Research Article

Textural Classification of Mammographic Parenchymal Patterns with the SONNET Selforganizing Neural Network

Daniel Howard,[1] Simon C. Roberts,[1] Conor Ryan,[2] and Adrian Brezulianu[3]

[1] *QinetiQ, St Andrews Road, Malvern, Worcestershire WR14 3PS, UK*
[2] *Department of Computer Science and Information Systems, College of Informatics and Electronics, University of Limerick, Ireland*
[3] *Faculty of Electronics and Telecommunications, "Gh.Asach" Technical University of Iasi, 700050 Iasi IS, Romania*

Correspondence should be addressed to Daniel Howard, dr.daniel.howard@gmail.com

In nationwide mammography screening, thousands of mammography examinations must be processed. Each consists of two standard views of each breast, and each mammogram must be visually examined by an experienced radiologist to assess it for any anomalies. The ability to detect an anomaly in mammographic texture is important to successful outcomes in mammography screening and, in this study, a large number of mammograms were digitized with a highly accurate scanner; and textural features were derived from the mammograms as input data to a SONNET selforganizing neural network. The paper discusses how SONNET was used to produce a taxonomic organization of the mammography archive in an unsupervised manner. This process is subject to certain choices of SONNET parameters, in these numerical experiments using the craniocaudal view, and typically produced O(10), for example, 39 mammogram classes, by analysis of features from $O(10^3)$ mammogram images. The mammogram taxonomy captured typical subtleties to discriminate mammograms, and it is submitted that this may be exploited to aid the detection of mammographic anomalies, for example, by acting as a preprocessing stage to simplify the task for a computational detection scheme, or by ordering mammography examinations by mammogram taxonomic class prior to screening in order to encourage more successful visual examination during screening. The resulting taxonomy may help train screening radiologists and conceivably help to settle legal cases concerning a mammography screening examination because the taxonomy can reveal the frequency of mammographic patterns in a population.

1. INTRODUCTION

Nationwide mammography screening (NMS) is the most successful method for the early detection of breast cancer [1]. There exist differing opinions concerning the details of an ideal NMS programme, with a body of opinion arguing that it should involve women over the age of 40 attending an annual mammography examination that obtains X-ray of each breast (for the two standard views), with technicians ensuring that the subject is imaged consistently each year to allow a comparison of the images over time, which is essential to mammography screening. Longitudinal comparisons of mammograms are quite revealing. Consider Figure 1, which pertains to a woman who assisted frequent screening for over 20 years. This longitudinal comparison llustrates the involution of the parenchyma with age of the subject as it is replaced by adipose tissue. Exceptions relate to extensive fibrosis or adenosis for which this involution to all intent and purposes cannot be observed [2]. Also, mammograms of subjects starting hormone replacement therapy appeared to us to restore an earlier appearance. By comparing images longitudinally, it may be possible to detect a lesion as an abnormality that grows in time in contrast to the gradual but perceivable retraction of the parenchyma.

An NMS programme also requires an integrated clinical team involving a pathologist, radiologist, oncologist and surgeon. Ideally, when a patient undergoes an invasive procedure then the radiologist should obtain X-ray images of sections of the biopsy specimen to visually compare these with the original mammogram so as to improve his or her abilities at detecting lesions.
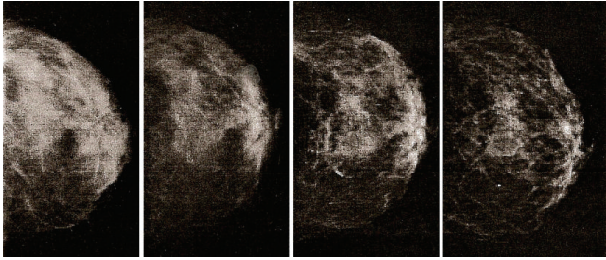
FIGURE 1: Longitudinal study showing the process of involution of the parenchyma with age of subject (digitized from L. Tabár archive). The CC view is shown from examinations in 1980, 1989, 1997, and 2002 (left to right).
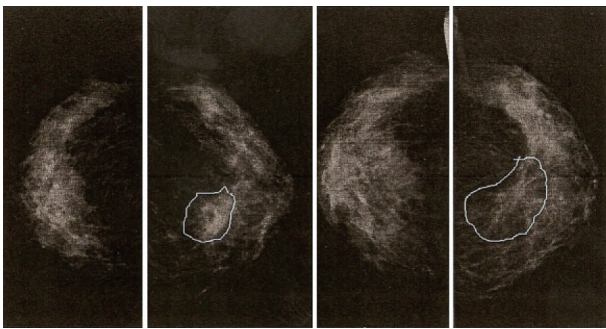

FIGURE 2: Left to right: longitudinal progression of a distortion (digitized from L. Tabár archive).
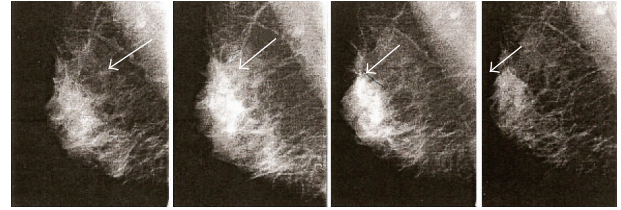

FIGURE 3: Left to right: longitudinal progression of a distortion (digitized from L. Tabár archive).

FIGURE 4: Left to right: longitudinal progression of a distortion (digitized from L. Tabár archive).

If the NMS programme were to encourage the same radiologist to screen the same women for a number of years, it is possible that fewer women would be recalled without good reason. When fewer women are recalled without good reason, higher levels of compliance and participation of women usually follow, and if so, then the benefits of screening: early detection and arguably lower cancer mortality rates should follow.

To date, computer-aided detection (CAD) of cancer does not appear to have achieved significant penetration in NMS services. A challenge of CAD is to help with difficult-to-perceive and subtle lesions also known as distortions, which are strong indications of breast cancer. Calcifications are not as significant because up to 80% of calcification occurrences in screening are benign [3].

