Prediction of Breast Cancer Biopsy Outcomes Using a Distributed Genetic Programming Approach

Simone A. Ludwig
Department of Computer Science
University of Saskatchewan
Canada
ludwig@cs.usask.ca

ABSTRACT

Worldwide, breast cancer is the second most common type of cancer after lung cancer and the fifth most common cause of cancer death. In 2004, breast cancer caused 519,000 deaths worldwide. In order to reduce the cancer deaths and thereby to increase the survival rates an automatic approach is necessary to aid physicians in the prognosis of breast cancer. The most effective method for breast cancer screening today is mammography. However, the predictions of the breast biopsy resulting from mammogram interpretation leads to approximately 70 % biopsies with benign outcomes, which could have been prevented. Therefore, an automatic method is necessary to aid physicians in the prognosis of mammography interpretations. The data set used is based on BI-RADS findings. Previous work has achieved good results using a decision tree, an artificial neural networks and a case-based reasoning approach to develop predictive classifiers. This paper uses a distributed genetic programming approach to predict the outcomes of the mammography achieving even better prediction results.

1. INTRODUCTION

Worldwide, breast cancer is the second most common type of cancer after lung cancer (10.4 % of all cancer incidence) and the fifth most common cause of cancer death. In 2004, breast cancer caused 519,000 deaths worldwide (7 % of cancer deaths; almost 1 % of all deaths) [1]. Breast cancer is the most common malignancy in women, except for non-melanoma skin cancers. It continues to be a major health care problem worldwide. Cancer occurs when cells in a part of the body begin to grow out of control. Normal cells divide and grow in an orderly fashion, but cancer cells do not. They continue to grow and crowd out normal cells. Although, there are many kinds of cancer, they all have in common this out-of-control growth of cells [2].

Different kinds of cancer can behave very differently. For example, lung cancer and breast cancer are very different diseases. They grow at different rates and respond to different treatments. That is why people with cancer need treatment that is aimed at their kind of cancer. Therefore, it is important to identify the type of cancer accurately, so that the correct treatment can be started.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
Copyright 20XX ACM X-XXXXX-XX-X/XX/XX ...$10.00.

Sometimes cancer cells break away from a tumor and spread to other parts of the body through the blood or lymph system. They can settle in new places and form new tumors. When this happens, it is called metastasis. Cancer that has spread in this way is called metastatic cancer.

Even when cancer has spread to a new place in the body, it is still named after the part of the body where it started. For example, if prostate cancer spreads to the bones, it is still called prostate cancer. If breast cancer spreads to the lungs, it is still breast cancer. When cancer comes back in a person who appeared to be free of the disease after treatment, it is called a recurrence.

Breast cancer is a cancer that starts in the tissues of the breast. There are two main types of breast cancer [3]:

- Ductal carcinoma starts in the tubes (ducts) that move milk from the breast to the nipple. Most breast cancers are of this type.
- Lobular carcinoma starts in parts of the breast, called lobules that produce milk.

The good news is that early detection and new treatments have improved survival rates of breast cancer. The 5-year survival rate for women diagnosed with cancer is 80 %. About 88 % of women diagnosed with breast cancer will survive at least 10 years. Unfortunately, women in lower social and economic groups still have significantly lower survival rates than women in higher groups. The good news is that women are living longer with breast cancer. Survivors must live with the uncertainties of possible recurrent cancer and some risk for complications from the treatment itself [4].

Randomized trials and population-based evaluations of screening mammography have shown that early detection of breast cancer via mammography greatly improves the chances of survival [23]-[26]. It is possible with mammography to identify cancer several years before physical symptoms are produced and therefore is recognized as the most effective breast cancer screening method which is available today. However, approximately 5-10 % of the mammography results are classified as abnormal or inconclusive until further examinations such as ultrasound imaging or breast biopsy lead to a final interpretation of normal or benign breast tissue. Only 10-30 % of all breast biopsies are reported to actually show a malignant pathology [27]. The unnecessary breast biopsies causes major mental and physical discomfort for the patients as well as unnecessary expenses spent for examinations.

