Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2012) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

Complex Adaptive Systems, Publication 2
Cihan H. Dagli, Editor in Chief
Conference Organized by Missouri University of Science and Technology
2012- Washington D.C.

PNN/GRNN Ensemble Processor Design for Early Screening of Breast Cancer

Walker H. Land, Jr.*[a], Xinpei Ma[a], Erin Barnes[a], Xingye Qiao[b], John Heine[c], Timothy Masters[d], Jin Woo Park[a]

[a] *Department of Bioengineering, Binghamton University, Binghamton, NY 13902-6000, USA.*
[b] *Department of Mathematical Sciences, Binghamton University, Binghamton, NY 13902-6000, USA.*
[c] *Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center & Research Institute, 12902 Magnolia Drive, MRC-CANCONT, Tampa, FL 33612, USA.*
[d] *TMAIC, Brackney, PA 18812, USA.*

Abstract

Breast cancer screening has reference to screening of asymptomatic, generally healthy women for breast cancer, to identify those who should receive a follow up check. Early screening can detect non-invasive ductal carcinoma in situ (called "pre breast cancer"), which almost never forms a lump and is generally non-detectible, except by mammography. This paper will describe the design and preliminary evaluation of this PNN/GRNN ensemble pre-screener, in the context of a possible pre-screening protocol, which may, if required, include other data. The results show that using the ensemble technique provides almost a 20% AUC increase over the average standalone PNN and almost 10% over the best performing PNN.

Keywords:Bioinformatics;Biomedical;Statistical neural networks;Breast cancer screening

1. Background

Carcinoma of the breast is second only to lung cancer as a tumor-related cause of death in women. It has been reported by the National Cancer Institute (NCI) that 211,300 new cases and 39,800 deaths will occur just in the US [1]. It also has been proposed that mortality from breast cancer could be decreased by up to 25% provided that all women in appropriate age groups are regularly screened [2]. Currently, the method of choice for early screening of breast cancer is conventional x-ray mammography, due to its general widespread availability, low cost, speed, and noninvasiveness. At the same time, while mammography is sensitive to the detection of breast cancer, it has a low positive predictive value (PPV), resulting in costly and invasive biopsies that are only 15%-34% likely to reveal malignancy at histological examination [3]. Computational intelligence has been applied to the problem of breast cancer **diagnosis** by several researchers. Earlier work by Floyd and Lo *et al*. [4] demonstrated the applicability of back propagation-trained Multiple Layer Feed Forward Neural Networks (MLFNs) to this task. Evolutionary-Programming (EP)-based feed forward networks were applied by Fogel *et al*. [5-8] and Land *et al*. [9]. These applications were able to achieve similar or better results than those trained by back propagation, but with much simpler architectures (*i.e.*, fewer nodes in the hidden layer). EP-based linear discriminate analysis [10], and an adaptive boosting/EP hybrid [11] also were investigated as potential screening / classification mechanisms. Finally, the Riverside Breast Cancer Committee encourages [12] yearly screening mammograms for women beginning at age 40. They encourage all women to talk with their primary care physician or gynecologist about whether high-

risk factors warrant earlier and more frequent screening. More detail and recent research summary on breast cancer may be found in [14].

2. GRNN Ensemble Formulation Summary

The objective is to design an ensemble processor that uses the gate variables to intelligently combine the outputs of competing models [15]. In the application section of this paper, we use Probabilistic Neural Networks (PNN) as the competing models. A background and history of ensemble processing may be found in [14]. Once the expected error of each competing prediction model is estimated, these expected errors are used to compute the weights for each model. When an unknown case is processed, the gate variables are used by the General Regression Neural Network (GRNN) to decide which competing models are likely to be the best for this particular case. These models are weighted more heavily than the likely inferior models. In particular, one has a training set composed of n cases. Each case $\mathbf{x}_i$ (i =1, …, n) consists of p gate variables: $x_{i,j}$ where j =1, …, p. These *gate variables* determine in some way the relative efficacy of the prediction models. The m competing prediction models provide outputs $q_{i,k}$ for the case $\mathbf{x}_i$ where k =1, …, m. The desired output (the target value) is y_i.