Architectural distortions are variable in shape and size and difficult to pick-up in early detection even by an experienced radiologist. Figures 2, 3, and 4 illustrate this typical progression for consecutive examinations over time, the X-ray images can be seen to differ most significantly with an impression of a "pulling" effect at the site of the distortion. These examples are very obvious but it is necessary to detect very subtle distortions early on. By observing the foremost radiologists at work it became apparent that anomalies in image texture (angles in the orientation of textural patterns) accounted for their intuitions about the presence of a lesion. Radiologists have used emotive terms to explain

how they pick-up on subtle anomalies and the detection of mammographic lesions has been described as "visual art" [3]. This together with the variability and elusiveness of architectural distortions motivated us to develop machine vision algorithms to construct a taxonomy of mammography textural patterns rather than to produce a more standard CAD tool. It was observed that unusual angular orientations of texture alerted the experienced radiologist to a subtle distortion. This observation, however, is an oversimplification for the purpose of this discussion, as clinical knowledge is essential for the detection of the lesion.

2. A COMPUTATIONAL TOOL FOR MAMMOGRAM CLASSIFICATION

How to best assist an NMS programme using a computational tool for image analysis? This is the research question that we set out to answer. In cancer screening, the number of normal cases far exceeds the number of suspected lesions (perhaps to a ratio of 50 to 1). This can lead to fatigue in human-facilitated screening and a lack of lesion examples with which radiologists are trained. Furthermore, the high variability of normal cases has led it to be said that a mammogram can equate to a fingerprint in its subject identification ability.

Our aim was to apply artificial intelligence (AI) and image analysis techniques to answer the aforementioned research question. In order to study the problem quantitatively, we deployed an accurate Lumysis scanner to digitize a large number of pristine film mammograms from a long-established archive that contains arguably the greatest density of high-quality mammograms in the world [4]. Leading

radiologists of the breast such as Wolf and Tabár considered "parenchymal types" [3, 5] to establish a subjective classification of mammograms into a few categories. This paper addresses the automation of this classification process by using an unsupervised, selforganizing method of AI known as SONNET (described in Section 3). The automated classification scheme is based on image textural analysis because, as previously mentioned, we surmised that textural patterns and orientations (the angles that can be perceived in this texture) are the important criteria of visual inspection for radiologists; they constitute a type of visual "algedonic alert" [6] to the presence of a subtle distortion in the mammogram.

A categorized organization of a mammogram archive by texture could have many potential uses. Learning and training are obvious but for example with the objective of higher accuracy in screening, the mammograms could be sorted by taxonomic pattern to enable the organized viewing that gives time for radiologists to become accustomed to the background texture. This normal texture could then more easily be compared against the anomalous texture of architectural distortions and lesions. Also, patients could be tracked longitudinally based on their progression through the taxonomy of parenchymal types [4, 5, 7]. This could provide additional information for a mammographic examination by allowing comparison to other patients who experienced a similar progression. The aberrations of normal breast development may be studied with this taxonomy. They could offer clues to reasons for a higher incidence of breast cancer in a population, for instance, according to lifestyle or genetic profile. Conceivably, a taxonomy could help to settle a legal case as it quantifies how common is the parenchymal pattern, to help settle a dispute concerning an early breast cancer warning that was missed.

3. THE SONNET CLASSIFIER

The artificial neural network known as SONNET [8] consists of an array of classifiers connected to an input field as shown in Figure 5. Input patterns are presented in turn to the input field and *typical* patterns are gradually encoded as follows. The constituent classifiers compete to encode each pattern such that the classifier with the best match to the current input tends to adapt itself more than the other classifiers. This *winning* classifier adapts itself by partially encoding the current input pattern on *weighted* excitatory connections from the input field. Furthermore, the classifier adapts weighted inhibitory connections to the other competing classifiers, thus allowing the winning classifier to suppress its competitors. The classifiers consequently diverge so that each responds only to input patterns which are similar to the pattern encoded on its excitatory connections.

SONNET is a selforganizing neural network based on adaptive resonance theory [2] that encodes classifications using unsupervised learning. The SONNET architecture is shown in Figure 5 where a field of input neurons is connected to a field of classification neurons via weighted connections. An input pattern is a pattern of relative neural activity in the input field at any given moment, and a set of input patterns
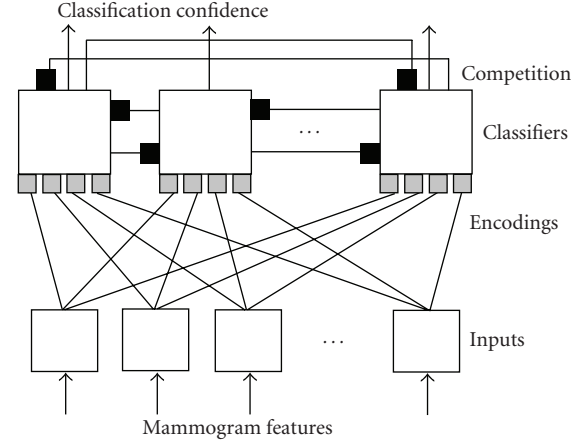

FIGURE 5: SONNET architecture.

is presented to SONNET by setting the activity of the input neurons to each pattern in turn for a fixed duration.

Each input neuron is connected to each classification neuron by an excitatory weighted connection, and the weight on each connection is continually adapted via a learning rule (this is similar to Hebbian learning [9, 10] which postulates that memory is stored in the synaptic weights and learning is the process that changes those weights) such that the connection becomes stronger when the two corresponding neurons are simultaneously active. Furthermore, higher activity on the two neurons causes the connection to become stronger more rapidly. The maximum rate at which the weights can change governs the network's learning speed and this is regulated by controlling parameters that are set prior to a SONNET run.

The relative pattern of excitatory weights on connections to a single-classification neuron represents the so-called *prototype* for that classifier. The excitatory input to a classification neuron is based on two measures. The first measure is based on the size of the excitatory weights so that a large excitatory input can be achieved when strong weights gate high-input activations. The second measure quantifies how well the prototype matches the current input pattern such that a large excitatory input can be achieved for a good match even when the prototype is represented by small excitatory weights. A large excitatory input to a classification neuron allows the neuron to gain a high activation in response to the current input pattern. This activation represents the confidence with which the neuron classifies the input pattern. The learning speed can be set to allow the prototype to form gradually from repeated exposure to input patterns, such that the prototype encodes a *generalization* for multiple similar input patterns. The classification neuron can then obtain a high activation when any one of these input patterns occurs and thus it classifies these patterns together.