In the past several years computer aided diagnosis (CAD) systems have been proposed that use lesion descriptions based on the BI-RADS [5] standard lexicon as input attributes to support the physician's decision to perform a breast biopsy or a short follow-up diagnosis on a suspicious region seen in a mammogram. In general, BI-RADS attributes are collected by different physicians trained at

different radiology centres providing values for a given BI-RADS attribute such as the mass shape, to a suspicious region seen in a mammogram. An artificial neural networks approach was proposed to deduce the diagnosis from BI-RADS descriptions [6, 7]. Alternative approaches based on case-based reasoning and Bayesian networks were later proposed [8]-[12]. The advantage of the case-based reasoning method over the earlier proposed approaches is the intelligible reasoning process that leads to the systems diagnosis outcome. A case-based reasoning CAD system reasons based on stored knowledge, i.e. given cases with associated ground truth) and its final diagnosis outcome is based on the ground truth of the stored cases that are most similar to the query case. Hence its reasoning process is much easier to comprehend for the physician than the artificial neural network approach whose model is internal and the reasoning can not be captured.

In this paper another approach is investigated for this problem based on genetic programming. Genetic programming is an evolutionary computation approach which is based on natural evolution. This approach will be explained in more detail and it will be compared to a decision-tree approach, a neural network approach and a case-based reasoning approach, which were proposed earlier.

The paper is structured as follows. First, some research regarding related work is outlined in Section 2. In Section 3, the proposed approach using genetic programming is introduced outlining the different parameters involved. The experiments and results are given in Section 4. Section 5 concludes this paper with a summary and analysis of the results obtained.

2. RELATED WORK

Related work on classifiers for the prediction of breast cancer biopsy outcomes using the Breast Imaging Recording and Data System (BI-RADS) of the American College of Radiology is vast. Different types of classifiers have been developed.

Artificial neural network was used to categorize benign and malignant breast lesions. The eighteen inputs to the network included 10 BI-RADS lesion descriptors and eight input values from the patient's medical history. The network was trained and tested on 206 cases (133 benign, 73 malignant cases). The positive predictive value of the biopsy from 35 % to 61 % with a relative sensitivity of 95 %, the specificity of the artificial neural network approach (62 %) was significantly greater than the specificity of radiologists (30 %) [6].

Another approach investigated is a case-based reasoning system which was designed to support the decision to perform biopsy in those patients who have suspicious findings on benign lesions. The system is designed to help decrease the number of benign biopsies without missing malignancies. Clinicians interpret the mammograms using a standard reporting lexicon. The case-based reasoning system compares these findings with a database of cases with known outcomes (from biopsy) and returns the fraction of similar cases that were malignant. This malignancy fraction is an intuitive response that the clinician can then consider when making the decision regarding biopsy. The system was evaluated using a round-robin sampling scheme and performed with an area under the receiver operating characteristic curve of 0.83, comparable with the performance of a neural network model [8].

Another case-based reasoning classifier was developed to predict biopsy results from BI-RADS findings. Case-based reasoning similarity was defined using either the Hamming or Euclidean distance measure over case features. Ten features represented each case: calcification distribution, calcification morphology, calcification number, mass margin, mass shape, mass density, mass size, associated findings, special cases, and age. The performance was evaluated using Round Robin sampling, Receiver Operating Characteristic (ROC) analysis, and bootstrap. Feature selection was performed over all possible feature combinations (1022) and similarity thresholds. Different features were identified for the different distance measures used, however, the ROC value was 0.90 in both cases [10].

A Bayesian network structure learning and probability estimation method was investigated to classify breast lesions. The method was compared to a naive Bayes classifier which was previously developed [28]. The proposed approach reflects the difference in the classification of biopsy outcomes and the invasiveness of malignant lesions for breast masses and microcalcifications. It was found that the difference between masses and microcalcifications should be taken into consideration when interpreting systems for automatic pathological classification of breast lesions [12].

