Archived at the Flinders Academic Commons:

http://dspace.flinders.edu.au/dspace/

'This is the peer reviewed version of the following article: Ghassempour, S., McLeod, P., Verma, B. and Maeder, A.J. (2012). Comparing data mining with ensemble classification of breast cancer masses in digital mammograms. In Proceedings of the Second Australian Workshop on Artificial Intelligence in Health. Aachen, Germany: CEUR-WS. AIH 2012. Sydney. Dec 2012, pp. 55-63.,

which has been published in final form at

http://ceur-ws.org/Vol-941/aih2012_GhassemPour.pdf

Comparing Data Mining with Ensemble Classification of Breast Cancer Masses in Digital Mammograms

Shima Ghassem Pour[1], Peter Mc Leod[2], Brijesh Verma[2], and Anthony Maeder[1]

[1] School of Computing, Engineering and Mathematics,
University of Western Sydney
Campbelltown, New South Wales, Australia
[2] School of Information and Communication Technology,
Central Queensland University
Rockhampton, Queensland, Australia
{shima.ghassempour,mcleod.ptr}@gamil.com,
b.verma@cqu.edu.au,a.maeder@uws.edu.au

Abstract. Medical diagnosis sometimes involves detecting subtle indications of a disease or condition amongst a background of diverse healthy individuals. The amount of information that is available for discovering such indications for mammography is large and has been growing at an exponential rate, due to population wide screening programmes. In order to analyse this information data mining techniques have been utilised by various researchers. A question that arises is: do flexible data mining techniques have comparable accuracy to dedicated classification techniques for medical diagnostic processes? This research compares a model-based data mining technique with a neural network classification technique and the improvements possible using an ensemble approach. A publicly available breast cancer benchmark database is used to determine the utility of the techniques and compare the accuracies obtained.

Keywords: latent class analysis, digital mammography, breast cancer, clustering, classification, neural network.

1 Introduction

Medical diagnosis is an active area of pattern recognition with different techniques being employed [17, 19, 12]. The expansion of digital information for different cohorts [15] has allowed researchers to examine relationships that were previously not uncovered due to the limited nature of information as well as a lack of techniques being available for the analysis of large data sets. Flexible data mining techniques have the capacity to predict disease and reveal previous unknown trends.

The question that arises is whether the relationships that are revealed by those techniques are as accurate or as comparable as techniques that are specifically developed for other purposes, such as a diagnostic system for a particular

disease or condition. This research aims at contrasting the cluster analysis technique (Latent Class Analysis) of Ghassem Pour, Maeder and Jorm [4] against a baseline neural network classifier, and then considers the effects of applying an ensemble technique to improve the accuracies obtained.

The organisation of this paper is that section two provides a background on the approaches that have been utilised for breast cancer diagnosis, sections three and four detail the proposed techniques for comparison, section five outlines the experimental results obtained and conclusions are presented in section six.

2 Background

Medical diagnosis is a problematic paradigm in that complex relationships can exist in the diagnostic features that are utilised to map to a resultant diagnosis about the disease state. In different cases the state of the disease condition itself can be marked by stages where the diagnostic symptoms or signs can be subtle or different to other stages of the disease. This means that there is often not a clean mapping between the diagnostic features and the diagnosis [13, 14].

Breast cancer screening using mammography provides an exemplar of this situation. Early detection and treatment have been the most effective way of reducing mortality [2] however Christoyianni et al. [1] noted that 10-30% of breast cancers remain undetected while 15-30% of biopsies are cancerous. Taylor and Potts [22] made similar observations in their research. There are many reasons why various cancers can remain undetected. These include the obfuscation of anomalies by surrounding breast tissue, the asymmetry of the breast, prior surgery, natural differences in breast appearance on mammograms, the low contrast nature of the mammogram itself, distortion from the mammographic process and even talc or powder on the outside of the breast making it hard to identify and discriminate anomalies. Even if an anomaly is detected, a high rate of false positives exist [17, 18].

Clustering has provided a widely used mechanism for organising data into similar groupings. The usage of clustering has also been extended to classifiers and detection systems in order to improve detection and provide greater classification accuracy. Kim et al. [9] developed a classifier based on Adaptive Resonance Theory (ART2) where micro-calcifications were grouped into different classes with a three-layered back propagation network performing the classification. The system achieved 90% sensitivity (Az of 0.997) with a low false positive rate of 0.67 per cropped image.