The classifiers compete to encode each input pattern such that the classifier with the best match to the current input tends to adapt itself more than the other classifiers, thus further improving its competitive advantage. Each classifier
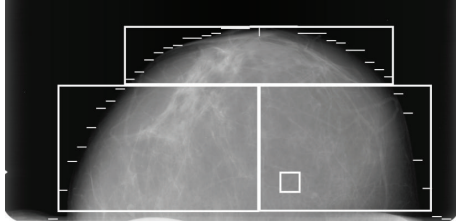
FIGURE 6: Mammogram regions.

is connected to all other classifiers by an inhibitory weighted connection that is again adapted via a learning rule such that a connection becomes stronger (i.e., more inhibitory) when the corresponding neurons are simultaneously active. Classifiers that have partially encoded similar patterns thus compete strongly against each other, causing one classifier to eventually suppress its competitors. The classifiers consequently diverge to encode different input patterns so that only one classification neuron achieves a high activation in response to each input pattern. When the input pattern changes, the activation of a previously excited classifier can decrease due to both passive decay and inhibition from competing classifiers that better represent the new input pattern.

SONNET performs real-time learning by continually adapting its weights in the selforganizing manner described above. The learning algorithm is unsupervised and there is no reinforcement of any kind from an external source to judge the emergent classifications against expected classifications. The network is initialized with random small weights and the classifiers compete such that the most common input patterns are encoded first and the less common patterns are encoded more gradually.

The selforganizing behaviour causes SONNET to be susceptible to the so-called *stability-plasticity dilemma* [2], which states that a network should always remain adaptive to learn new patterns (i.e., have plasticity) without degrading well-formed encodings for previously learned patterns (i.e., have stability). SONNET achieves plasticity due to the aforementioned learning algorithm but it also achieves stability by reducing the learning speed at a single classifier when the size of the classifier's excitatory weights become large. A classifier can only gain large excitatory weights after it has encoded a good representation for one or more input patterns, and a stable classifier is said to be *committed* with excitatory weights that constitute a long-term memory of the encoding.

For the current application, each input pattern represented features extracted from a mammogram. A set of mammograms was selected with which to *train* SONNET, and each presentation of the full mammogram set is known as an epoch. SONNET typically learned by adapting itself over many epochs until a *stable* set of classifiers could classify each mammogram with a significant degree of confidence. The order of mammogram presentation was randomized on each epoch. This reduced the likelihood of an unstable

classifier from oscillating between similar yet significantly different potential classes.

SONNET is a highly dynamic system which is controlled by many parameters as discussed in other recent research presentations in [11–13]. It is a fully unsupervised system which encodes classes via selforganization in response to the input patterns. However, the manual specification of SONNET's controlling parameters allows a degree of supervision. For example, a number of parameters govern SONNET's learning speed which in turn influences the number of classes encoded. The greatest learning speed produces *one-shot learning* where SONNET simply memorizes each input pattern. Slower learning produces broader classes, where a single classifier can represent multiple similar patterns by forming encodings that generalize the characteristic features of the class.

Multiple SONNET runs were conducted using different randomized initial weights on the connections within the network. This allowed different encodings to form on each run. SONNET's controlling parameters were also varied on different runs to change the learning speed. SONNET comprised at most 80 classifiers though the actual number was set in accordance with the learning speed. Each run terminated after 100 epochs but the final epoch did not necessarily represent the optimum SONNET state. Section 4.5 explains how the optimum SONNET runs and epochs were identified.

4. DEVELOPING A MAMMOGRAM TAXONOMY USING UNSUPERVISED CLASSIFICATION

The development of an unsupervised classification scheme to produce a mammogram taxonomy had to address the following issues: input feature extraction; input feature selection to produce a minimal set of features which best characterize the input cases; input feature preprocessing prior to presentation to the classification system; classifier development; and the definition of classification performance measures in order to compare the classifiers resulting from different SONNET runs. These issues are discussed in the following subsections.

4.1. Mammogram feature extraction

450 mammograms were chosen for the current study. These mammograms represented the CC left and right views for 225 different patients. The mammograms were X-rayed between 1990 and 2002; and they were of a highly consistent top quality. Most mammograms displayed normal breast tissue but 49 of the patients had been diagnosed as having breast cancer. Subtle cancerous lesions were evident in the mammograms corresponding to these patients.

The breast tissue in a mammogram must be segmented from the background before mammogram features can be extracted. This was achieved by locating maximal brightness gradients to produce multiple hypotheses for the actual breast margin. The best hypothesis was identified by optimizing contour shape and smoothness. The location of the nipple was also estimated to ascertain three different regions within the breast as shown in Figure 6. These regions are the retroareolar region (behind the nipple), an axillar region
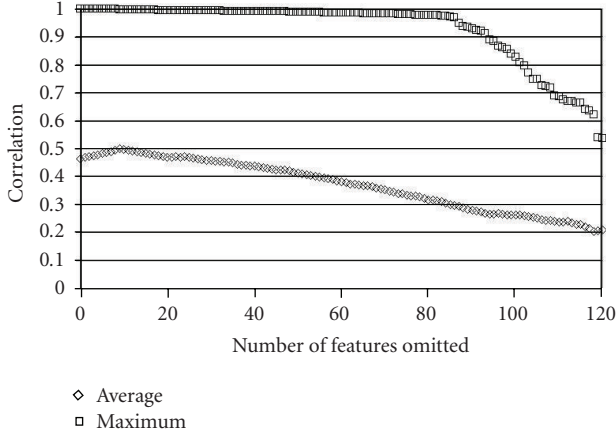
Figure 7: Average and maximum correlations between mammogram features against the number of features omitted.

(outer) and a medial region (inner). The identification of the breast margin allowed equivalent regions to be defined on different mammograms by specifying positions relative to the nipple location.

Standard image processing techniques were used to extract the following information from each of the three regions: brightness distribution, contrast distribution, and textural measures. Brightness was calculated in a 10 mm square (as depicted in Figure 6) which was swept over each region to give the brightness distribution as represented by minimum, maximum, average, and standard deviation values. The same procedure was used to calculate the contrast distribution.

Textural measures were calculated by accumulating co-occurrence matrices over each region. The following 9 textural features were calculated from each matrix: angular second moment, inverse difference moment, contrast, entropy, sum entropy, difference entropy, and three correlation measures. Furthermore, co-occurrence matrices were generated for each of four orientations: 0, 45, 90, and 135 degrees. Hence 12 matrices were generated; four orientations for the three regions.

