International Journal of Medical, Medicine and Health Sciences
ISSN: 2517-9969
Vol:1, No:2, 2007

Pattern Recognition Techniques Applied to Biomedical Patterns

Giovanni Luca Masala

***Abstract*—**Pattern recognition is the research area of Artificial Intelligence that studies the operation and design of systems that recognize patterns in the data. Important application areas are image analysis, character recognition, fingerprint classification, speech analysis, DNA sequence identification, man and machine diagnostics, person identification and industrial inspection. The interest in improving the classification systems of data analysis is independent from the context of applications. In fact, in many studies it is often the case to have to recognize and to distinguish groups of various objects, which requires the need for valid instruments capable to perform this task. The objective of this article is to show several methodologies of Artificial Intelligence for data classification applied to biomedical patterns. In particular, this work deals with the realization of a Computer-Aided Detection system (CADe) that is able to assist the radiologist in identifying types of mammary tumor lesions. As an additional biomedical application of the classification systems, we present a study conducted on blood samples which shows how these methods may help to distinguish between carriers of Thalassemia (or Mediterranean Anaemia) and healthy subjects.

***Keywords*—**Computer Aided Detection, mammary tumor, pattern recognition, dissimilarity

I. Introduction

PATTERN recognition, the act of taking in raw data and making an action based on the category of the pattern , has been crucial for our survival, and over the past tens of millions of years we have evolved highly sophisticated neural and cognitive system for such tasks[1]. It encloses sub-disciplines like discriminant analysis, feature extraction, error estimation and cluster analysis (together known as statistical pattern recognition).

The conceptual boundary between feature extraction and proper classification can be somewhat arbitrary [1]. An ideal feature extractor would yield a representation that would make the job of the classifier trivial; conversely, an omnipotent classifier would need the help of a sophisticated feature extractor. The traditional goal of the feature extractor is to characterize an object to be recognized by measurements whose values are very similar for objects in the same category, and very different for an object in different categories. This lead to the idea of seeking distinguishing features, which are invariant to irrelevant transformation of the input. Generally, we want the features to be invariant to translation, whether horizontal or vertical because rotation is also irrelevant for classification, we would also like features to be invariant to the rotation. Therefore, we may also want the features to be invariant to scale. In general, features that describe properties such as shape, colour and texture are invariant to translation, rotation and scale.

All the classifiers need a normalization of the most significant features to reduce the time of calculation and the cost of the measures, which can remarkably improve the performances of a classifier and reduce the noise of the data:
In general, normalization methods standardize data to the range [-1, 1] or [0, 1] or regularize data to have zero mean and unity standard deviation.

Also a feature reduction can be achieved through classical method [1],[7] like a Minimal Entropy where the features with minimum entropy are selected, so the uncertainty of the measures is reduced or Divergence that is a measure of separability of the classes found on the research of the minimum classification error. Component Analysis [1] is a very important unsupervised approach to finding the right features: Principal Component Analysis (PCA) algorithm uses the Karhunen Loève transformation to reduce the features using the main eigenvalues of the covariance matrix of the mean vectors. It is possible to implement PCA through neural network with a three layer auto-encoder net (Sanger's algorithm) [1]-[3], which works well with high dimensional features vectors. If data represent complicated interactions of features, then the linear subspace may be a poor representation and nonlinear component may be needed. It is possible to implement a NonLinear Component Analysis (NLCA) through a five layers neural network with two layers of nonlinear units [1], [4]. While PCA and NLCA seek directions in a features space that best represent data in a sum-squared error sense, Independent Component Analysis (ICA) instead seeks directions that are most independent from each other. In particular a Fast Independent Component Analysis (FASTICA) is used to provide a computationally quick method of estimating the unobserved independent components [1],[5]. Often it is possible to have good performances with a Linear Discriminant Analysis (LDA) or Fisher's linear discriminant [1],[7] that supply a preliminary linear classification of the features by the use of training set. Furthermore many linear algorithms which use data in terms

This work in large part was supported by the Italian National Institute of Nuclear Physics (INFN) with the project MAGIC-5 (Medical Application on Grid Infrastructure Connection). The work on Thalassemia dataset was supported by the Regione Sardegna, Italy, through the "Consorzio 21".

G. L. Masala is with the Struttura Dipartimentale di Matematica e Fisica dell'Università di Sassari and Sezione INFN di Cagliari, Italy, Via Vienna 2, Sassari, 07100, Italy, phone:+39079229486, fax: +39079229482; E-mail: giovanni.masala@ca.infn.it

of dot products only can be non-linearised by substituting a kernel function for the dot products. To this purpose a Kernel Component Analysis is used [6]. Other important applications make use of the Kohonen Maps or Self Organizing Features Maps (SOM), that is two-layers NN fully connected to a large number of outputs corresponding to the points along the target lines [1]-[2],[8]. Another approach uses similarity or dissimilarity data representation [1], [9]. This representation is realized starting from a previous set of data that is compared with a set of prototypes, obtaining therefore a new set on which the next elaborations can be done.

