Significance Analysis of Qualitative Mammographic Features, Using Linear Classifiers, Neural Networks and Support Vector Machines

Michael Mavroforakis[1], Harris Georgiou[1], Nikos Dimitropoulos[2], Dionisis Cavouras[3], Sergios Theodoridis[1]

[1] Informatics & Telecommunications Dept., Univ. of Athens, TYPA buildings, Univ. Campus, 15771, Athens, Greece.

[2] Medical Imaging Department, EUROMEDICA Medical Center, 2 Mesogeion Avenue, Athens, Greece.

[3] Medical Image and Signal Processing Lab., Dept. of Medical Instruments Technology, STEF, Technological Educational Institution of Athens, Ag. Spyridonos Street, Egaleo, 12210, Athens, Greece.

Michael Mavroforakis, 43 Knossou str, Glyfada 16561, Athens, Greece.

Tel: +30 210 9648663, Email: mmavrof@di.uoa.gr

Significance Analysis of Qualitative Mammographic Features, Using Linear Classifiers, Neural Networks and Support Vector Machines

Abstract

Advances in modern technologies and computers have enabled digital image processing to become a vital tool in conventional clinical practice, including mammography. However, the core problem of the clinical evaluation of mammographic tumors remains a highly demanding cognitive task. In order for these automated diagnostic systems to perform in levels of sensitivity and specificity similar to that of human experts, it is essential that a robust framework on problem-specific design parameters is formulated. This study is focused on identifying a robust set of clinical features that can be used as the base for designing the input of any computer-aided diagnosis system for automatic mammographic tumor evaluation. A thorough list of clinical features was constructed and the diagnostic value of each feature was verified against current clinical practices by an expert physician. These features were directly or indirectly related to the overall morphological properties of the mammographic tumor or the texture of the fine-scale tissue structures as they appear in the digitized image, while others contained external clinical data of outmost importance, like the patient's age. The entire feature set was used as an annotation list for describing the clinical properties of mammographic tumor cases in a quantitative way, such that subsequent objective analyses were possible. For the purposes of this study, a mammographic image database was created, with complete clinical evaluation descriptions and positive histological verification for each case. All tumors contained in the database were characterized according to the identified clinical features' set and the resulting dataset was used as input for discrimination and diagnostic value analysis for each one of these features. Specifically, several standard methodologies of statistical significance analysis were employed to create feature rankings according to their discriminating power. Moreover, three different classification models, namely linear classifiers, neural networks and support vector machines, were employed to investigate the true efficiency of each one of them, as well as the overall complexity of the diagnostic task of mammographic tumor characterization. Both the statistical and the classification results have proven the explicit correlation of all the selected features with the final diagnosis, qualifying them as an adequate input base for any type of similar automated diagnosis system. The underlying complexity of the diagnostic task has justified the high value of sophisticated pattern recognition architectures.

Keywords

Mammography; Tumor Characterization; Automated Diagnosis; SVM Classifiers.

1. Introduction

Breast cancer is the most common cancer type and the second most common death cause in women in civilized world. Screening mammography, for detecting early breast cancer in asymptomatic women, increases the likelihood for cure and long-term survival. However, in cases of indeterminate mammographic findings, breast biopsy may be required. Avoiding unnecessary biopsies is important due to the discomfort, cost and probable breast scars inflicted upon the patients, which may cause diagnostic difficulty in future mammographic examinations.

The diagnostic and clinical evaluation of mammographic images constitutes a difficult and complex cognitive task, which requires advanced levels of expertise and knowledge by the trained physicians. Mammographic screening, for the identification of abnormalities and the pathological characterization of breast tissue, is a visual task that combines several aspects and x-ray findings, presented in various areas of the mammographic image, as well as external data available through each patient's clinical history. Specific clinical findings, such as the morphological properties and fine-scale structural information of the underlying tissue, are the key factors in characterizing the severity of every mammographic tumor, i.e., its benign or malignant nature [01]. Modern computer technology can be used to implement automatic image processing and analysis of various aspects of these findings, thus supporting effectively the expert's evaluation as a valuable suggestive tool. However, the exact task of tissue characterization and classification of a tumor, as probable benign or probable malignant, is extremely complex and includes advanced inference mechanisms [02-03]. Consequently, computer-aided diagnosis (CAD) systems focus on specific aspects of the diagnostic process, such as the identification and analysis of microcalcifications or the detection of irregular tissue structures, each suggestive of specific abnormalities [04]. Therefore, it is essential that these morphological and textural properties are defined in detail as a specific list of qualitative features that can be formulated into a robust set of corresponding quantitative measurements.

This study focuses on three core issues: (a) to investigate efficient mammographic features, already used in current clinical practice for the pathological evaluation of mammographic tumors, (b) to assess their diagnostic value with objective statistical and classification methods, and (c) to formulate a robust quantitative model for using them as the input for any automated image analysis method. Specifically, a complete and coherent set of clinical features was constructed, by exploiting significant pathological

factors, related to mammographic abnormalities and directly or indirectly suggestive of probable malignant cases of breast tumors. The information content of these features is related to the mammographic image itself, namely the morphological and textural properties of the tumor's area, or external data obtained by the patient's clinical history. Subsequently, this set of qualitative descriptive estimations, supplied by the expert physician's subjective evaluation, was quantified and translated into a robust dataset. This dataset was used in statistical and classification analysis schemes, employing a wide range of discrimination evaluations, ranging from standard significance tests, to advanced pattern recognition architectures. The results obtained during this analysis can be used for objective comparative studies, as well as to produce ranking lists of clinical features in accordance to their true diagnostic value. These features and their relative discriminative power constitute the input specifications and guidelines, which are essential for any type of automated tumor diagnosis system that is based on morphological, textural or descriptive datasets.