The above processing resulted in 132 image features for each mammogram. Note that the mammogram for the left breast was flipped horizontally before processing to map the axillar and medial regions onto those for the right breast, and to give the appropriate orientations for the textural features.

4.2. Mammogram feature preprocessing

The extracted image features constitute an input feature vector that can be presented to SONNET's input field. However, each dimension of the input feature vector must be normalized so that each feature varies over the same range. This prevents individual dimensions from dominating the input feature space. For example, suppose dimension X ranged from 0 to 255 and dimension Y ranged from 0 to 1, then without normalization X would dominate Y in the input vector so that Y would effectively be negligible. Furthermore, the normalization improves the

discrimination between input cases. In the above example, without normalization each input case would typically be represented by a vector where X is two orders of magnitude greater than Y. Consequently, the input cases would appear more similar to each other than if each input dimension was normalized.

The mammogram features were linearly scaled to range from 0 to 1 by analyzing the mammogram set for each input dimension independently. For a single input dimension, the minimum and maximum values across the mammogram set were discovered and these were used to normalize each input case.

4.3. Mammogram feature selection

The 132 features extracted from each of the 450 mammograms were analyzed to produce a minimal set of features which best characterized the mammograms. The procedure for this was as follows:

(i) calculate the correlation between each pair of features across the set of mammograms,

(ii) identify the most correlated pair of features; features X and Y,

(iii) omit feature X if it has the least deviation across the set of mammograms, else omit feature Y,

(iv) repeat from step 2 whilst the highest correlation is above a prescribed threshold.

This procedure produced the correlations shown in Figure 7. It can be seen from the maximum correlations that many of the features were highly correlated. These correlations corresponded to the same type of textural features taken from the same mammogram regions, but where the features pertained to different textural orientations. For example, the entropies in the retroareolar region at 45 degrees and 135 degrees were highly correlated. The maximum correlation between nontextural features was 0.87.

The figure shows that the omission of highly correlated features tended to reduce the average correlation between features after an initial increase in this average. The discrimination between mammograms improves as the average correlation between the features is minimized. However, as the average correlation tends to continually decrease a correlation threshold must be set to terminate feature omission. This threshold was set by considering the distance between mammograms in feature space.

Section 4.1 explained that each feature was scaled to range from 0 to 1, hence the maximum distance between two mammograms in feature space was the square root of the number of features used. For example, the maximum distance for the original 132 features was 11.5. Therefore, for a given number of features, the distance between mammograms can be calculated and then normalized by the maximum potential distance.

Figure 8 displays the variation in the average normalized distance between mammograms as highly correlated features were omitted. Similarly, Figure 9 displays the variation in
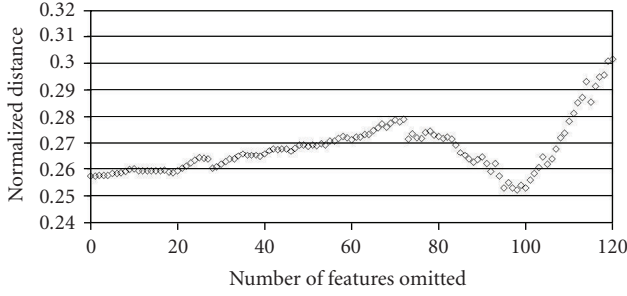
FIGURE 8: Average normalized distance between mammograms against the number of features omitted.
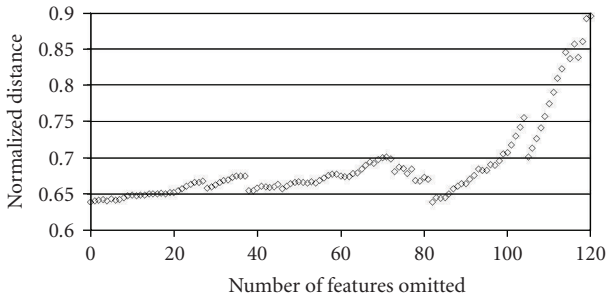

FIGURE 9: Maximum normalized distance between mammograms against the number of features omitted.

the maximum normalized distance between mammograms. The normalized distance between mammograms should be maximized to improve the discrimination between mammograms. The figures show that the normalized distances increase slightly as more features were omitted. However, excessively omitting features would restrict the information captured from the mammograms. Consequently, a correlation threshold of 0.98 was set to terminate feature omission. This caused 79 features to be omitted which approximately corresponds to local maxima in Figures 8 and 9.

In summary, mammogram features were omitted from the input vector in order to minimize the correlation between the remaining features, whilst maximizing the normalized distance between the mammograms in feature space. A correlation threshold of 0.98 was set to limit the highest correlation between mammogram features. This caused 79 features to be omitted and thus retained 53 features. SONNET's input field therefore consisted of 53 input neurons where each neuron represented a specific type of mammogram feature.

4.4. Classification performance measures

This section defines performance measures to compare different mammogram classifications. The measures in the first subsection are general to any classification task whereas those in the second subsection are specific to mammogram classification.

4.4.1. Distance in input feature space

A set of input cases can be conceived as a set of points in input feature space. Thus the performance of a classification scheme can be quantified by considering the distances between input cases in input feature space. These distances give rise to the following rule. *Input cases which receive the same classification should be proximate in feature space, whereas cases which are classified differently should be distant from each other.* Hence, the classification task becomes a multiobjective optimization problem which is required to minimize the average within-class distance between case-pairs, whilst maximizing the average between-class distance.

Performance measures can be formulated for the current task by considering two mammograms i and j which are a distance d_{ij} apart in input feature space. Suppose that these mammograms are classified as being of type χ_i and χ_j respectively, and that the corresponding classification confidences are c_i and c_j. It is more important for mammograms which are classified with a high confidence to be consistent with the above rule, than it is for mammograms classified with a lower confidence. Hence, the distance d_{ij} should be weighted by the confidences c_i and c_j.