In literature [1]-[22] many statistical methods of supervised classification are available. The supervised classification is fundamentally based on the theory of the probability: by using the probability density of the values of the feature of the various classes to determine the most probable. The probability density can be known or estimated on the basis of the training-set whose classes are known. The majority of supervised classifications are implicitly or explicitly based on this concept.

Generally, the approach to linear separable classes is that classifiers with linear discriminative functions work very well. If instead the problem is not linearly separable then the classifiers with linear discriminative functions are in trouble and have low performances. Later the method of the Support Vector Machine (SVM) will be introduced, which is an optimal evolution of the theory of linear discriminative functions; with the right conditions in the determination of hyperplane separation of the classes, it works also on the non-linear separable cases. Many classifiers exist based on the theory of the Bayesian decision but in many cases due to the parametric methods of the theory of the Bayesian decision, is not applicable in how the available knowledge is not easily expressible in probabilistic terms. In such cases, techniques of simplified classification are used. Such techniques concur, for example, to determine the discriminating functions for various classes beginning from samples. This technique will be exemplified when the K-Nearest Neighbours (K-NN) are introduced. Moreover, there exist other non-algorithmic methods such as, artificial neural networks. They were created in order to reproduce typical activities of the human brain such as, perception of images, acknowledgment of shapes, understanding of language, motor coordination, and etc,. Recent success of neural networks are fundamentally used to understand the mechanisms that regulate the nervous system in the hope to realize parallel architectures capable to carry out difficult tasks in respect to sequential architectures. Among the various types of existing neural networks we will use those adapted for the supervised classification: the Feed-Forward Neural Networks (FF-NN) with the algorithm of back propagation.

The analysis carried out is an original work of systematic application of classification methods to different types of data. In this article we report the results obtained with some classifiers as a Feed Forward Neural Network, a K-Nearest Neighbours and a Support Vector Machine[1]-[2],[7]-[8],[10]-[18], used on two different dataset. The first dataset is obtained from the database of digitized mammographic images of the project MAGIC-5 [19]-[22]. The second dataset [23]-[24] is extracted from a database of patient clinical records based on a thalassemia screening carried out on Public School's students.

More generally, in this work it is proved that the best data analysis algorithm is dependent on data but is independent of the nature of the problem. Therefore the intercrossed use of the proposals methods is convenient for the search of the best performances. To this purpose, classifiers conceptually much various between them are used. Moreover an efficacious process of optimization of the classifiers architecture is necessary in order to obtain good quality results.

Furthermore it is also put in evidence that the representation of data is not negligible. We show a new approach based on dissimilarity representation (a transformation of the vectors of features in a space of distances from set of prototypes) that can increase the performances of the classification systems.

II. General Methods

For the two different datasets we make a comparative study of the following classifiers:

- K-Nearest Neighbours (K-NN). For this type of deterministic classifier, it is necessary to have a training set which is not too small, and a good discriminating distance. KNN performs well in multi-class simultaneous problem solving. There exists an optimal choice for the value of the parameter K, which brings to the best performance of the classifier. This value of K is often approximately close to $N^{1/2}$.
- Feed-Forward Neural Networks (FF-NN). The selected supervised classifier is a Multi Layer Perceptron, back-propagation network, trained with gradient descent learning rule with "momentum", so as to quickly move along the direction of decreasing gradient, thus avoiding oscillations around secondary minima.
- Support Vector Machine (SVM). This algorithm creates a hyperplane that separates the data into two classes with the maximum-margin. Given training examples labeled either "yes" or "no", a maximum-margin hyperplane is identified which splits the "yes" from the "no" training examples, such that the distance between the hyperplane and the closest examples (the margin) is maximized. There is way to create non-linear classifiers by applying the kernel trick to maximum-margin hyperplanes. The resulting algorithm is formally similar, except that every dot product is replaced by a non-linear kernel function. This allows the algorithm to fit the maximum-margin hyperplane in the transformed feature space. The transformation may be non-linear and the transformed space high dimensional; thus though the classifier is a hyperplane in the high-dimensional feature space it may be non-linear in the original input space.

III. Mammographic Dataset

The mammographic database of the Magic-5 project contains approximately 5000 digitized images. From this database a dataset of ROI is obtained, as shown in table I.

TABLE I
COMPOSITION OF THE MAMMOGRAPHIC DATASET

	Number of total samples (ROI)	Number of positive samples (ROI)
Training set	235	145
Testing set	238	147

The mammographic images (18x24 cm^2, digitized by a CCD linear scanner with a 85 μm pitch and 4096 gray levels) are fully characterized: pathological ones have a consistent description which includes the radiological diagnosis and the histological data, while non pathological ones correspond to patients with a follow up of at least three years [19].