Other researchers such as Mohanty, Senapati and Lenka [16] explored the application of data mining techniques to breast cancer diagnosis. They indicated that data mining medical images would allow for the collection of effective models, rules as well as patterns and reveal abnormalities from large datasets. Their approach was to use a hybrid feature selection technique with a decision tree classifier to classify breast cancer. They utilised 300 images from the MIAS database. They achieved a classification accuracy of 97.7% however their dataset images contained microcalcifications as well as mass anomalies.

3 Latent Class Analysis and Data Mining

Latent Class Analysis (LCA) has been proposed as a mechanism for improved clustering of data over traditional clustering algorithms like k-means [11]. LCA classifies subjects into one of K unobserved classes based on the observed data, where K is a constant and known parameter. These latent or potential classes are then refined based upon their statistical relationships with the observed variables.

LCA is a probabilistic clustering approach: although each object is assumed to belong to one cluster, there is uncertainty about an object's membership of a cluster [11, 10]. This type of approach offers some advantages in dealing with noisy data or data with complex relationships between variables, although as an iterative method there is always some chance that it will be susceptible to noise and in some cases fail to converge.

An advantage of using a statistical model is that the choice of the cluster criterion is less arbitrary. Nevertheless, the log-likelihood functions corresponding to LC cluster models may be similar to the criteria used by certain non-hierarchical cluster techniques [18]. Another advantage of the model-based clustering approach is that no decisions have to be made about the scaling of the observed variables: for instance, when working with normal distributions with unknown variances, the results will be the same irrespective of whether the variables are normalized or not.

Other advantages are that it is relatively easy to deal with variables of mixed measurement levels (different scale types) and that there are more formal criteria to make decisions about the number of clusters and other model features [3]. We have successfully applied LCA for cases in health data mining when the anomalous range of variables results in more clusters than have been expected from a causal or hypothesis based approach [5]. This implies that in some cases LCA may be used to reveal associations between variables that are more subtle and complex.

Unsupervised clustering requires prior specification of the number of clusters K to be constructed, implying that a model for the data is necessary which provides K. The binary nature of the diagnosis problem implies that K=2 should be used in ideal circumstances, but the possibility exists that allowing more clusters would give a better solution (e.g. by allowing several different classes within the positive or negative groups). Consequently a figure of merit is needed to establish that the chosen K value is optimal. In this research the Bayesian Information Criteria (BIC) is determined for the mass dataset in order to gauge the best number of clusters.

Repeated application of the clustering approach can also lead to different solutions due to randomness in starting conditions. In this work we used multiple applications of the clustering calculations to allow improvement in the results, in an ensemble-like approach. Our improvement strategy was based on selection of the most frequent membership of classes per element, over different numbers of clustering repetitions.

Comparing Data Mining with Ensemble Classification

4   Neural Network and Ensemble Methods

Neural networks have been advocated for breast cancer detection by many researchers. Various efforts to refine classification performance have been made, using a number of strategies involving some means of choice between alternatives. Ensembles have been proposed as a mechanism for improving the classification accuracy of existing classifiers [6] providing that constituents are diverse.

Zhang et al. [23] partitioned their mass dataset obtained from the DDSM into several subsets based on mass shape and age. Several classifiers were then tested and the best performing classifier on each subset was chosen. They used SVM, k-nearest neighbour and Decision Tree (DT) classifiers in their ensemble and achieved a combined classification accuracy of 72% that was better than any individual classifier.

Surrendiran and Vadivel [21] proposed a technique that could determine what features had the most appropriate correlation on classification accuracy and achieved 87.3% classification accuracy. They achieved this by using ANOVA DA, Principal Component Analysis and Stepwise ANOVA analysis to determine the relationship between input feature and classification accuracy.

Mc Leod and Verma [14] utilised a clustered ensemble technique that relied on the notion that some patterns could be readily identified through clustering (atomic). Other patterns that were not so easily separable (non-atomic) were classified by a neural network. The classification process involved an initial lookup to determine if a pattern was associated with an atomic class however for non-atomic classes a neural network ensemble that had been created through an iterative clustering mechanism (to introduce diversity into the ensemble) was employed. The advantage of this technique is that the ensemble was not adversely affected by outliers (atomic clusters). This technique was applied to the same mass dataset as utilised in this research and achieved a classification accuracy of 91%.