A classifier based on the likelihood ration was developed. The data set used for this classifier contained 670 cases (245 malignant) with 16 features from BI-RADS findings and patient history findings. A separate database of 151 (43 malignant) validation cases were collected that were previously unseen by the classifier. Performance evaluation methods included ROC, round-robin, and leave-one-out bootstrap sampling [11].

Another paper investigated using different similarity measures for a case-based reasoning classifier to predict breast cancer. It also used the BI-RADS description of a lesion to predict the outcomes of breast cancer. The similarity measures used were Euclidean distance and Hamming distance. The result was that Euclidean distance measure produced a greater ROC area than the Hamming distance, with significant results. A ROC area of 0.82 ± 0.01 was achieved with the dataset collected by the Duke University Medical Center. [9]

3. PROPOSED APPROACH

The origins of evolutionary computation reach back to the 50's of the last century. Genetic programming, in itself, was not considered until the middle of the 80's. The term first appeared in [18], the main development took place in the early and middle 90's, particularly through work by Koza [19].

Genetic programming uses the concepts of genetics and Darwinian natural selection to generate and evolve entire computer programs. Genetic programming largely resembles genetic algorithms in terms of its basic algorithm. The notions of mutation, reproduction (crossover) and fitness are essentially the same, however, genetic programming requires special attention when using those operations. While genetic algorithms are concerned with modifying fixed-length strings, usually associated with parameters to a function, genetic programming is concerned with actually creating and manipulating the (non-fixed length) structure of the program (or function).

Therefore, genetic programming is more complex than genetic algorithms [20] and works as follows. In genetic programming one wants to find solutions to some problem in the form of a computer program. It is a stochastic search strategy that is particularly powerful in the circumstances where one cannot make any assumptions about the characteristics of the solution. The solution is developed by first creating a number of initial programs, which are then recombined and changed in each evolution step. The set of programs is referred to as the population; any single program is an individual. Run of evolution is the term used to describe the whole process of finding a solution.

Before starting an evolution, one has to define (at least) the following:

- Fitness measure/function: a function that evaluates how close a program is to the optimal solution.
- Population size: the number of programs that are supposed to be used for evolving a solution.
- Function and terminal set: functions, constants, and variables the programs are allowed to use.
- Genetic operators: selection, crossover and mutation operators and the probability for using the later two. There are a variety of different selection, crossover and mutation operators available to choose from.
- Termination criterion: the evolution usually either ends if a sufficiently good solution is found, or if the maximum number of iterations is reached.

After setting these parameters, the initial population can be created. Unless one already has some idea about how the solution might look like, the programs are built randomly. Each evolution step works as follows: Until a certain percentage of the population size (crossover rate) is reached, new programs are constructed as follows: two programs are selected according to the chosen selection method. The programs are 'crossed over', that means certain parts of them are swapped.

In tree-based genetic programming, a subtree is selected in each program and the two subtrees are swapped. The remaining part of the new population consists of copied programs from the old population (reproduction is the term used for copying old programs) or newly created programs. With a certain probability, the mutation rate, an individual is changed. Mutation can have various forms, most commonly it only changes one function/terminal in a program to a different one. This process is repeated until the termination criterion is reached. The result of the run is usually the program with the best fitness value found during the whole evolution [21].

In summary, genetic programming performs the following steps:

Step 1: Assign the maximum number of generations to be run and probabilities for cloning, crossover and mutation.

Step 2: Generate an initial population of computer programs of size N by combining randomly selected functions and terminals.

Step 3: Execute each computer program in the population and calculate its fitness with an appropriate fitness function. Designate the best-so-far individual as the result of the run.

Step 4: With the assigned probabilities, select a genetic operator to perform cloning, crossover or mutation.

Step 5: If cloning operator is chosen, select one computer program from the current population of programs and copy it into a new population. If crossover operator is chosen, select a pair of computer programs from the current population, create a pair of offspring programs and place them into the new population. If mutation operator is chosen, select one computer program from the current population, perform mutation and place the mutant into the new population.