2. Materials and Methods

The current study was based on four distinct issues: (1) Create a thorough list of abnormal findings regarding diagnostic evaluation of a mammogram, especially related to image textural and morphological features of the underlying tissue. From this list, the most prominent and content-rich features were to be selected, according to their suitability for automatic extraction through image processing algorithms. (2) Create a specialized mammographic image database, containing clearly identifiable and histologically verified cases of benign or malignant tumors. All cases were evaluated and annotated in relation to the previously defined list of important clinical features. (3) Analyze the newly constructed set of mammographic images in relation to the feature list, focusing especially on investigating the importance, comprehensiveness and consistency of each one of these features when correlated with the verified final diagnosis. (4) Investigate the performance of individual features, as well as subsets of combined features, when used as real training datasets for various classifier architectures, i.e., linear classifiers, neural networks (NN) and support vector machines (SVM).

2.1 Mammographic Features List

The first phase of the study included an extensive research through various aspects of identifying and evaluating numerous radiologic findings in mammographic images, related to benign and malignant abnormalities. The investigation was conducted by enumerating and documenting all the morphological and textural tissue characteristics, which are recognized and evaluated by the experts when they conduct a clinical diagnosis [05-07]. Furthermore, an additional list of other important features, like patient's age and clinical background, were also included in this list. Some of the features, like the presence of suspicious masses or microcalcifications, are normally related directly to abnormalities, while others, like the exact location and size of the mass, are usually evaluated as intermediate suggestive indications of benignancy or malignancy [02-03].

The complete list of the 31 features, along with direct indications of benign and malignant biopsy results, is summarized in Table 1. The features were grouped in categories according to the general type of abnormality they refer to. The "CPU" column refers to the capability of relating the corresponding features to image processing algorithms, which can automatically extract specific content-related information. Advanced algorithms for automated mammographic lesion detection have been proposed, however their level of sensitivity and specificity, as well as their fine-scale accordance to the corresponding expert's detailed description of tumor boundaries, is still under investigation [04, 08].

It is obvious that some of the above features, although very important, are not directly related to the mammographic image by itself and, thus, they have to be provided as external annotation data for each case by the physician [05-06]. Furthermore, not all of them are related to the clinical characterization of tumors, which is the main concern of the current study. Subsequently, a robust, content-rich subset of features was constructed, using selected features that are highly related to tumor benignancy or malignancy and, at the same, time refer to textural and morphological characteristics of the tumor, i.e. to objective image properties. These feature selections were also based on the general requirement that the features can be automatically extracted and processed. In this case, the features that are extracted from the image can be linked directly or indirectly to morphological or textural properties of the tissue inside and around the tumor area, as it appears on the image itself [09-12]. Qualitative or descriptive features were scaled in numerical ranges or percentages, in order to acquire quantitative data values.

Both the final feature list selections, as well as the exact quantification scales, were defined in cooperation with an expert physician in order to ensure complete and detailed clinical results.

The final set of 9 clinical features was the base for the annotation list, which was used to describe and document the expert's clinical evaluation for each mammographic image in the database. Specifically, (1) the presence of tumors, (2) the presence of microcalcifications, (3) the tumor density, (4) the percentage of fat within the tumor, (5) the tumor boundary vagueness, (6) the tumor homogeneity, (7) the tumor morphological shape type, (8) the patient's age, as well as (9) the final histologic diagnosis, were included. As the patient's age remains a feature of high clinical importance, it was also included in the final annotation list as a unique "external" data, although it cannot be referred directly from the mammographic image itself [05-06]. Finally, the morphological shape type refers to the classification of the tumor's shape in one of four predefined shape categories, related to tumor's boundary roughness and stellate or lobulated outline. These four categories, illustrated in Figure 1, are defined as round, lobulated, micro-lobulated and stellate, and their ranking is directly related to their pathology, from benign to malignant, respectively [13-16].

These quantified properties are essentially explicit information related to specific types of malignant mammogram abnormalities, including architectural distortion, clusters of microcalcifications, lobulated or stellate masses, as well as skin oedema [07]. Thus, the initial annotation list constitutes a complete dataset that provides significant diagnostic data, which has been used for further statistical and clustering analysis. The final annotation list, containing the selected features and quantification scales, is presented in Table 2.

2.2 Mammographic Image Database

The second phase of the study included the creation of a thorough mammographic image database, especially designed to focus on cases of tumor presence with histological verification as benign or malignant by an expert physician. The requirement for patient's clinical history and positive histological verification of the benignancy or malignancy of each case was assessed as one of extreme importance for the quality and validity of the subsequent results. Thus a new, special-purpose image set was assembled, using cases of mammographic tumors with complete radiologic evaluation and histologic diagnosis. The initial set contained several hundreds standard mammograms and it was used

as a base for the final selection of tumor cases with positive clinical verification by surgical biopsy and histologic examination. The selected subset was constructed in accordance to the general requirement for complete and unbiased statistical distributions over all the radiologic findings investigated in the study.

The selected mammogram films were digitized at a typical resolution of 63 μm (400 dpi) with 8-bit graylevel depth, in order to retain fine scale textural and structural tissue characteristics. Furthermore, some additional post-processing was applied uniformly over all the selected images, using optimized unsharp filtering for image enhancement with minimal spectrum alteration. The resulting images were evaluated and verified by the expert as acceptable in terms of image quality and resolution. The final set of 130 images of histologically confirmed lesions (46 benign and 84 malignant) was used as the base for all the subsequent analysis presented in this study, with no reduction in spatial resolution or graylevel depth.