For the gate variables and model outputs of a trial case $\mathbf{x}$ (which is a future case other than the training set), just one subscript is used: x_j where j =1, …, p, are the values of the observed *gate variables*, and q_k where k=1, …, m, are the computed outputs of the m competing prediction models for this trial case.

Define $D(\mathbf{x},\mathbf{x}_i)$ the weighted Euclidean distance between training case $\mathbf{x}_i$ and the trial case $\mathbf{x}$, using p sigma weights for p gate variables. Then the GRNN ensemble's predicted squared error for model k may be shown to be:

$$\hat{e}_k(\mathbf{x}) = \frac{\sum_{i=1}^{n} (y_i - q_{i,k})^2 \exp(-D(\mathbf{x}, \mathbf{x}_i))}{\sum_{i=1}^{n} \exp(-D(\mathbf{x}, \mathbf{x}_i))} \tag{1}$$

It is desired that the final prediction for case $\mathbf{x}$ is a linear combination of the outputs of the competing models:

$$\hat{y} = \sum_{k=1} w_k q_k \tag{2}$$

Here w_k is the weight for the k-th model. If the models have the (desirable) property that their predictions are unbiased, the following condition must be imposed:

$$\sum_{k=1}^{m} w_k = 1 \tag{3}$$

It can be shown that the linear combination of unbiased estimators having minimum mean-squared error uses weights proportional to the reciprocal of each estimator's variance. If the predicted squared error calculated in (1) is used in place of the variance, the following formula is derived for the weights:

$$w_k = \frac{1/\hat{e}_k}{\sum_{l=1}^{m} 1/\hat{e}_l} \tag{4}$$

The GRNN ensemble processor is trained (*i.e.*, the p sigma weights in the distance metric $D(\mathbf{x},\mathbf{x}_i)$ are optimized) in the usual leave-one-out cross validation manner as follows. To evaluate the quality of a sigma vector, a case is removed from the training set and the formulas just shown are used first to estimate the competing models' errors, then compute the w_k weights, then weight the competing models to get the grand prediction, and finally compute the error of this grand prediction (compared with the true value of the omitted case). This procedure is repeated for each training case. The sigma vector providing minimum root mean square (RMS) error is found.

3. Differential Evolution

Price and Storn [13] reported on a variation of genetic optimization called *differential evolution*. This variation is much more appropriate than traditional genetic methods when optimizing a multivariate function. It is especially

valuable when the scaling in the different dimensions are not commensurate, a situation commonly found in both PNN and GRNN training using poorly prescaled or highly correlated variables. Consequently, differential evolution was used to train the GRNN ensemble processor.

Differential evolution [15] is similar to traditional genetic optimization in that it starts with a collection of parameter sets that will be called the *source population*. The individuals comprising this population are combined with each other via *crossover* and subjected to *mutation* to produce the members of the *destination population*. The members of the destination population, taken as a group, would generally be expected to be superior to the members of the source population. By repeating this process enough times, the best member of the final population should be close to the global optimum.

Several important differences exist between traditional **genetic optimization** and **differential evolution,** where the most important difference is in the nature of the mutation. In traditional genetic optimization, mutation takes the form of a random perturbation of a fixed type, such as flipping bits in a binary representation of a parameter set, or adding random numbers to specific individual parameters. This approach, however, fails to account for the fact that what might be a small perturbation for one parameter might be gigantic for another. Also, random bit flipping can be extremely destructive. On the other hand, differential evolution avoids these problems by using the source population itself to determine the nature and degree of mutation, by randomly selecting a pair of individuals and computing the difference between their parameter vectors. This difference vector is multiplied by a fixed constant (around 0.5 or so) and added to the individual being mutated. When optimization begins, the average difference will be about the same for all variables being optimized. But as generations pass, the difference will tend to adapt to the natural scaling of the problem. Variables having a large natural scale will be distributed over a larger range in the population, so mutations for these variables will also be relatively large. As convergence approaches, those variables having a narrow and well-defined range around the minimum will have small variation among the population members, resulting in their mutations being relatively small. This automatic adaptation significantly improves behavior of the algorithm as convergence nears. (Note: Another way to address this problem is to normalize the feature covariate set to [0,1] or [-1,1], etc.)