Step 6: Repeat Step 4 until the size of the new population of computer programs becomes equal to the size of the initial population N.

Step 7: Replace the current (parent) population with the new (offspring) population.

Step 8: Go to Step 3 and repeat the process until the termination criterion is satisfied.

Distributed GP evolves not only one but several populations in parallel, after a fixed number of generations a predefined number of individuals is exchanged. This guarantees the diversity in each population that tends to lead to better results as after some iterations many programs are just variations of the best solution, because this one is preferably selected for crossover and reproduction, and therefore, the evolution does not result in new solutions per say. Injecting newly created programs could be used, but these have a lower fitness and are not often considered for crossover. However, adding other programs of similar fitness, but possibly having a very different structure, is more likely to create good new solutions. Therefore, a distributed approach with 10 populations is chosen, each of size 100. After each 100 generations 10 programs are moved to a different population, which is repeated 10 times.

For this research investigation, both, the normal GP and also the distributed GP approach are used. In order to achieve good prediction accuracy, the following genetic programming parameters were chosen: tournament selection of size 4, population size 1,000, minimal initial depth 5, maximal crossover depth 12, crossover rate 0.7, mutation rate 0.3. Basic mathematical operations (addition, subtraction, multiplication, division), comparison operators (<,>), if-statement, logarithm and exponential made up the function set. For those parameters that were analysed, the best performing value was used for all following tests. 30 test runs were performed for each parameter value in order to guarantee statistically distributed values.

The Java Genetic Algorithms Package (JGAP) [22] was chosen as the programming platform. JGAP is a Genetic Algorithms and Genetic Programming package written in Java. It is designed to require minimum effort to use, but is also designed to be highly modular. It provides basic genetic mechanisms that can be used to apply evolutionary principles to solve problems and was used and expanded for this research investigation.

4. EXPERIMENTS AND RESULTS

4.1 Mammography Data Set

The mammography data set from the UCI machine learning repository [13] was taken for this investigation. The database consists of data used from modern full-field digital mammograms and contains cases which have been acquired during the period of 2003 to 2006. 515 (53.6 %) of these mass regions are benign and 446 (46.4 %) are malignant. The features recorded are patient's age and BI-RADS descriptions, as well as the mass density. In particular, the following feature including the predictive outcome is given as follows [25]:

1. BI-RADS assessment: 1 to 5 (ordinal)
2. Age: patient's age in years (integer)
3. Shape: mass shape: round=1, oval=2, lobular=3, irregular=4 (nominal)
4. Margin: mass margin: circumscribed=1, microlobulated=2, obscured=3, ill-defined=4, spiculated=5 (nominal)
5. Density: mass density: high=1, iso=2, low=3, fat-containing=4 (ordinal)
6. Severity: benign=0 or malignant=1 (binominal, outcome)

Figure 1 shows the histogram of the age attribute showing the benign and malignant regions. It shows that malignant regions are observed in persons of age 40 to 90 years with most cases in the 70 year group.

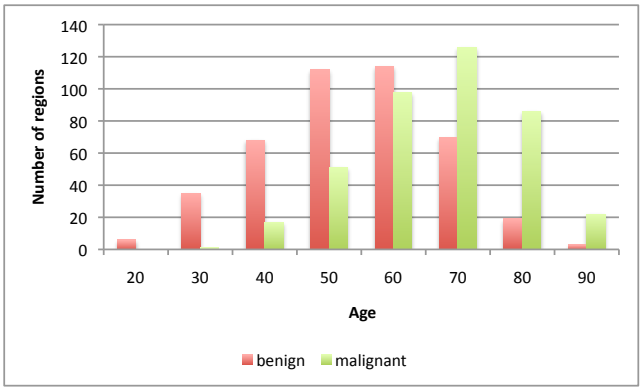

Figure 1: Histogram of age attribute showing the benign and malignant regions

	DT	ANN	CBR
$A(z)$	0.838±0.017	0.847±0.017	0.857±0.016
$A(z)_{0.9}$	0.477±0.060	0.521±0.055	0.505±0.063
$Spec_{0.95}$	0.298±0.076	0.338±0.070	0.313±0.054