The ensemble utilised in this research was created by fusing together (using the majority vote algorithm) constituent neural networks that were created by varying the number of neurons in the hidden layer to create diverse networks for incorporation into an ensemble classifier.

5   Experimental Results

The experiments were conducted for LCA and neural network techniques and the related ensemble approaches using mass type anomalies from the Digital Database of Screening Mammography (DDSM) [7]. The features used for classification purposes coincided with the Breast Imaging Reporting and Data System (BI-RADS) as this is how radiologists classify breast cancer. The BI-RADS features of density, mass shape, mass margin and abnormality assessment rank are used as they have been proven to provide good classification accuracy [20]. These features are then combined with patient age and a subtlety value [7].

Experiments were performed utilising the clustering technique of Ghassem

Comparing Data Mining with Ensemble Classification

Pour, Maeder and Jorm [4] on this dataset. This was achieved using the Latent Gold® software package. The first step was to utilise the analysis feature of LatentGold® to calculate the BIC value and the classification error rate. This information appears in Table 1 below, with Npar designating the resulting parameter value associated with the LCA.

Table 1. LCA Cluster optimisation based on Classification Error.

Clusters	BIC	Npar	Classification Error
2	1238.8	30	0.0303
3	1240.6	38	0.0403
4	1241.8	46	0.0446
5	1254.1	54	0.0470

Minimisation of BIC and the Classification Error determines the best number of clusters for the LCA analysis in terms of classification accuracy and this was found to be 2 clusters. Nevertheless, it might be expected that some further complexity could be identified in higher numbers of clusters, where multiple clusters may exist for either positive or negative classes. The results obtained when cases of more than 2 clusters were merged to form the dominant positive and negative classes, are detailed in Table 2. These results show the instability

Table 2. LCA Classification Technique Accuracy.

Clusters	Accuracy %
2	87.2
3	56.7
4	43.2
5	32.8

of LCA classification for this dataset at higher numbers of clusters, for example the 2-cluster solution gives better accuracy than the 3-cluster solution (merging into 2 clusters) and so forth. From this we conclude that the natural 2-cluster solution is indeed optimal.

In order to provide a comparison, further experiments were performed using a neural network and then applying an ensemble classifier. The neural network and ensemble techniques were implemented in MATLAB® utilising the neural network toolbox. The parameters utilised are detailed in the Table 3 below. Experiments were first performed with a neural network classifier alone, in order to provide a baseline for measuring the classification accuracy on the selected dataset. The results obtained are detailed in Table 4 below. Further experiments were then performed utilising an ensemble technique with a summary of the neural network test results using ten-fold cross validation, as detailed in Table 5 below.

Comparing Data Mining with Ensemble Classification

Table 3. Neural network configuration parameters.

Parameter	Value
Hidden Layers	1
Transfer Function	Tansig
Learning Rate	0.05
Momentum	0.7
Maximum Epochs	3000
Root Mean Square Goal	0.001

Table 4. Neural network classification technique accuracy.

Hidden Neurons	Accuracy (%)
13	80
25	80
52	90
111	79

Table 5. NN-ensemble classification technique accuracy.

Networks	Hidden Neurons in Ensemble	Accuracy (%)
6	24,5,15,32,31,43	94
10	24,5,15,32,31,43,50,75,38,59	96.5
13	24,5,15,32,31,43,50,75,38,59,68,79,116	98
15	24,5,15,32,31,43,50,75,38,59,68,79,116,146,14	96

Experiments were also performed for the ensemble-like optimising of results from the LCA technique. It is difficult to match this process directly with the complexity used for the NN-ensemble experiments, so the number of repetitions has been modelled on plausible choice based on dataset size of 100 cases. The results for these experiments are shown in Table 6 below.

Table 6. LCA-ensemble classification technique accuracy.

Repetitions	Accuracy (%)
10	87
20	89
40	91
70	94

6   Discussion and Conclusions

Examination of the results from Tables 1 to 6 demonstrates that the accuracy obtained with the LCA technique is below that of the baseline classification

Comparing Data Mining with Ensemble Classification

performed with the neural network. However an ensemble oriented approach enabled improvement of the results from both techniques.