Subsequently, every mammographic image in the database was evaluated by two expert radiologists and all the important findings were recorded separately for each case, using the annotation list that was created during the previous phase of the study. As the tumor's shape is one of the most important morphological properties for clinical characterization, it was essential that shape type information was provided by the expert and registered into the annotation list. In addition, for further morphological and textural analysis capabilities at some later stage, it was crucial that every tumor was clearly described and registered by defining its boundary outline. In order to obtain tumor boundaries of high quality and detail, a manual segmentation was applied. Specifically, each tumor was manually described by the radiologists using a high-resolution digitizer device and stored as an embedded boundary descriptor via alpha channel data. These boundary descriptions were used for further independent work on the definition of mass inclusion masks and boundary zones for textural features extraction at these areas of interest.

2.3 Statistical Analysis

Statistical analysis was conducted on the data obtained through the annotation list in three groups: benign cases, malignant cases and all cases. For each of these groups, normal distribution approximation parameters, i.e., mean value and variance, were calculated and the results were

investigated under statistical significance analysis and projected aliasing errors [17]. Specifically, for mean values calculated separately for each case grouping, significance ranges were estimated according to the current group size and variance. Subsequently, feature distributions for each group were approximated by normal distributions and the statistical error was calculated according to the optimal limiting value separating two different classes, i.e., benign and malignant cases, using each individual feature. In other words, for every individual feature contained in the annotation list, an optimal decision threshold was calculated and the corresponding classification errors for the benign and malignant cases were used as an indicative measurement of statistical aliasing. As an example, the use of tumor boundary sharpness, as the sole discrimination feature for bimodal normal distribution modeling, is illustrated in Figure 2.

The features were also processed through Univariate significance analysis, specifically T-test [18]. Multivariate significance analysis was also applied, using the Multivariate Analysis of Variance (MANOVA) method [19]. In both cases, every feature was investigated separately under statistical dependence hypotheses in relation to the diagnosis and the results formulated a quantitative ranking, regarding the correlation between each feature and the diagnosis.

2.4 Classification Analysis

In order to assess the discriminative power of each one of the qualitative clinical measurements, several classification schemes were applied against the verified diagnosis for each case. Pattern recognition techniques include various types of decision-theoretic approaches for data analysis and classification, and have been proven extremely valuable to real problems of high complexity such as the task of mammographic diagnosis. In this study, several of the standard linear and non-linear classifiers were used in order to evaluate both the true performance of these features, as well as the overall complexity of the problem itself [18].

Both linear and non-linear classification architectures were employed for every dataset configuration. Specifically, optimization of the best feature set was investigated by applying exhaustive search through all the combinations of features, in order to identify the ones that yield maximum discrimination capability and optimal performance. Furthermore, the performance of each classifier

was evaluated through extensive use of k-fold cross validation techniques, specifically leave-one-out and leave-k-out methods [20].

For linear classification testing, three standard models were considered. Linear Discrimination Analysis (LDA) was applied in the form of classifier, using iterative subsets of the initial training set and employing leave-one-out classification for every individual pattern in the set [21]. A Minimum Distance Classifier (MDC) with Mahalanobis distance function was employed in combination with Least-Squares data transformation for better statistical compactness, yielding the Least-Squares Minimum Distance Classifier (LSMD) that was used in this study [18].

Additionally, a typical Nearest Neighbor classifier with variable neighborhood size (K-NN) was employed, using the neighborhood size K as an optimization parameter [18].

From the various types of typical non-linear classifiers, two representative neural architectures were considered. A Multi-Layered Perceptron (MLP) neural network model was used, using the back-propagation algorithm for training, employing topology optimization and various choices for the neuron activation functions, specifically softmax, hyperbolic tan and hard limiter [22]. Similarly to the MLP, a Radial Basis Function (RBF) neural network architecture was also employed as a kernel-based alternative, using Gaussian activation functions and optimized topology [23]. For neural network classifiers, no feature reduction was necessary, as the neurons of the (trained) input layer could be examined in order to discard features that correspond to input weights with zero or near-zero values [24]. In other words, the architecture of the neural networks favors the automatic ranking of the inputs during the training phase, in a way that the final classifier can be examined in order to identify significant and non-significant features.

For more advanced investigation of the feature set, typical Support Vector Machine (SVM) models were applied in relation to the final diagnosis. Specifically, the C-SVC model was used in combination with standard RBF kernel functions, optimizing the penalty factor (C) and the Gaussian spread parameter (σ) during training [25]. SVM classifiers employed limited feature set optimizations, using iterative runs of enlarging inclusions of several features, available on the feature ranking lists created by MANOVA significance analysis. Due to the statistical importance of the shape type feature, all classifications were considered both with and without the inclusion of this specific feature.

3. Results

Preliminary analysis on the initial dataset has confirmed the strong statistical correlation between morphological shape type and verified diagnosis of breast tumors in the mammograms. Specifically, the first two types of morphology, round and lobulated tumors, exhibited 17% and 5% of malignancy, respectively, within the same class. On the contrary, micro-lobulated and stellate types exhibited 95% and 97% of malignancy, respectively, as illustrated in Table 3. When combining the round and lobulated cases, the overall percentage of malignancy was 12%, while for combined micro-lobulated and stellate cases, the overall percentage of malignancy was 96%, as illustrated in Table 4. This high statistical dependency of specific morphological features of each tumor with its verified pathology confirms the clinical value of its shape when conducting a pathological evaluation of a mammogram. It should be noted that if the shape type feature were to be used as the sole input for predicting the final diagnosis, an accuracy rate just over 93% could be achieved.