Table 1: ROC Performance of three approaches [25, 26]

	GP	DGP
$A(z)$	0.859±0.032	0.860±0.032
$A(z)_{0.9}$	0.503±0.028	0.514±0.028
$Spec_{0.95}$	0.290±0.015	0.271±0.015

Table 2: ROC Performance of proposed approach (normal (GP) and distributed version (DGP))

4.2 Comparison Approaches

The proposed genetic programming approach is compared with the approaches of the researchers who provided the data set [25, 26]. Therefore, a direct comparison is possible. The genetic programming approach is measured in two ways; first, the normal version is run and then the distributed version is run.

4.2.1 Decision Tree Approach (DT)

Decision tree classification is a very common and heavily used classification technique. A decision tree is built where nodes represent attributes and leafs represent the values of the attributes. Terminal nodes are called leaf nodes and represent the classification outcome, e.g. "benign" or "malignancy". There are many different decision learning algorithms available, but the most used one is the ID3 algorithm proposed by Quinlan [14]. The ID3 algorithm uses the technique of information gain as a quantitative measure to construct the tree from a set of examples. The approach which is compared in this research is an extension of the ID3 algorithm called C4.5 [15]. It supports numeric attributes in addition to nominal ones and employs a pruning method to make the decision tree more compact, and it deals with missing data.

4.2.2 Artificial Neural Network Approach (ANN)

An Artificial Neural Network is a mathematical model that tries to simulate the structure and functional aspects of biological neural networks. It consists of an interconnected group of artificial neurons and processes information. ANN can be used to model complex relationships between inputs and outputs or to find patterns in data. As mentioned previously, ANN is the state-of-the-art approach in medical diagnosis, especially in the diagnosis of breast cancer. The ANN used to compare the genetic algorithm approach consists of a three-layer, feed-forward network and it is trained using backpropagation. The layers consist of an input layer with one input node per attribute, a hidden layer with four nodes, and an output layer with a single output node.

4.2.3 Case-based Reasoning Approach (CBR)

A case-based reasoning system stores expertise as a library of cases with known outcomes. Each case is a region of interest containing a mammographic abnormality together with a set of input attributes as described in the following. The known outcome is the biopsy result (benign and malignant) for the abnormality. To generate a diagnosis proposal for a new region, the BI-RADS attributes of the query region are matched against the BI-RADS attributes of all the regions in the case library using a similarity metric. The most similar regions are retrieved for the database. The known classifications of the retrieve regions are then used to suggest a solution for the query region (benign or malignant) based on a decision rule.

4.3 Performance Evaluation

The performance of the approaches is best described in terms of their sensitivity and specificity quantifying their performance related to false positive and false negative instances. In particular, sensitivity gives the percentage of the true positive fraction, and specificity is calculated by 1 - false positive fraction, which allows one to draw a ROC curve [16, 17]. The ROC curve is generated by varying a confidence threshold for cases being classified as malignant over the range from zero to one. The area under the ROC curve A(z) is used as a metric for the performance of the system by using numeric integration. For breast cancer prediction high sensitivity is usually considered more important than high specificity, i.e. it is better to falsely classify a benign region as malignant rather than to miss a breast cancer by classifying a malignant region as benign. Hence, we also use the area under the partial ROC curve for a sensitivity of 0.9 or greater, denoted as $A(z)_{0.9}$, and the specificity $Spec_{0.95}$ at the given sensitivity 0.95 as performance measures.

4.4 Results

The results from the comparison of the proposed approach, both the normal and the distributed version, with the three given approaches are shown in Table 1 and 2, as well as shown graphically in Figure 2. It can be seen that both the GP and DGP approach achieve higher ROC values. The best ROC value achieved by the given approaches is the one for CBR which achieved a value of 0.857±0.016. The GP approach achieves a ROC value of 0.859, whereas the DGP approach achieves a slightly higher ROC value of 0.860.

