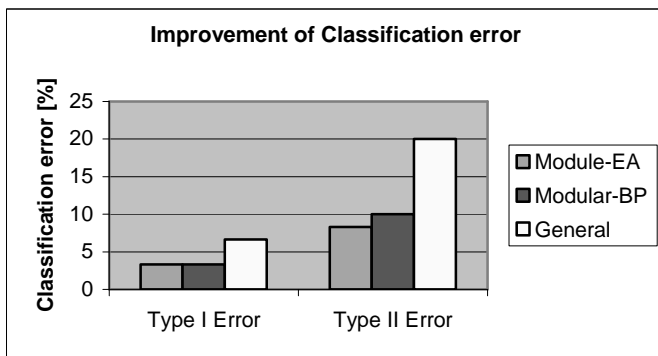
Figure4: Improvement of Classification error

Figure 5 shows the comparison of time complexity in the Modular neural network selection over the general selection model for 10-point cross validation.

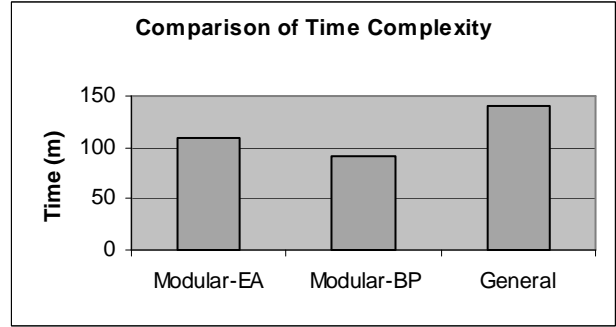
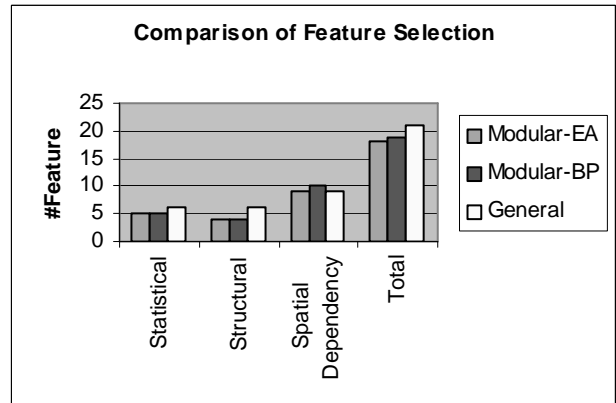
Figure5: Comparison of Time Complexity

Figure 6 shows the comparison of the number of feature selected in each category by the two models. The total number of features selected by the Modular selection model is less than the total number of features selected by the general selection model.

Figure6: Comparison of Feature Selection

IV. CONCLUSION

In this paper a novel modular framework was proposed that is suitable for multi-category feature selection. The selection module uses a combination of GA and neural network classifier. The classifier is a Evolutionary Strategy. The developed structure is compared with earlier work. We have tested the approach with the Digital Mammogram dataset. We have used three categories of statistical, structural and dependency features. A classification accuracy of 88% was achieved.

REFERENCES

1. I. Christoyianni, E. Dermatas, G. Kokkinakis, Neural Classification of Abnormal Tissue in Digital Mammography using Statistical Features of the Texture, *IEEE, ICECS'99*, Pafos, Cyprus, vol.1, pp. 117-120, 1999.
2. J. Martin, M. Moskowitz, and J. Milbrath, Breast cancer missed by mammography. *AJR*, vol. 132, p. 737, 1979.
3. L. Kalisher, Factors influencing false negative rates in xeromammography, *Radiology*, vol. 133, p. 297, 1979.
4. Z. Huo, M. Giger, C. Vyborny, D. Wolverton, R. Schmidt, K. Doi, Computer-Aided Diagnosis: Automated Classification of Mammographic Mass Lesions,

¹ The time is the total time for training the 10-point cross validation training set.

- Proceedings of the 3rd Int. Workshop on Digital Mammograph*, Chicago, IL, pp. 207-211, 1996.
5. H. Chan, D. Wei, M. Helvie, B. Sahiner, D. Adler, M. Goodsitt., and N. Petrick, Computer-Aided Classification of Mammographic Masses and Normal Tissue: Linear Discriminant analysis in Texture Feature Space, *Phys. Med. Biol.*, vol.40, pp. 857-876, 1995.
 6. D. Meersma, P. Scheunders , and D. Van Dyck, Detection of Microcalcifications using Neural Networks, *Proceedings of the 3rd Int. Workshop on Digital Mammograph*, Chicago, IL ,pp. 97-103, 1996.
 7. A. P. Dhawan, Y. Chite, C. Bonasso, and K. Wheeler, Radial-Basis-Function Based Classification of Mammographic Microcalcifications Using Texture Features, *IEEE Engineering in Medicine and Biology & CMBEC*, pp. 535-536, 1995.
 8. A. K. Jain, and D. Zongker, Feature selection: evaluation, application and small sample performance, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 19, no. 2, pp. 153 – 158, 1997.
 9. R. Kohanvi, and G. H. John, Wrappers for feature subset selection, *Artificial Intelligence*, vol. 97, pp. 273 – 324, 1997.
 10. R. Ghosh, M. Ghosh, J. Yearwood, A Modular Framework for Multi category feature selection in Breast Cancer recognition, *Accepted in 12th European Symposium on Artificial Neural Networks*, 2004.