3.1 Statistical significance analysis

Global statistics of the dataset were calculated for every individual feature in relation to the final diagnosis. The properties and differences in the resulting bimodal normal distributions, for benign and malignant cases, revealed the discriminating power of each individual feature, as well as the corresponding significance ranges for the mean values. The mean value and standard deviation of each feature were used for constructing a bimodal normal distribution statistical model, while the corresponding aliasing error between the two kernels was used to evaluate the statistical separability of benign and malignant cases, according to this feature. Figure 2 presents a graphical display of using boundary sharpness for this type of statistical modeling. Table 5 summarizes the complete results of these tests, while Table 6 summarizes all types of statistical discrimination modeling, including T-test, F-test and bimodal normal distribution modeling. Although some tests were not applicable to specific features due to statistical limitations (e.g., zero variance or non-separable Gaussian distributions), early conclusions on the discrimination power of each individual feature could already be drawn from these early tests.

Furthermore, in order to produce feature rankings, which take into account statistical dependencies between the individual features, MANOVA was applied, investigating the discriminating power of each feature against the final diagnosis, as well as its independency to all the other features.

Table 7 summarizes the feature rankings for all statistical significance analysis methods applied in this study. The results obtained by bimodal normal distribution modeling, Univariate and MANOVA were generally similar and consistent, producing feature rankings with little differences in the exact ordering with regard to the importance of each feature.

3.2 Classification results

Classifications results were used as guidelines for evaluating the performance of individual features, as well as identifying optimal feature combinations. Classification accuracy rates were thoroughly investigated for all classifier models and comparative results were obtained against the final diagnosis.

3.2.1 Individual features evaluation

Datasets of single feature inclusions were constructed for conducting discriminating power analysis against diagnosis, using a typical LSMD classifier. The complete results for individual feature classification configurations are summarized in Table 8. Similarly to the results already obtained by statistical significance analysis, the morphological shape type of the tumor proved to be the most correlated feature with regard to final diagnosis. When patient's age and shape type features were excluded, optimal feature combinations, which were selected by the classifier, included tumor's boundary sharpness, fat inclusion percentage and tumor homogeneity, yielding a maximum accuracy rate of 86,9%. The introduction of patient's age into the set of input features also produced optimal configurations, achieving 89,2% accuracy. Classification results analysis showed that for feature combinations without shape type inclusion, the accuracy ranged from 87,7% up to 91,5%, while for feature combinations that included shape type information, the accuracy ranged from 91,5% up to 93,1%, essentially verifying the explicit discriminating value of this specific feature.

3.2.3 Comparative classifier performance

For a more realistic performance analysis for optimized feature combinations, a wide range of linear and non-linear classifiers were used. Linear classifiers included exhaustive search through all feature

subsets for identifying optimal feature combinations, while NN and SVM classifiers used full feature sets or optimal feature subsets, already available through the linear classifiers. Table 9 summarizes the highest accuracy rates achieved by each classifier, with and without the inclusion of shape type information.

The LDA classifier exhibited an accuracy rate of 87,69% when using an optimally selected feature set of patent's age and tumor's boundary sharpness. When the shape type feature was included in the input, the optimal feature set was constituted of the tumor's boundary sharpness and shape type features, and the accuracy rates were raised up to 93,85%. It should be noted that this particular accuracy rate was marginally higher than the statistical dependency of tumor's shape type versus its final diagnosis, namely 93,08%, thus proving the importance of combining several features to construct optimal feature subsets.

Similarly to the LDA, the optimized K-NN classifier achieved an accuracy rate of 91,54% when using all the available features except shape type and 93,08% when including shape type information. In fact, the optimal feature subset in the second case was constituted only by the shape type feature itself, essentially implementing the statistical classifier of grouping round and lobulated cases as probable benign, and micro-lobulated and stellate cases as probable malignant. There was no clear indication regarding the overall optimal value for the neighborhood size, although most configurations of high accuracy employed sizes of 3 to 8 neighboring samples.

For the LSMD classifier, success rates were similar to the K-NN. Specifically, an accuracy of 89,23% was achieved when using all the available features except shape type information and 93,08% when using the shape type feature as well. As in the case of K-NN, the shape type feature dominated the optimal feature subset and the resulting classification scheme is analogous to the suggestive statistical grouping of round and lobulated cases as probable benign, and micro-lobulated and stellate cases as probable malignant.

Both the MLP classifier and the RBF neural classifiers employed full feature sets and optimized size for the hidden layer. The MLP classifier yielded an accuracy of 91,54% when no shape type information was available, using 1 or 4 hidden units and linear activation function. The inclusion of the shape type feature did not affect the overall accuracy rate, although it resulted in many more

configurations achieving this maximum efficiency. Similarly, the RBF classifier achieved an accuracy of 90,77% without shape type feature inclusion (using 8 hidden units) and 91,54% with shape type feature inclusion (using 5 hidden units).

The SVM classifier, employed in this study as a representative candidate of this family, was the C-SVC model (penalty-driven SVM classifier) with radial basis kernel function (RBF). For feature lists with shape type information excluded, the accuracy rates achieved by the SVM classifier were 93,85% even when using only the first 4 features from the ranking list, namely tumor's boundary sharpness, patent's age, percentage of fat inclusion and tumor homogeneity. For feature lists including the shape type property, the accuracy rates achieved by the SVM classifier were 94,62%, conclusively higher than the statistical correlation between shape type and diagnosis. As expected, the shape type feature was included in all optimal feature subsets achieving this level of performance. Although no clear conclusions can be drawn regarding the exact choices on the values of SVM parameters C and σ, analysis of the various SVM configurations have shown that the values of the penalty factor C, namely between 1 and 10, were inversely proportional to the corresponding values of function spread parameter σ, namely from 0,1 down to 0,01.