In order to examine the results more closely the sensitivity, specificity and positive predictive value have been calculated for the best performing results for each of the trialled techniques, shown below in Table 7.

Sensitivity is the True Positive diagnosis divided by the True Positive and False Negative components. Sensitivity can be thought of as the probability of detecting cancer when it exists.

Specificity is the True Negative component divided by the True Negative component plus the False Positive component. Specificity can be thought of as the probability of being correctly diagnosed as not having cancer.

Positive Predictive Value (PPV) is the True Positive component divided by the True Positive component plus the False Positive component. PPV is the accuracy of being able to identify malignant abnormalities. The latent class analysis

Table 7. Performance results for the proposed techniques.

Technique	Performance(%)		
	Sensitivity	Specificity	PPV
Latent Class Analysis	80.5	93.9	95.0
LCA-ensemble	82.7	95.2	96.0
Neural Network	91.6	88.4	90.0
NN-ensemble	97.0	97.9	99.0

technique was not as sensitive as the neural network but had better specificity and a higher positive predictive value than the neural network. Both ensemble approaches resulted in substantially better performance, which of course must be traded off against the increased computational cost. The NN-ensemble technique performed the best with good sensitivity, specificity and a high positive predictive value.

The flexibility of clustering techniques such as LCA provides a mechanism for gaining insight from large data repositories. However once patterns in the data become evident it would appear that other less flexible but more specialised techniques could be utilised to obtain analysis at a higher degree of granularity of the data in question.

A summary of the overall performance of the techniques employed in this paper are presented in Figure 1. The optimal LCA-ensemble result, while less than the optimal NN-ensemble result, is obtained with somewhat less processing effort and complexity, and further improvement may be possible.

Future work could look at extending the comparison of LCA with other data mining algorithms to determine their applicability. Breast cancer represents only one problem domain and applying these methods to other datasets would be a logical extension. Our future research will include more experiments with LatentGold® on other breast cancer datasets to determine how different numbers of clusters produce different classification results for a more detailed analysis.
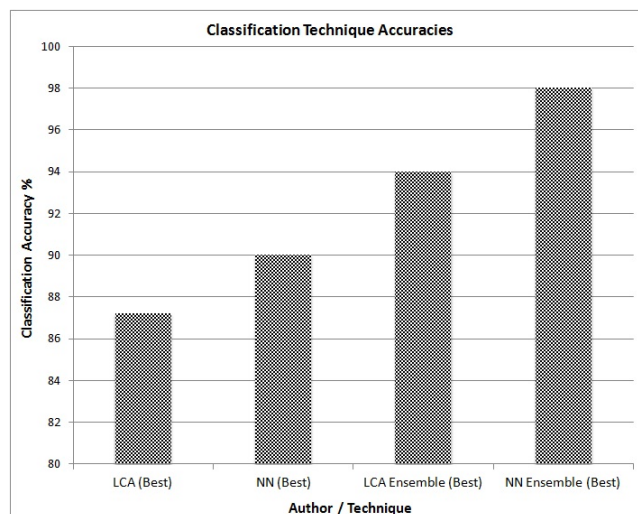
Comparing Data Mining with Ensemble Classification

Fig. 1. Comparative Classification Accuracies.