Another important difference is that differential evolution does not involve selection of parents based on fitness. Instead, fitness determines which children are kept. In particular, one parent, called the *primary parent* is selected deterministically: each individual in the source population is chosen as a primary parent exactly once.

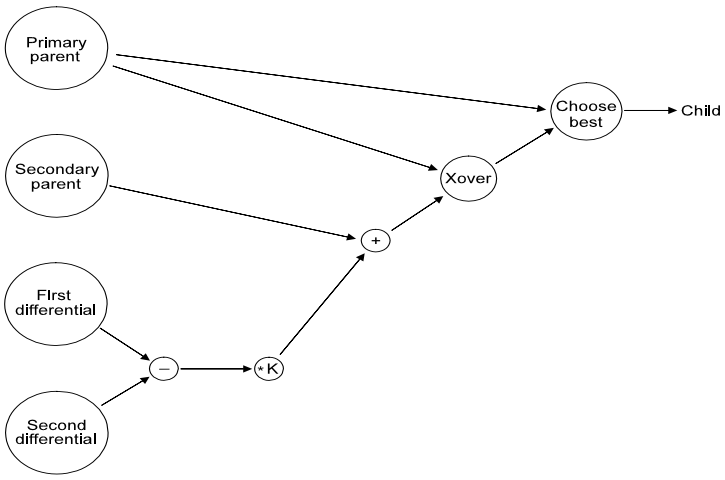

Fig.1 Differential Evolution Overview

The other parent, called the *secondary parent*, is randomly chosen. Two other individuals which make up the *differential pair* are also selected randomly. These two are subtracted and their difference is multiplied by a small fixed constant. This scaled difference vector is added to the secondary parent to induce mutation. Ordinary crossover

is applied to the primary parent and the mutated secondary parent. The resulting child's fitness is compared to that of the primary parent. The winner becomes a new member of the destination population. This entire process is illustrated in Figure 1 above.

3. Data Base

The data set consisted of ~200 samples, approximately equally divided between cases and controls**.** Patient Information Data, which included retrospective non-randomized information from the Patient Breast Questionnaire, Diagnostic Exams, Pathology data, Magnetic Resonance (MR) breast examinations with the de-identified per HIPAA regulations (the Health Insurance Portability and Accountability Act of 1996) MR images and the truth-files generated by the radiologist. Two hundred patient cases were collected retrospectively (from 01/01/2005 to 06/01/06), 100 of which were negative (benign) and 100 were positive (malignant). All cancer-positive cases were pathology verified. The patient population represents a subset of those patients encountered in diagnostic mammography that are undergoing diagnostic Magnetic Resonance Mammography (MRM). The data is retrospective and therefore was acquired with no specific imaging protocol. This collection of MRM subjects are patients who (1) have a known diagnosis of cancer (to determine extent of disease, evaluate for synchronous lesions), (2) have equivocal mammogram (suspicious mammogram finding but not clearly localized) or (3) require follow-up after therapy (for example, dense breasts in an XRT/ lumpectomy patient). The collection also consists of patients that were scanned prior to their surgeries (known cancer positive patients). The cases that have an equivocal mammogram are usually benign and are followed by repeat MRM, ultrasound, or x-ray mammogram. If an equivocal finding looks suspicious it is biopsied. The patient data are non-image features, which are: (1) patient age, (2) race/ethnicity, (3) weight, (4) age of menarche, (5) duration of menstruation, (6) age of menopause (7) number of full term deliveries (8) duration of hormone replacement therapy (9) family history of breast cancer (yes or no), and (10) cigarette smoking. The database was pared-down by discarding subjects for a particular trial that had open features fields for a given model arrangement (for either feature non-applicable reasons or when the feature was not-available).