4. Discussion

Results from statistical significance analysis reveal the high correlation between most of the qualitative features in the annotation list and the final diagnosis. Morphological shape type of the tumor's outline exhibits the highest dependency in relation to the diagnosis, yielding the specific feature as adequate to provide discrimination capability for correctly classifying benign and malignant cases with success rates up to 93%. Several other features, such as tumor boundary sharpness, tumor homogeneity, as well as patient's age, have proven as important clinical aspects of the expert's evaluation. All the features contained in the annotation list exhibited some degree of correlation to the diagnosis, thus qualifying them as plausible for automatic diagnosis systems, although in most cases optimal combinations of several features have to be used, instead of single features.

The discriminating value of each individual feature was confirmed by several statistical significance properties, including T-test, F-test, MANOVA, bimodal normal distribution error estimation and mean

value significance range, as well as real classification runs using a typical LSMD classification scheme. Although the feature ranking lists, created using the results of these analysis methods, differ slightly on the exact ordering of the features, some basic conclusions could be drawn with regard to the overall quality of each individual feature. Specifically, the morphological shape type of the tumor, namely its classification in one of the round, lobulated, micro-lobulated or stellate categories, has been established as the most important feature. Round and lobulated cases have been proven highly correlated to benignancy, while micro-lobulated and stellate cases have shown high correlation to malignancy, as Table 3 and Table 4 show. For combined features configurations, optimized sets containing the shape type feature exhibited 3% to 4% higher success rates than the ones without it, when used in real classification schemes of both linear and non-linear models. Subsequently, tumor boundary sharpness or fuzziness was clearly suggestive to benign or malignant tumors respectively, as Table 8 shows. Another important feature was the overall density of the tissue that constitutes the tumor, as dense tissue samples of abnormal physiology are typically related to malignant growth rate of the cells in those areas [02, 07]. Tumor's homogeneity and percentage of fat inclusion were also important when combined together or with some other feature of high discriminative quality, although none of them could provide high success rates when used individually. Patient's age has been used as a good representative of features provided as "external" annotative information by the physician, as it is not directly related to the informative content of the mammographic image itself, but rather on clinical data from patient's history. However, it proved to be a very important aspect of the overall clinical evaluation of each case. Finally, the indicative feature of microcalcifications' presence was included in several optimal feature combinations, although it shown minimal discriminative value when used individually to predict benignancy or malignancy.

All features, except patient's age and tumor's shape type, refer directly or indirectly to textural properties of the tumor area as it appears in the mammographic image. However, the shape type was evaluated as the most important feature contained in the annotation list. This means that, in order to have a fully automated diagnosis system that is based on objective feature measurements of various textural properties, several of these features have to be optimally combined. In any case, both morphological and textural features can be formulated into a well-defined set of extraction functions, capable of implementing objective estimators of various morphological and textural properties of each tumor, as it appears in the mammographic image. This would be very helpful to the estimation of the

correct radiologic diagnosis as an adjunct tool to the physician and the in future a complementary system combined with a CAD.

With regard to best classifier performance, the efficiency of non-linear architectures over linear equivalents was proven in almost all cases. Regarding the prediction of the final diagnosis, the performance of all classifiers was evaluated separately when including or excluding the shape type, which was the most dominant feature in the set. For feature sets without any shape type information inclusion, optimized K-NN classifier achieved the best results over LDA and LSMD alternatives. Similar feature sets containing the shape type feature produced results with no significant preference towards any of these three classifiers. These results were closely matched or exceeded by several MLP and RBF NN configurations, especially in the case of excluding shape type information. Although the best accuracy rates in some cases were produced by linear, instead of non-linear, classifiers, it should be noted that NN classifiers used only complete feature sets or feature combinations already calculated as optimal for linear equivalents. Both MLP and RBF models required a larger number of hidden units when shape type information was excluded, while the inclusion of the specific feature essentially simplified the discrimination process and thus concluded in topologies with lesser hidden units. The overall performance of MLP architectures was marginally higher than RBF equivalents, employing much smaller hidden layers and greater degrees of generalization.

The SVM classification schemes yielded overall maximum accuracy rates, both when the shape type feature was excluded or included in the input vector, higher than the corresponding maximum rates of any other linear or NN alternative investigated in this study. Thus, a representative application of advanced SVM models, compared to several linear and NN classification schemes, is suggestive to their superiority in classification problems that exhibit high degree of non-linearity in the training datasets.

5. Conclusion

The problem of identifying image features characterizing the overall morphology and fine-scaled structural properties of the tissue in mammographic tumors, as well as external clinical data, was investigated using objective statistical analysis and pattern recognition approaches. Therefore, although

the initial descriptive data were qualitative in nature, the translation into quantitative values and their thorough processing via advanced pattern analysis algorithms, produced objective evaluations and discriminating power estimations of their true efficiency.

All the selected features have shown some degree of dependency to the final diagnosis, while some of them, such as morphological shape type, provided discriminating levels high enough to be used even individually for tumor classification schemes. Optimal feature sets, employed in advanced non-linear classification architectures, like SVM classifiers, provided accuracy rates up to almost 95%, thus proving their efficiency and making such systems plausible for clinical application. All features investigated in this study, except patient's age, are related to morphological and textural properties on the mammographic image itself, therefore a completely automated diagnosis system, using the same content-rich descriptive features, is feasible.