The same trend can be seen for the ROC area for sensitivity equals or greater than 90 % ($A(z)_{0.9}$). The specificity values, on the other hand, achieved by the GP and DGP approach are slightly smaller, compared to the given three approaches. However, for the prediction of breast cancer high sensitivity is usually considered as more important than high specificity.

Each case in the data set represents a BI-RADS number ranging from 2 (definitely benign) to 5 (highly suggestive of malignancy) that are assigned by the physicians. Assuming that all cases with BI-RADS numbers greater than or equal to a given value (varying from 2 to 5) are malignant and the other cases benign, sensitivities and associated specificities can be calculated. The following sensitivity values were obtained as shown in Table 3 (as well as in

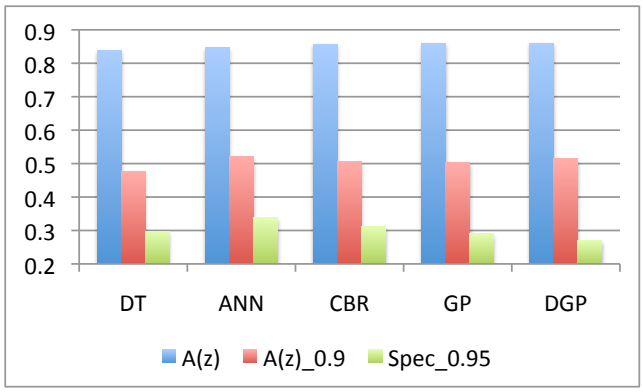

Figure 2: ROC performance and specificity of all approaches

BI-RADS	Sens.	lower	upper
≥ 2	0.839713	0.800184	0.872846
≥ 3	0.839713	0.800184	0.872846
≥ 4	0.841346	0.801839	0.874394
≥ 5	0.898773	0.859519	0.928310

Table 3: Sensitivity values including 95 % confidence interval for different BI-RADS categories

Figure 3). The sensitivity values together with their 95 % confidence interval values are shown for the different BI-RADS levels. The sensitivity values are higher for greater bands of BI-RADS categories, which suggests that the prediction of benign and malignant cases increases.

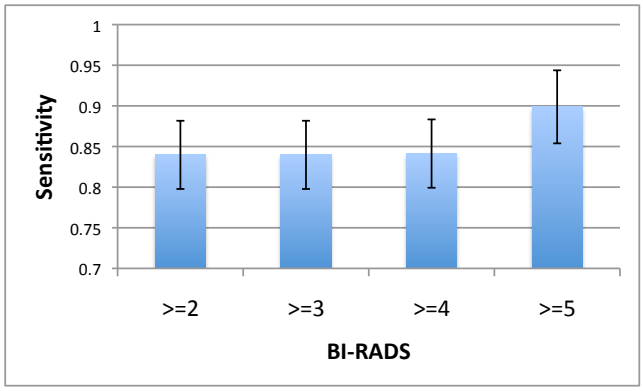

Figure 3: Sensitivity values for different BI-RADS categories

A similar trend for the specificity values can be observed in Table 4 (as well as in Figure 4). The 95 % confidence values are also given. Please note, that for the BI-RADS category 5 or higher, the false positives and the true negatives were zero and therefore, no specificity values could be calculated.

5. CONCLUSION

Given that mammogram interpretation leads to 70 % of biopsies with benign outcomes which could have been prevented, researchers have come up with an automatic technique called classifiers to predict breast cancer outcomes from mammography. Classifiers based on artificial neural networks, decision tree classifiers,

BI-RADS	Spec.	lower	upper
≥ 2	0.120393	0.091196	0.156962
≥ 3	0.122500	0.092814	0.159644
≥ 4	0.121693	0.091329	0.159970
≥ 5	NaN	NaN	NaN

Table 4: Specificity values including 95 % confidence interval for different BI-RADS categories