The average distance over a set of mammograms can now be calculated. The average within-class distance, D_w, would be calculated over the set of mammograms which received the same classification (i.e., $\chi_i = \chi_j$), whereas the average between-class distance, D_b, would be calculated over the set of mammograms which received different classifications (i.e., $\chi_i \neq \chi_j$). These average distances are calculated as follows:

$$D_w = \frac{\sum_{\chi_i = \chi_j} c_i c_j d_{ij}}{\sum_{\chi_i = \chi_j} c_i c_j}, \quad (1)$$

$$D_b = \frac{\sum_{\chi_i \neq \chi_j} c_i c_j d_{ij}}{\sum_{\chi_i \neq \chi_j} c_i c_j}. \quad (2)$$

4.4.2. Patient-wise mammogram comparison

A patient should receive the same classification for their left and right CC mammograms, and this notion was confirmed by casual subjective observation. This notion can be tested by analyzing the distances between pairs of mammograms in feature space. Figure 10 shows the distances between mammogram-pairs in the reduced feature space of 53 feature types. Comparison between the left and right mammograms for the same patient produced the lower line, where each point corresponded to a single patient and the points were ranked according to increasing distance.

Each "diff D1" point was produced by comparing the right views between two different patients, and this comparison was repeated for all combinations of patient-pairs. Similarly, the "diff S1" points were produced by comparing pairs of left views of different patients. The points were again ranked according to increasing distance. There was no significance in comparing the left or right views individually and so the corresponding points overlap to appear as the upper line in Figure 10.
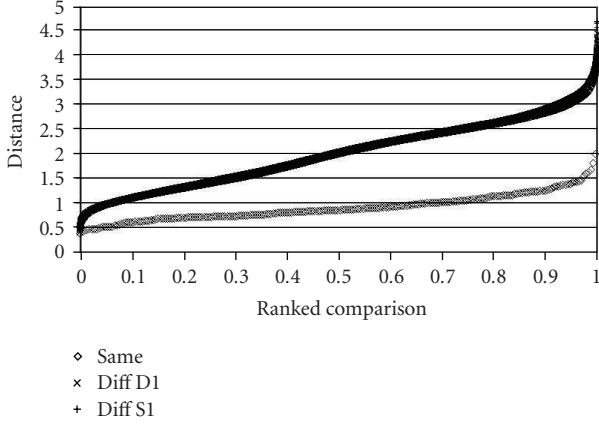
Figure 10: Comparing the distances between mammograms using 53 features: (a) *same*, the left and right views within patients, (b) diff D1, the right views between patients, and (c) diff S1, the left views between patients.
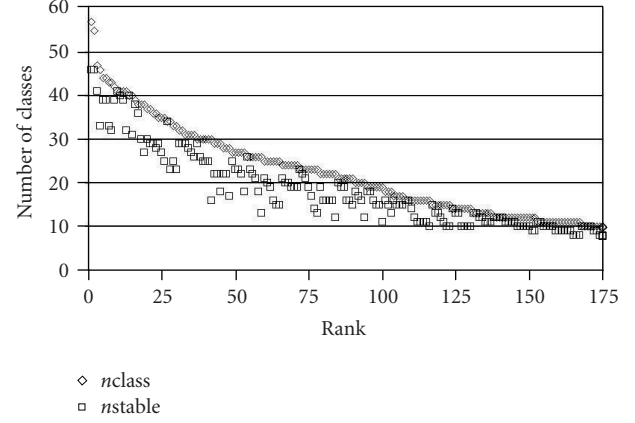

Figure 11: The number of classes formed on the best SONNET epochs. *nClass* is the total number of classes formed and *nStable* is the number of stable classes. The epochs are ranked according to descending *nClass* and then descending *nStable*.

The figure shows that within-patient distances were typically less than between-patient distances. Approximately 70% of the within-patient comparisons gave distances less than 1, and all these comparisons gave distances less than 2. Conversely, only 5% and 50% of the between-patient comparisons gave distances less than 1 and 2 respectively.

The results in this section justify the use of patient-wise performance measures for the mammogram classifiers. However, because the two mammograms for a particular patient can differ significantly, patient-wise performance measures should be used only as secondary measures. For example, patient-wise performance measures could be used to compare classifiers which are indistinguishable when using the measures based on distances in mammogram feature space. Note that patient-wise performance measures did not actively drive SONNET's development, but instead the measures were used to assess classification performance after development.

4.5. Discovering optimum classifications

Section 3 stated that multiple SONNET runs were conducted for 100 epochs. Any of these epochs could represent the optimum SONNET state, where many stable classifiers separate the mammogram set into clearly distinguishable classes. Every epoch was assessed according to various performance measures and this posed a multiobjective optimization problem. The number of candidate optimum SONNET epochs was reduced by discovering the *Pareto front* across the performance measures. Consequently, none of these candidate epochs could be dominated by another epoch on *every* performance measure. The Pareto front was discovered across the following dimensions:

(i) average within-class distance D_w (1),

(ii) average between-class distance D_b (2),

(iii) the number of classes encoded,

(iv) the classification confidences,

(v) the fraction of patients which received the same classification for their two mammograms.

D_w was minimized whereas all of the other performance measures were maximized.

5. MAMMOGRAM CLASSIFICATION PERFORMANCE

This section discusses the performance of SONNET in establishing a mammogram taxonomy. The optimum classifications from multiple SONNET runs were judged using the performance measures described in Section 4.5.

5.1. Number of mammogram classes

Casual observation of the mammogram set can roughly indicate the number of taxonomic classes involved but it is difficult to precisely specify the number of required classes. However, the current study focused on developing a maximal number of classes to discover the typical subtleties which discriminate mammogram classes. Various SONNET parameters control the number of classes encoded. These parameters were varied to analyze the number of classes which most commonly formed, and this number was deemed to correspond to the most natural taxonomic decomposition of the mammograms.

Figure 11 displays the number of classes formed on the best SONNET epochs. These epochs relate to many different SONNET runs but a single run could also produce multiple best epochs. The epochs were ranked according to the number of classes formed and the number of these which were *stable*. The resulting rank numbers are used to identify the best epochs in the subsequent discussion.