Statistical and classification analysis results have shown that, although the selected feature sets were in fact content-rich with regard to their diagnostic value, the diagnostic process itself remains a complex and demanding task. The high degree of non-linearity employed in the discrimination of the input data with regard to diagnosis prediction suggests that automatic diagnosis systems should implement powerful pattern recognition models of non-linear and highly adaptive architecture. Future work should be focused on designing specialized image processing algorithms for efficient automatic extraction of morphological and textural features, combined with robust implementations of advanced classification architectures, such as SVMs. Automated diagnosis of breast mammographic abnormalities, combined with CAD systems, which indicate suspicious lesions in mammograms, will be a very powerful tool in the hands of the mammographic departments and the reporting physicians, especially the less experienced ones.

References

[01] Bocchi L, Coppini G, et al., Tissue characterization from X-ray images. Med Eng Phys 1997; 19(4): 336-42.

[02] Eagan R L, Breast imaging: Diagnosis and morphology of breast diseases. Philadelphia: Saunders, 1988.

[03] Robbins S L, Angell M, Kumar V, Basic Pathology. Philadelphia: Saunders, 1981.

[04] Christoyianni I, Dermatas E, Kokkinakis G, Fast detection of masses in computer-aided mammography. IEEE Sig Proc Mag 2000; Jan: 54-64.

[05] Orsi C D, Getty D J, Swets J A, et al., Reading and decision aids for improved accuracy and standardization of mammographic diagnosis. Radiology 1992; 184: 619-22.

[06] Wu Y, Giger M L, Doi K, et al., Artificial neural networks in mammography: Application to decision making in the diagnosis of breast cancer. Radiology 1993; 187: 81-7.

[07] Ackerman L V, Mucciardi A N, et al., Classification of benign and malignant breast tumors on the basis of 36 radiographic properties. Cancer 1973; 31: 342-52.

[08] Huai Li, Yue Wang, Ray Liu K J, et al., Computerized radiographic mass detection – Part I: Lesion site selection by morphological enhancement and contextual segmentation. IEEE Trans Med Im 2001; 20(4): 289-301.

[09] Haralick R M, Shanmugam K, Dinstein I, Textural features for image classification. IEEE Trans Sys Man Cyb 1973; SMC-3(3): 610-21.

[10] Haralick R M, Statistical and structural approaches to texture. Proc IEEE 1979; 67(5): 786-804.

[11] Galloway M, Texture analysis using gray level run lengths. Comp Graph Im Proc 1975; 4: 172-9.

[12] Mavroforakis M, Georgiou H, Cavouras D, Dimitropoulos N, Theodoridis S, Mammographic Mass Classification Using Textural Features and Descriptive Diagnostic Data, 14th International Conference on Digital Signal Processing (DSP-2002), July 1-3, 2002, Santorini, Greece.

[13] Bruce L M, Adhami R R, Classifying mammographic mass shapes using the wavelet

transform modulus-maxima method. IEEE Trans Med Im 1999; 18(12): 1170-7.

[14] Georgiou H, Cavouras D, Dimitropoulos N, Theodoridis S, Mammographic mass shape characterization using neural networks, 2nd European Symp. on Biom.Eng. and Med. Phys., Patras, Greece, 2000, BME-13.

[15] Kilday J, Palmieri F, Fox M D, Classifying mammographic lesions using computerized image analysis, IEEE Trans Med Im 1993; 12(4): 664-9.

[16] Georgiou H, Mavroforakis M, Cavouras D, Dimitropoulos N, Theodoridis S, Multiscaled Mammographic Mass Shape Analysis and Classification Using Neural Networks, 14th International Conference on Digital Signal Processing (DSP-2002), July 1-3, 2002, Santorini, Greece.

[17] Cheeseman P, Stutz J, Bayesian Classification (AutoClass): Theory and Results. NASA Ames Research Center: Artificial Intelligence Research Branch, 1995.

[18] Theodoridis S, Koutroumbas K, Pattern Recognition (2nd Ed.). USA: Academic Press, 1999.

[19] Leeden R, Vrijburg K, Leeuw J, A review of two different approaches for the analysis of growth data using longitudinal mixed linear models: Comparing hierarchical linear regression (ML3, HLM) and repeated measures designs with structured covariance matrices (BMDP5V). The Statistical Software Newsletter, SSNinCSDA 1996; 20: 583-605.

[20] Stone M, Cross-validatory choice and assessment of statistical predictions. J R Stat Soc 1974; B36(1): 111-47.

[21] Devroye L, Gyorfi L, Lugosi G, A Probabilistic Theory of Pattern Recognition. New York: Springer-Verlag Inc, 1996.

[22] Haykin S, Neural Networks: A Comprehensive Foundation, 2nd/Ed. New Jersey: Prentice Hall, 1999.

[23] Poggio T, Girosi F, Networks for approximation and learning. Proc. IEEE 1990; 78: 1481-97.

[24] Reed R, Prunning algorithms – A survey. IEEE Trans Neural Networks 1998; 4(5): 740-7.

[25] Christianini N, Shawe-Taylor J, An Introduction to Support Vector Machines.

Cambridge, UK: Cambridge University Press, 2000.

Figure 1 – Morphological shape type representations of mammographic tumors.

Figure 2 – Likelihood probability distributions of tumor boundary sharpness versus diagnosis and the corresponding bimodal normal distribution. Interference between the two normal distribution curves constitutes the statistical error probability due to aliasing between the two classes. The true aliasing error was calculated by applying the specific decision threshold for boundary sharpness indicator in the current set of 130 mammographic images.

Table 1 – Clinical findings and features normally implicated in mammogram evaluation.