4. Results

PNNs (Probabilistic Neural Networks) were classifiers that were used as inputs to the GRNN ensemble. The PNN/GRNN ensemble refers to the GRNN ensemble with the PNNs as inputs. The PNN/GRNN had an area under the receiver operating characteristic (ROC) curve (AUC) value of 83.08%. This outperforms the standalone PNNs by an average of 19.56% AUC and outperforms the best standalone PNN by 9.06% AUC. This result displays that the GRNN ensemble has the potential to improve the performance of the PNN to a level that can be used in clinical situations.

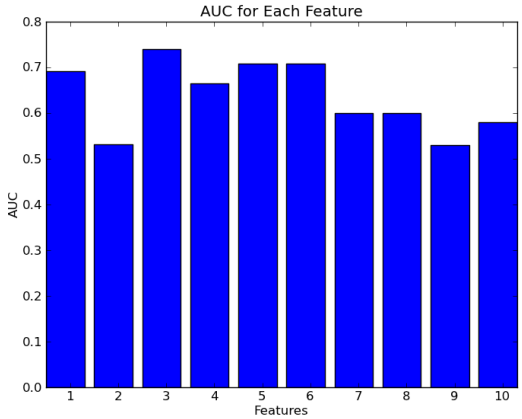

Fig. 4. AUC for each feature

The figure above illustrates each feature's AUC value. Each feature corresponds to (1) patient age, (2) race/ethnicity, (3) weight, (4) age of menarche, (5) duration of menstruation, (6) age of menopause (7) number of full term deliveries (8) duration of hormone replacement therapy (9) family history of breast cancer, and (10) cigarette smoking.

Features	1	2	3	4	5	6	7	8	9	10
1	1.000	-0.262	0.408	0.413	0.470	0.111	-0.162	0.239	0.297	0.299
2	-0.262	1.000	0.010	0.196	0.149	0.262	0.047	0.096	-0.317	-0.007
3	0.408	0.010	1.000	0.011	0.212	0.567	0.263	0.259	-0.018	-0.003
4	0.413	0.196	0.011	1.000	0.473	0.184	0.081	0.263	0.237	0.466
5	0.470	0.149	0.212	0.473	1.000	0.161	0.185	0.112	0.157	0.522
6	0.111	0.262	0.567	0.184	0.161	1.000	0.238	0.040	-0.070	0.025
7	-0.162	0.047	0.263	0.081	0.185	0.238	1.000	-0.100	0.031	0.133
8	0.239	0.096	0.259	0.263	0.112	0.040	-0.100	1.000	-0.254	-0.126
9	0.297	-0.317	-0.018	0.237	0.157	-0.070	0.031	-0.254	1.000	0.348
10	0.299	-0.007	-0.003	0.466	0.522	0.025	0.133	-0.126	0.348	1.000

Table 1. Correlation Matrix of Features

This table shows the correlation matrix between the results of the PNN. This shows that there is low correlation between each PNN output. Using clinical variables separately as inputs to the PNN and combining the outputs of PNN using GRNN ensemble seems to be a logical idea since it can make use of the lack of correlation between the variables.

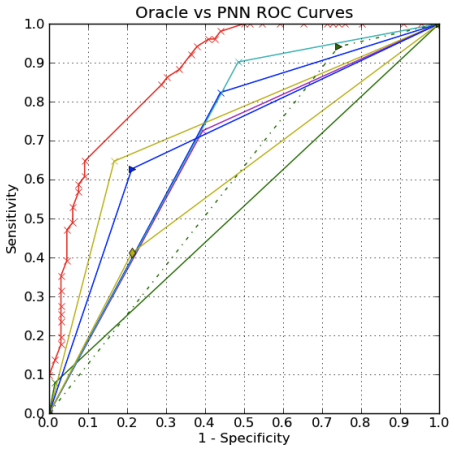

Fig. 5. GRNN Ensemble and PNN ROC Curves. The red curve on the top is from the GRNN Ensemble.

The ROC curve above shows the comparison between each PNN and the GRNN ensemble. The ensemble makes a marked improvement over the input PNNs.