Classes became stable in SONNET after their encoding had been refined by sufficient past experience. Unstable classes were always present however, to enable SONNET to adapt to changes in the input patterns. Therefore, the proportion of SONNET's classes which were stable represents the maturity of the overall network and the quality of the
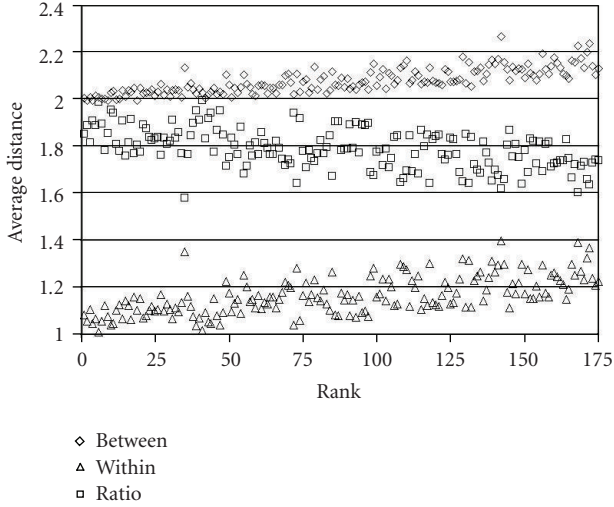
FIGURE 12: Average distance in input feature space for between-class mammograms and within-class mammograms for the best SONNET epochs. The ratio of between-class distance over within-class distance is also shown. The epoch ranking corresponds to Figure 11.

encodings. Hence, the best results in Figure 11 are those with the maximum number of *stable* classes.

Figure 11 shows that it was difficult for more than 40 stable classes to form. SONNET parameters were investigated to produce more classes but this resulted in SONNET memorizing individual mammograms instead of clustering them with other mammograms. SONNET commonly produced between 20 and 30 classes suggesting that this represents the most natural taxonomic breakdown. SONNET parameters could be set to form fewer, broader classes but these classifications obscure the subtleties which discriminate mammogram classes.

5.2. Class tightness

Figure 12 shows the average distance in input feature space for mammograms classified differently (between-class, D_b given by (2)) and for mammograms classified the same (within-class, D_w given by (1)). These distances were plotted for each of the best SONNET epochs which were ranked in Figure 11.

The SONNET epochs with low-rank numbers produced many narrow classes and thus yielded

  (i) a low average within-class distance because mammograms had to be highly proximate in feature space to be clustered together,

  (ii) a low average between-class distance because the class tightness allowed similar mammograms to be classified differently.

Conversely, SONNET epochs with high-rank numbers formed fewer, broader classes, and thus yielded higher within-class and between-class distances.

Reference distances were calculated to create a context within which to consider the between-class and within-class

distances. The average distance between all the mammograms, D_{av}, was 2.00 and the maximum distance, D_{max}, was 4.66. (These reference distances can be seen in Figure 10.) Figure 12 shows that the between-class distances for low-rank numbers approximately equalled D_{av} and that the maximum between-class distance was approximately half D_{max}.

A further reference distance can be calculated by considering a classification where each patient is distinct, such that their two mammograms are classified the same with a confidence of 1. This would yield $D_w = 0.90$ and $D_b \approx D_{av}$. Figure 12 shows that the within-class distances for low-rank numbers were slightly greater than 0.9, which was expected as each class clustered approximately 10 mammograms together.

The best classifications minimized the within-class distance yet maximized the between-class distance, therefore the ratio of between-class distance over within-class distance should be maximized. Figure 12 includes this ratio for each ranked epoch, and shows that the best epochs produced a ratio of almost 2 by developing relatively tight classes ($D_w \approx$ 1), whilst retaining typical between-class distances ($D_b \approx D_{av}$).

5.3. Patient-wise performance measures

Section 4.4.2 justified the use of patient-wise performance measures to quantify the extent to which the two mammograms for each patient received the same classification. This patient-wise performance measure was used as a secondary measure to discriminate classifiers which were similar when judged using other performance measures.

Figure 13 displays the fraction of patients whose mammograms were classified the same for the best SONNET epochs. This fraction was approximately 40% for the epochs with narrow classes (low-rank numbers), as these encodings captured the subtleties which differentiate the mammograms for a single patient. Conversely, the epochs with broad classes (high-rank numbers) classified approximately 75% of the patients as being the same for their two mammograms.

5.4. Mammogram taxonomy

The best result was deemed to be the SONNET epoch with rank number 6 in the previous discussion. This result produced 39 stable classes, where the first was formed on the 2nd epoch of the run and the last was formed on the 84th epoch.

This result produced relatively narrow classes with an average within-class distance of 1.01 whilst retaining a typical average between-class distance of 2.00. Consequently, this result yielded a relatively high ratio in Figure 12 of 1.98. Approximately 37% of the patients had their two mammograms classified the same.

The chief features that discriminated class encodings were two textural features, namely angular second moment and contrast. Figures 14 to 19 are examples of the mammogram classes that were encoded by SONNET.
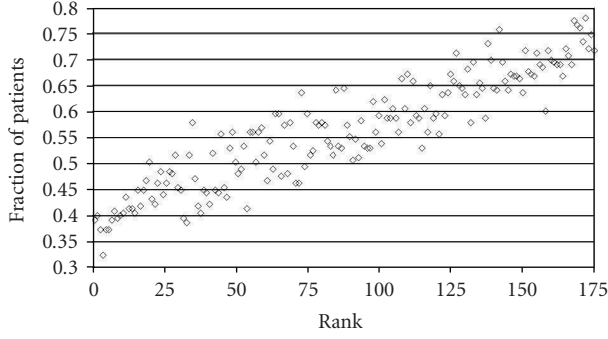
FIGURE 13: The fraction of patients whose two mammograms received the same classification for the best SONNET epochs. The epoch ranking corresponds to Figure 11.
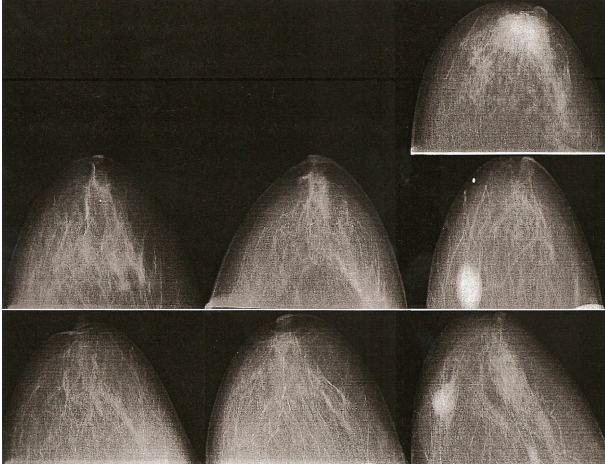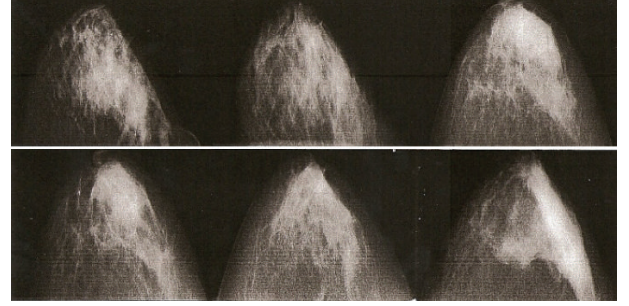