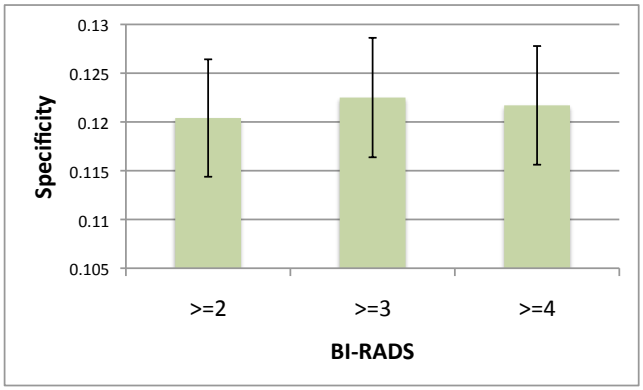

Figure 4: Specificity values for different BI-RADS categories

and case-based reasoning approaches have been proposed in the past.

This paper introduced a new method based on genetic programming for building a classifier. The genetic programming technique was investigated in two ways. First, normal genetic programming was used, and then a distributed version was developed and examined. One data set was investigated, which has been used in another research investigation and therefore, a fair comparison could be conducted.

Both genetic programming approaches outperform the previous techniques used. A ROC value of 0.859 and 0.860 was achieved by the normal and the distributed approach respectively, whereas the former best technique scored 0.857. One drawback which the genetic programming method has however, is that the training procedure for training the classifier takes a very long time. Depending on the settings of the parameters one training run can take approximately 40 minutes. Therefore, if the classifier needs to be trained "on the fly", then the genetic programming method will most likely not be the first choice, even though it has shown to achieve higher accuracy for the prediction of breast cancer mammograms.

Comparing the transparency of the models of the classifiers, the decision tree generates a decision tree which is quiet natural and easily understood. The neural network approach does not reveal the model, as it is seen as a black box. The case-based reasoning process is also not intuitive, however, given that the distance between exemplars are used, it is at least possible to observe. The genetic programming approach builds a computer program in the form of a mathematical equation for the classification of benign and malignant outcomes. This might not be as intuitive as the decision tree approach, however, the formal representation can easily be used in software systems.

6. ACKNOWLEDGMENTS

The author would like to thank the Natural Sciences and Engineering Research Council Canada (NSERC) for the partial funding

of this research.

7. REFERENCES

[1] World Health Organization, "Fact sheet No. 297: Cancer", http://www.who.int/mediacentre/factsheets/fs297/en/index.html, last retrieved June 2010.

[2] K. Nouri "Skin Cancer" McGraw-Hill, 2007.

[3] Calgary Breast Health Program http://www.calgaryhealthregion.ca/ breasthealth/breastcancer/typesofbreastcancer.htm, last retrieved June 2010.

[4] Breast Cancer Prognosis http://www.healthcentral.com/breast-cancer/, last retrieved June 2010.

[5] American College of Radiology Breast Imaging Reporting and Data System Bi-RADS, Atlas 2006.

[6] J. A. Baker, P. J. Kornguth, J. Y. Lo, M. E. Williford, and C. E. Floyd, "Breast cancer: Prediction with artificial neural network based on BI-RADS standardized lexicon", Radiology 196, 817-822, 1995.

[7] M. K. Markey, J. Y. Lo, R. Vargas-Voracek, G. D. Tourassi, and C. E. Floyd "Perception error surface analysis: A case study in breast cancer diagnosis", Comput. Biol. Med. 32, 99-109, 2002.

[8] C. E. Floyd, J. Y. Lo, and G. D. Tourassi, "Case-based reasoning computer algorithm that uses mammographic findings for breast biopsy decisions", AJR, Am. J. Roentgenol. 175, 1347-1352, 2000.

[9] A. O. Bilska-Wolak and C. E. Floyd, "Investigating different similarity measures for a case-based reasoning classifier to predict breast cancer", Proc. SPIE 4322, 1862-1866, 2001.

[10] A. O. Bilska-Wolak and C. E. Floyd, "Development and evaluation of a case-based reasoning classifier for prediction of breast biopsy outcome with BI-RADS lexicon", Med. Phys. 29, 2090-2100, 2002.