References

1. Christoyianni, I., Koutras, A., Dermatas, E., Kokkinakis, G.: Computer Aided Diagnosis of Breast Cancer in Digitized Mammograms. Computerized Medical Imaging and Graphics 26(5), 309-319 (2002)
2. DeSantis, C., Siegel, R., Bandi, P., Jemal, A.:Breast Cancer Statistics, 2011. CA: A Cancer Journal for Clinicians 61(6), 408-418 (2011)
3. Fraley, C., Raftery, A.: Model-based Clustering, Discriminant Analysis, and Density Estimation. Journal of the American Statistical Association 97(458), 611-631(2002)
4. Ghassem Pour, S., Maeder, A., Jorm, L.: Constructing a Synthetic Longitudinal Health Dataset for Data Mining. DBKDA 2012, The Fourth International Conference on Advances in Databases, Knowledge, and Data Applications.86-90 (2012)
5. Ghassem Pour, S., Maeder, A., Jorm, L.: Validating Synthetic Health Datasets for Longitudinal Clustering. The Australasian Workshop on Health Informatics and Knowledge Management (HIKM 2013) 142, to appear (2013)
6. Gou, S., Yang, H., Jiao, L., Zhuang, X.: Algorithm of Partition Based Network Boosting for Imbalanced Data Classification. The International Joint Conference on Neural Networks (IJCNN).1-6. IEEE (2010)
7. Heath, M., Bowyer, K., Kopans, D., Moore, R., Kegelmeyer, P.: The Digital Database for Screening Mammography. Proceedings of the 5th International Workshop on Digital Mammography.212-218 (2000)
8. Hofvind, S., Ponti, A., Patnick, J., Ascunce, N., Njor, S., Broeders, M., Giordano, L., Frigerio, A., Tornberg, S.: False-positive Results in Mammographic Screening for Breast Cancer in Europe: a literature review and survey of service screening programmes. Journal of Medical Screening 19(1), 57-66 (2012)
9. Kim, J., Park, J., Song, K., Park, H.: Detection of Clustered Microcalssifications on Mammograms Using Surrounding Region Dependence Method and Artificial Neural Network. The Journal of VLSI Signal Processing 18(3),251-262 (1998)

Comparing Data Mining with Ensemble Classification

10. Lanza, S., Flaherty, B., Collins, L.: Latent Class and Latent Transition Analysis. Handbook of Psychology. 663-685 (2003)
11. Magidson, J., Vermunt, J.: Latent Class Models for Clustering: A Comparison with k-means. Canadian Journal of Marketing Research 20(1), 36-43 (2002)
12. Malich, A., Schmidt, S., Fischer, D., Facius, M., Kaiser, W.: The Performance of Computer-aided Detection when Analyzing Prior Mammograms of Newly Detected Breast Cancers with Special Focus on the Time Interval from Initial Imaging to Detection. European Journal of Radiology 69(3),574-578 (2009)
13. Mannila, H.: Data mining: Machine learning, Statistics, and Databases. Proceedings of Eighth International Conference on Scientific and Statistical Database Systems.2-9 IEEE (1996)
14. McLeod, P., Verma, B.: Clustered Ensemble Neural Network for Breast Mass Classiffcation in Digital Mammography. In: The International Joint Conference on Neural Networks (IJCNN). 1266-1271 (2012)
15. Mealing, N., Banks, E., Jorm, L., Steel, D., Clements, M., Rogers, K.: Investigation of Relative Risk Estimates from Studies of the Same Population with Contrasting Response rates and Designs. BMC Medical Research Methodology 10(1), 10-26 (2010)
16. Mohanty, A., Senapati, M., Lenka, S.: A Novel Image Mining Technique for Classification of Mammograms Using Hybrid Feature Selection. Neural Computing & Applications. 1-11 (2012)
17. Nishikawa, R., Kallergi, M., Orton, C., et al.: Computer-aided Detection, in its present form, is not an Effective aid for Screening Mammography. Medical Physics 33(4), 811-814 (2006)
18. Nylund, K., Asparouhov, T., Muthen, B.: Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. Structural Equation Modeling 14(4), 535-569 (2007)
19. Oh, S., Lee, M., Zhang, B.: Ensemble Learning with Active Example Selection for Imbalanced Biomedical Data Classification. IEEE/ACM Transactions on Computational Biology and Bioinformatics 8(2), 316-325 (2011)
20. Sampat, M., Bovik, A., Markey, M.: Classification of Mammographic lesions into BIRADS Shape Categories Using the Beamlet Transform. In: Proceedings of SPIE, Medical Imaging: Image Processing. 16-25. SPIE(2005)
21. Surrendiran, B., Vadivel, A.: Feature Selection Using Stepwise ANOVA, Discriminant Analysis for Mammogram Mass Classification. International Journal of Recent Trends in Engineering and Technology 3, 55-57 (2010)
22. Taylor, P., Potts, H.: Computer Aids and Human Second Reading as Interventions in Screening Mammography: two systematic reviews to compare effects on cancer detection and recall rate. European Journal of Cancer 44(6), 798-807 (2008)
23. Zhang, Y., Tomuro, N., Furst, J., Raicu, D.: Building an Ensemble System for Diagnosing Masses in Mammograms. International Journal of Computer Assisted Radiology and Surgery 7(2), 323-329 (2012)