FIGURE 15: Example of a taxonomic class.
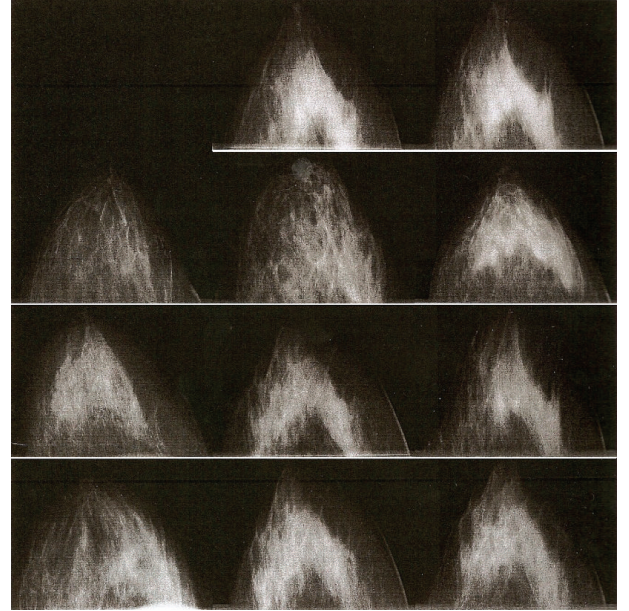

FIGURE 14: Example of a taxonomic class.

FIGURE 16: Example of a taxonomic class.

6. REFINING THE MAMMOGRAM TAXONOMY

6.1. Input feature selection

The main weakness with the current classification scheme is considered to be the manual specification of the image features which were extracted from the mammograms. The feature types and the regions from which they were extracted were designed to capture a priori knowledge about mammography, for instance, the importance of the retroareolar region. However, this manual specification necessarily requires arbitrary decisions, for instance, the quantitative position of the retroareolar region.

The image features and their corresponding region boundaries could be *automatically evolved* to produce an optimal input feature space. Two aspects of this are

(i) capturing *a priori* mammographic knowledge, for example, characteristic positions of lobular units, and

(ii) producing a high-quality feature space, for example, a minimal set of features with maximal orthogonality.

Other mammogram views could be used to extract input features in addition to the craniocaudal projection, for example, the mediolateral oblique view could be used.

6.2. Using control mammograms

Control mammograms could be exploited to refine the mammogram taxonomy. Control mammograms should be selected to represent clearly distinguishable mammogram classes. The classification scheme should initially be developed on these cases alone to shape classifier encodings. More ambiguous mammograms could then be introduced for subsequent classifier development. In order to achieve this, the classification scheme must be capable of *incrementual learning* and it must also address the so-called *stability-plasticity dilemma* [2]. SONNET satisfies these requirements.

6.3. Alternative classification schemes

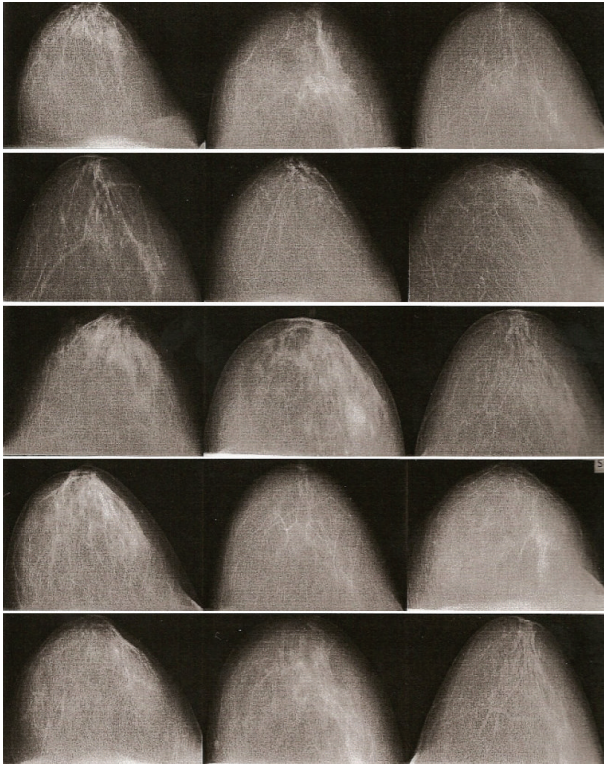A supervised classification scheme could be used to allow performance measures to actively drive classifier

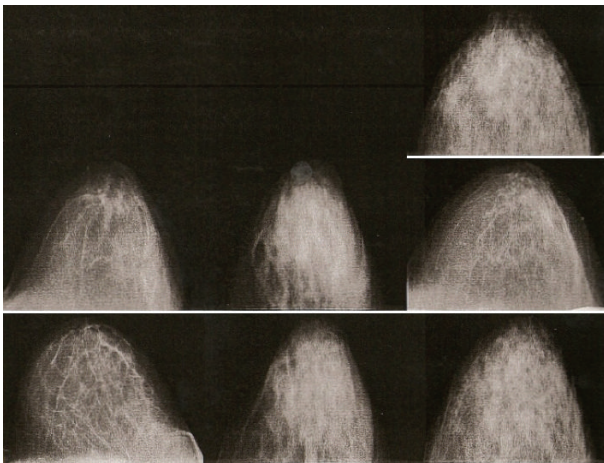FIGURE 17: Example of a taxonomic class.

FIGURE 18: Example of a taxonomic class.

FIGURE 19: Example of a taxonomic class.

development in a manner consistent with the passive discovery of optimal SONNET classifications, as outlined in Section 4.5.