5. Conclusions

The PNN/GRNN ensemble had an AUC value of 83.08%. This outperforms the standalone PNNs by an average of 19.56% AUC and outperforms the best standalone PNN by 9.06% AUC. This result displays that the PNN/GRNN ensemble has the potential to improve the performance of the PNN to a level that can be used in clinical situations such as a prescreener for breast cancer. This will reduce the overall cost of diagnosing patients as well as conserving clinician's time.

However, additional testing and evaluation must be conducted to evaluate performance against intra and inter observer variability (*i.e.* demographics, life "style", additional data sets, etc.).

References

1. National Cancer Institute, *SEER Cancer StatisticsReview 1975-2000*, Table I-1, Estimated New Cancer Cases and Deaths for 2003, 2003.

2. P. Strax, "Make Sure That You Do Not Have BreastCancer," St. Martin's, NY, 1989.

3. J. Y. Lo, J. A. Baker, P. J. Kornguth, J. D. Iglehart, and C. E. Floyd, "Predicting Breast Cancer Invasion with Artificial Neural Networks on the Basis of Mammographic Features," *Radiology*, Vol. 203, pp. 159-163, 1997.

4. C. E. Floyd, J. Y. Lo, A. J. Yun, D. C. Sullivan, and P. J. Kornguth, "Prediction of Breast Cancer Malignancy Using an Artificial Neural Network," *Cancer*, Vol. 74, pp. 2944-98, 1994.

5. D. B. Fogel, E. C. Wasson, and E. M. Boughton, "Evolving Neural Networks for Detecting Breast Cancer," *Cancer Letters*, Vol. 96, pp. 49-53, 1995.

6. D. B. Fogel, E. C. Wasson, E. M. Boughton, and V.W. Porto, "A Step Toward Computer-Assisted Mammography using Evolutionary Programming and Neural Networks," *Cancer Letters*, Vol. 119, pp. 93-97, 1997.

7. D. B. Fogel, E. C. Wasson, E. M. Boughton, and V.W. Porto, "Evolving Artificial Neural Networks for Screening Features from Mammograms," *Artificial Intelligence in Medicine*, Vol. 14, pp. 317-326, 1998.

8. D. B. Fogel, E. C. Wasson, E. M. Boughton, V. W. Porto, and P. J. Angeline, "Linear and Neural Models for Classifying Breast Masses," *IEEE Trans. Medical Imaging*, Vol. 17, No. 3, pp. 485-488, 1998.

9. W. H. Land, Jr., T. Masters, J. Y. Lo, and D. McKee, "Using Evolutionary Computation to Develop Neural Network Breast Cancer Benign/Malignant Classification Models," *4th World Conference on Systemics, Cybernetics and Informatics*, Vol. 10, pp. 343-7, 2000.

10. D. B. Fogel, E. C. Wasson, E. M. Boughton, V. W. Porto, and P. J. Angeline, "Linear and Neural Models for Classifying Breast Masses," *IEEE Trans. Medical Imaging*, Vol. 17, No. 3, pp. 485-488, 1998.

11. W. H. Land, Jr., T. Masters, and J. Y. Lo, "Application of a New Evolutionary Programming/Adaptive Boosting Hybrid to Breast Cancer Diagnosis," IEEE Congress on Evolutionary Computation Proceedings, 2000.

12. http://www.reversideonline.com/services/csncer/cancer-screening-recommendations.cfm

13. K. Price, and R. Storn, "Differential Evolution", *Dr. Dobb's Journal*, pp. 18-24, April, 1997.

14. W. H. Land, Jr., D. Margolis, M. Kallergi, and J. J. Heine, (2010) 'A kernel approach forensemble decision combinations with two-view mammography applications', *Int. J. Functional Informatics and Personalised Medicine*, Vol.2, 2010

15. W. H. Land, L. Wong, D.W. McKee, T. Masters, F. R. Anderson, " Breast Cancer Computed Aided Diagnosis (CAD) Using a Recently Developed SVM/GRNN Oracle Hybrid"Paper number 425, for SMCC'03, 0-7803-7952-7-7/03/S17,2003