FEATURES LIST	MORPHOLOGICAL DATA	TEXTURAL DATA	OTHER	CPU	DOCTOR
TUMOR					
Intramammary node			✓		✓
Size (general view)	✓			✓	✓
Inclusion of fat (%)		✓		✓	✓
Degree of irregularity	✓				✓
Type of irregularity	✓	✓	✓		✓
Stellate border	✓			✓	✓
Indistinct border		✓	✓	✓	✓
Density (hypo/iso/hyper)		✓		✓	✓
Homogeneity		✓		✓	✓
Location			✓	✓	✓
Diameter	✓			✓	✓
Boundary shape (type)	✓			✓	✓
MICROCALCIFICATIONS					
Size of cluster (general view)	✓			✓	✓
Number of elements		✓		✓	✓
Shape of cluster	✓			✓	✓
Variability of size of elements		✓		✓	✓
Irregular shape of elements		✓	✓		✓
Linear or branching elements	✓			✓	✓
SECONDARY SIGNS					
Architectural distortion			✓		✓
Asymmetric density		✓	✓		✓
Skin thickening or retraction			✓		✓
Regional calcifications		✓	✓		✓
PREVIOUS HISTORY					
Availability			✓		✓
Comparability			✓		✓
Existence of abnormality in previous study			✓		✓
CORRELATION WITH CLINICAL FINDINGS					
Availability			✓		✓
Correlation: location of clinical findings with radiographic study			✓		✓
Correlation: size/extent of clinical findings with radiographic study			✓		✓
Level of suspicion due to clinical findings			✓		✓
OTHER DATA					
Age			✓		✓
Benign / Malignant (histological)			✓		✓

Table 2 – Final annotation list and quantification details.

Qualitative Feature	Range
Patient's Age	(true age)
Mass Existence	Yes / No
Microcalcifications Existence	Yes / No
Fat Percentage	0%…100%
Boundary Sharpness	0%…100%
Mass Density	L (hypo) / M (iso) / H (hyper)
Mass Homogeneity	1…10
Mass Shape Type	1 (round) / 2 (lobulated) / 3 (micro-lobulated)/ 4 (stellate)
Hisologic Diagnosis	B (benign) / M (malignant)

Table 3 – Distribution of the 4 morphological shape types against diagnosis. Percentages are calculated per column.

	Round	Lobulated	Micro-lobulated	Stellate
Benign	25 83%	18 95%	2 5%	1 3%
Malignant	5 17%	1 5%	41 95%	37 97%

Table 4 – Distribution of the “1+2” and “3+4” grouped morphological shape types against diagnosis. Percentages are calculated per column.

	Round + Lobulated	Micro-lobulated + Stellate
Benign	43 88%	3 4%
Malignant	6 12%	78 96%

Table 5 – Statistics of benign and malignant cases. Cells indicating "-" mean that the specific parameter could not be calculated due to zero variance. All confidence ranges for mean values were calculated for significance level (alpha) 0,95.

BENIGN (cases: 46)	Mean	Standard Deviation	Skewness	Kurtosis	Mean Conf.Range
Patient's age	47,457	8,939	0,119	1,623	± 2,583
Microcalcifications Presence	0,783	0,417	-1,417	0,006	± 0,121
Fat% Inclusion	0,126	0,270	2,208	3,539	± 0,078
Boundary Sharpness	0,808	0,205	-2,925	8,514	± 0,059
Tumor Density	0,326	0,701	-0,555	-0,781	± 0,203
Tumor Homogeneity	7,109	1,464	-0,951	0,979	± 0,423
Tumor Shape Type	1,543	0,690	1,324	2,215	± 0,199
MALIGNANT (cases: 84)	Mean	Standard Deviation	Skewness	Kurtosis	Mean Conf.Range
Patient's age	57,631	9,079	0,098	-0,188	± 1,942
Microcalcifications Presence	0,810	0,395	-1,605	0,590	± 0,084
Fat% Inclusion	0,000	0,000	-	-	-
Boundary Sharpness	0,255	0,264	1,017	-0,032	± 0,056
Tumor Density	0,798	0,485	-2,420	5,260	± 0,104
Tumor Homogeneity	5,381	1,605	0,266	-0,328	± 0,343
Tumor Shape Type	3,310	0,776	-1,398	2,394	± 0,166

Table 6 – Test statistics of benign versus malignant cases. For F-test values, cells indicating "-" mean that the specific parameter could not be calculated due to zero variance. Class boundary values indicating "N/S" mean that the clustering of the specific feature could not be qualified as bimodal normal distribution model, i.e. it could not produce any linear discrimination on the base classes.

STATISTICS (BEN.vs.MAL)	Classes Boundary	T-test Value	F-test Value	Statistical Errors	Error Probability
Patient's age	47,791	1,704e-08	0,926	29	22,31%
Microcalcifications Presence	N/S	0,721	0,660	-	-
Fat% Inclusion	N/S	0,003	-	-	-
Boundary Sharpness	0,601	1,041e-24	0,067	15	11,54%
Tumor Density	N/S	0,127e-03	3,801e-03	-	-
Tumor Homogeneity	7,006	1,181e-08	0,505	34	26,15%
Tumor Shape Type	2,226	3,890e-24	0,392	9	6,923%

Table 7 – Feature ranking lists produced by T-test, bimodal normal distribution errors and MANOVA evaluations against diagnosis. For non-linearly separable bimodal normal distribution cases, the overall shape and aliasing of the underlying distributions are considered instead of true misclassification errors.