An *evolutionary computing* (EC) technique could form an alternative classification scheme. EC is a flexible and adaptable technique and consequently it could combine a number of the processing stages detailed in the above protocol for producing a mammogram taxonomy. For example, EC could optimize its own subset of input feature types by processing raw image features directly.
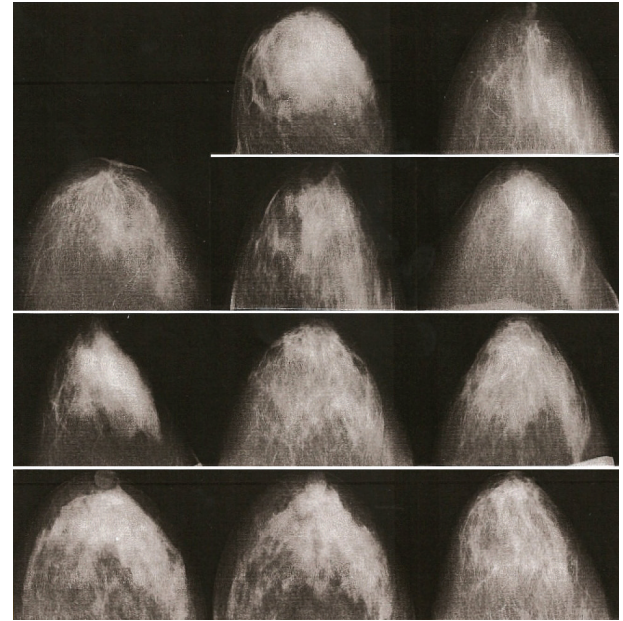
7. CONCLUSIONS

This study has developed a mammogram taxonomy by using an unsupervised classification scheme called SONNET. The encoded mammogram classes captured typical subtleties which discriminate mammograms. SONNET's controlling parameters were varied to govern the coarseness of the taxonomies. The developed classification scheme is considered to be a successful prototype but the scheme's efficacy is yet to be established.

The study shows promise for researching automated computational tools to assist with the detection of mammographic abnormalities. A mammogram taxonomy can be exploited to aid the detection of cancerous lesions via asymmetry identification [14], that is, by identifying anomalies between a patient's left and right mammograms. The evidence for cancerous lesions within the complex breast tissue can be very subtle, so mammogram features must capture localized information in a contextual manner, that is, *multiscale* features are required.

The authors have developed an evolutionary computation approach to discover multiscale features in imagery for a target detection application [15, 16]. This scheme used a *data crawler* which was evolved to gather evidence to discriminate target objects from nontarget objects. The crawler focused on low-level features in its immediate vicinity and processed these in the context of higher-level features collected over the crawler's trail.

As the data crawler has been developed for target detection in imagery, it is highly transferable to the problem of lesion detection in mammograms. The crawler could scrutinize mammogram areas which possess the greatest asymmetry and thus focus on candidate lesions. The evolutionary approach allows the crawler to discover its own multiscale features which best locate lesions.

The search for multiscale features over a diverse set of mammograms represents a very challenging problem, due to the high dimensionality of the potential search space. Hence, it is desirable to segregate the problem into multiple subproblems with less diversity. This can be achieved by exploiting the mammogram taxonomy as a preprocessing stage. This stage would classify a patient's mammograms, and thus would allow a data crawler to be evolved to specialize in only these taxonomic classes. Multiple crawlers could then be evolved, each of which specializes on its own subset of classes. Hence, the taxonomy would greatly constrain the search space in order to optimize asymmetry identification, and consequently, lesion detection.

ACKNOWLEDGMENT

REFERENCES

[1] S. W. Duffy, R. A. Smith, R. Gabe, L. Tabár, A. M. Yen, and T. H. Chen, "Screening for breast cancer," *Surgical Oncology Clinics of North America*, vol. 14, no. 4, pp. 671–697, 2005.

[2] G. Carpenter and S. Grossberg, "ART2: self-organization of stable category recognition codes for analog input patterns," *Applied Optics*, vol. 26, no. 23, pp. 4919–4930, 1987.

[3] L. Tabár, *Teaching Course of Mammography: Diagnosis and In-depth Differential Diagnosis of Breast Diseases*, Mammography Education, Cave Creek, Ariz, USA, 2006.

[4] D. Howard, S. C. Roberts, and L. Tabár, "Mammography taxonomy for improvement of lesion detection rates," in *Proceedings of the 6th International Workshop on Digital Mammography (IWDM '02)*, pp. 27–32, Bremen, Germany, June 2002.

[5] N. Jamal, K.-H. Ng, L.-M. Looi, et al., "Quantitative assessment of breast density from digitized mammograms into Tabar's patterns," *Physics in Medicine and Biology*, vol. 51, no. 22, pp. 5843–5857, 2006.

[6] S. Beer, *Brain of the Firm*, John Wiley & Sons, New York, NY, USA, 2nd edition, 1981.

[7] N. F. Boyd, C. Wolfson, M. Moskowitz, et al., "Observer variation in the classification of mammographic parenchymal patterns," *Journal of Chronic Diseases*, vol. 39, no. 6, pp. 465–472, 1986.

[8] A. Nigrin, *Neural Networks for Pattern Recognition*, Bradford Books, MIT Press, Cambridge, Mass, USA, 1993.

[9] D. O. Hebb, *The Organization of Behavior*, John Wiley & Sons, New York, NY, USA, 1949.

[10] O. Paulsen and T. J. Sejnowski, "Natural patterns of activity and long-term synaptic plasticity," *Current Opinion in Neurobiology*, vol. 10, no. 2, pp. 172–179, 2000.

[11] S. Harford, "Automatic segmentation, learning and retrieval of melodies using a self-organizing neural network," in *Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR '03)*, Baltimore, Md, USA, October 2003.

[12] J. A. Marshall and V. S. Gupta, "Generalization and exclusive allocation of creditin unsupervised category learning," *Network: Computation in Neural Systems*, vol. 9, no. 2, pp. 279–302, 1998.

[13] M. W. Spratling and M. H. Johnson, "Neural coding strategies and mechanisms of competition," *Cognitive Systems Research*, vol. 5, no. 2, pp. 93–117, 2004.

[14] D. Scutt, G. A. Lancaster, and J. T. Manning, "Breast asymmetry and predisposition to breast cancer," *Breast Cancer Research*, vol. 8, no. 2, article R14, 2006.

[15] D. Howard, S. C. Roberts, and C. Ryan, "Machine vision: exploring context with genetic programming," in *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO '02)*, pp. 756–763, Morgan Kaufmann, New York, NY, USA, July 2002.

[16] D. Howard, S. C. Roberts, and C. Ryan, "Pragmatic genetic programming strategy for the problem of vehicle detection in airborne reconnaissance," *Pattern Recognition Letters*, vol. 27, no. 11, pp. 1275–1288, 2006.