[11] A. O. Bilska-Wolak, C. E. Floyd, J. Y. Lo, and J. A. Baker, "Computer aid for decision to biopsy breast masses on mammography: Validation on New Cases", Acad. Radiol. 12, 671-680, 2005.

[12] M. K. Markey, E. A. Fischer, and J. Y. Lo, "Bayesian networks of BIRADS descriptors for breast lesion classifications", in International Conference of the IEEE Engineering in Medicine and Biology Society San Francisco, California, 2004, pp. 3031-3034.

[13] D. J. Newmann, S. Hettich, C. L. Blake, and C. J. Merz, "UCI repository of machine learning databases", Univeristy of California, Irvine, Dept. of Information and Computer Sciences, http://www.ics.edu/simmlearn/MLRepository.html, 1998.

[14] J. R. Quinlan, "Induction of decision trees", Mach. Learn. 1, 81-106, 1986.

[15] J. R. Quinlan, C4.5: Programs for Machine Learning Kaufmann, San Francisco, 1993.

[16] L. B. Lusted, "Signal detectability and medical decision-making", Science, 171, 1217-1219, 1971.

[17] J. P. Egan, Signal Detection Theory and ROC Analysis Academic, New York, 1975.

[18] N. L. Cramer, "A representation for the Adaptive Generation of Simple Sequential Programs", Proceedings of an International Conference on Genetic Algorithms and the Applications, Grefenstette, 1985.

[19] J. R. Koza, "Genetic Programming: On the Programming of Computers by Means of Natural Selection", MIT Press, Cambridge, 1992.

[20] W. Banzhaf, P. Nordin, R. E. Keller, F. D. Francone, "Genetic programming: an introduction on the automatic evolution of computer programs and its applications", Morgan Kaufmann Publishers, 1998.

[21] R. Poli, W. B. Langdon, N. F. McPhee. "A Field Guide to Genetic Programming", http://www.gp-eld-guide.org.uk, last retrieved June 2010.

[22] Java Genetic Algorithms Package (JGAP), source code, http://jgap.sourceforge.net, last retrieved June 2010.

[23] L. L. Humphrey, M. Helfand, B. K. Chan, and S. H. Woolf, "Breast cancer screening: A summary of the evidence for the U.S. Preventive Services Task Force", Ann. Intern Med. 137, 347-360, 2002.

[24] L. Tabar, M. F. Yen, B. Vitak, H. H. Tony Chan, R. A. Smith, and S. W. Duffy, "Mammography service screening and mortality in breast cancer patients: 20-years follow-up before and after introduction of screening", Lancet 361, 1405-1410, 2003.

[25] S. W. Duffy, L. Tabar, and H. H. Chen, "The impact of organized mammography service screening on breast carcinoma mortality in seven Swedish counties", Cancer 95, 458-469, 2002.

[26] L. Tabar, B. Vitak, H. H. Chen, M. F. Yen, S. W. Duffy, and R. A. Smith, "Beyond randomized controlled trials: Oraganized Mammographic screening substantially reduces breast carcinoma mortality", Cancer 91, 1724-1731, 2001.

[27] M. Elter, R. Schulz-Wendtland, T. Wittenberg "The prediction of breast cancer biopsy outcomes using two CAD approaches that both emphasize an intelligible decision process", Med Phys. 2007 Nov. 34(11):4164-72.

[28] M. Elter, A. Horsch "CADx of mammographic masses and clustered microcalcifications: a review", Med Phys. 2009 Jun. 36(6):2052-68.

[29] D. B. Kopans, "The positive predictive value of mammography", AJR, Am. J. Roentgenol. 158:521-526, 1992.

[30] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian Network Classifiers", Machine Learning, 1997. 29:131-163.

APPENDIX

Preferred allocation of reviewing expertise:
Computing, Medicine

Topics covered in paper:

- Systems for cognitive and decision support
- Public health informatics
- Engineering medical data