T-test ranking	Bin.Dist.Err. ranking	MANOVA ranking
Boundary Sharpness	Mass Shape Type	Mass Shape Type
Mass Shape Type	Boundary Sharpness	Boundary Sharpness
Patient's Age	Patient's Age	Patient's Age
Mass Homogeneity	Mass Homogeneity	Fat Percentage
Mass Density	Mass Density	Microcalcifications?
Fat Percentage	Fat Percentage	Mass Homogeneity
Microcalcifications?	Microcalcifications?	Mass Density

Table 8 – True discrimination efficiency of individual features through LSMD classification.

Qualitative Feature	LSMD succ% - diag.
Mass Shape Type	93,1%
Boundary Sharpness	86,1%
Fat Percentage	74,6%
Mass Density	73,1%
Mass Homogeneity	73,1%
Patient's Age	68,5%
Microcalcifications?	60,8%

Table 9 – Success rates of all classifiers against diagnosis prediction, with and without shape type input.

TARGET: TUMOR DIAGNOSIS		
Classifier Model	Accuracy (%) without shape type info	Accuracy (%) including shape type info
LDA	87,69%	93,85%
K-NN	91,54%	93,08%
LSMD	89,23%	93,08%
MLP	91,54%	91,54%
RBF	90,77%	91,54%
C-SVC/RBF	93,85%	94,62%

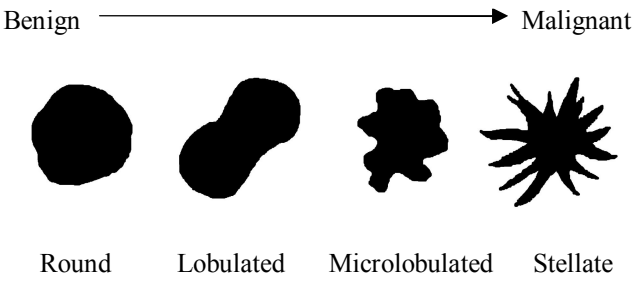

Figure 1 – Morphological shape type representations of mammographic tumors.

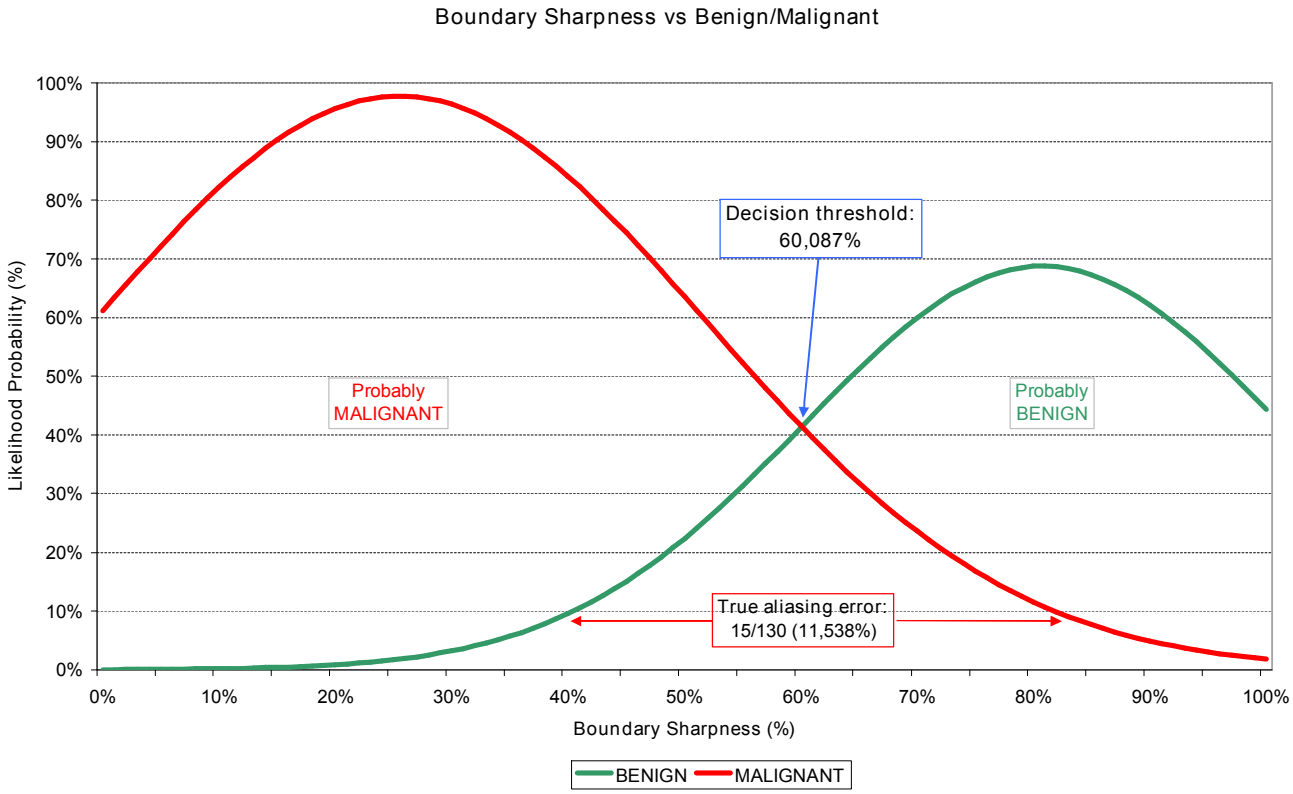

Figure 2 – Likelihood probability distributions of tumor boundary sharpness versus diagnosis and the corresponding bimodal normal distribution. Interference between the two normal distribution curves constitutes the statistical error probability due to aliasing between the two classes. The true aliasing error was calculated by applying the specific decision threshold for boundary sharpness indicator in the current set of 130 mammographic images.