IV. Computer Aided Detection

The Computer Aided Detection (CADe) system presented here is an expert system based on two preliminary steps before classification: a ROI-hunter and a features extractor. The focus is on the automated analysis of massive lesions, i.e. the search for rather 'large objects' in the mammographic image, usually characterized by peculiar shapes.

The ROI-hunter was described in ref. [20]. The aim of this stage is to reduce the data amount to process by searching for Regions Of Interest (ROIs) that include a lesion with high probability. Only selected regions are stored for the next processing steps, rather than the whole mammogram as shown in fig. 1.

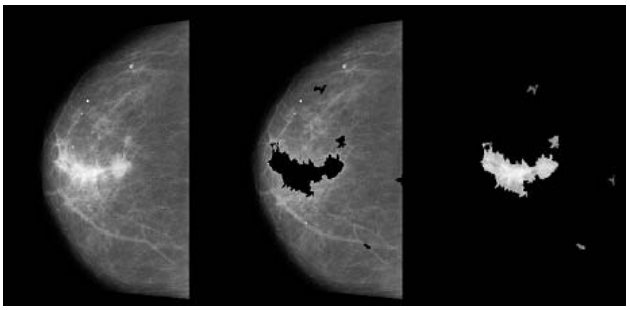

Fig. 1 The original image (left), the image without the ROI (middle) the extracted ROI (right)

The features extractor module: In this paper 16 features are extracted from the segmented masses. For each ROI we consider the minimal rectangular portion of the image which fully includes the ROI. The co-occurrence matrix is constructed from the image by estimating the pairwise statistics of pixel intensity, thus relying on the assumption that the texture content information of an image is contained in overall or average spatial relationship between pairs of pixel intensities [21]-[22]. Let us define the distance d between two pixels of the image as the minimum number of steps for going from one pixel to the other, where steps in the horizontal, vertical and diagonal directions are allowed. Two pixels at distances d and polar angle α are said to have a polar separation (d, α). Let G be the number of grey levels in the image ($G = 2^n$ for an n-bit image). For a given polar separation (d, α) a co-occurrence matrix M is a G x G matrix, which elements p_{ij} represent the fraction of pixels with grey levels i and j and polar separation (d, α).

In our work [21]-[22] we considered only displacements d = 1 at quantized angles $\alpha = k\pi/4$, with k = 0,1,2,3. Textural features can be derived from the co-occurrence matrix and used in texture classification in place of the single co-occurrence matrix elements. These features are shown in table II below. The values of these features are sensitive to the choice of the direction α. So using 4 co-occurrence matrices and 4 features for each matrix the record to be classified is composed by 16 features.

TABLE II
FEATURES EXTRACTED FROM THE CO-OCCURRENCE MATRIX

contrast:	$\sum_{i,j}(i-j)^2 \cdot p(i,j)$
homogeneity:	$\sum_{i,j}\frac{p(i,j)}{1+\lvert i-j\rvert}$
entropy:	$-\sum_{i,j}\ln[p(i,j)]\cdot p(i,j)$
energy:	$\sum_{i,j}p(i,j)^2$

V. Features Representation

Every example from the dataset for disposition after the module of features extraction becomes a numerical vector of real values correspondent to a class (known for the training vectors and unknown for those of test) and therefore comes untied through this abstraction from the original specific problem. Therefore, a representation of the features is needed for disposition in the information space. After it is possible to apply a classifier that decides which class model belongs to each of these vectors.

On each of these blocks it is possible to perform operations in order to maximize the results of the classification. In particular, it is very important to model the problem classes well. In fact, it is necessary to define a homogenous set able to distinguish the classes between of them.

Once the classes are defined it is possible to emphasize them through the operations on the vectors of the features. Generally, all the classifiers need a normalization of the features (most significant ones) to reduce the time of calculation and the cost of the measures which can remarkably improve the performances of a classifier and reduce the noise of the data [1]-[2]. For this purpose in next paragraph, a new approach is proposed by data transformation of dissimilarity that is a transformation of the features space in a distances space with respect to a set of prototypes. The use of the dissimilarities is especially interesting when features are

difficult to obtain or when they have low discriminative power. The representation based on dissimilarity relations between objects is an alternative to the traditional feature-based description.

A. Dissimilarity representation

Pattern recognition relies on the description of regularities under observation of object classes [9]. How this knowledge is extracted and represented is of crucial importance for learning. It is believed that alternative representations of feature-based descriptions of objects should be studied as they may capture different characteristics of the problem we want to analyze.

Proximity underscores the description of a class as a group of objects possessing similar characteristics. This implies that the notion of proximity is more fundamental than the notion of a feature or of a class. Thereby, it should play a crucial role in class constitution. This proximity should be possibly modelled such that a class has an efficient and compact description. For a number of years, the principle followed was that which advocates the learning from dissimilarity representations. They are derived from pairwise object comparisons, where the shared degree of commonality between two objects is captured by a dissimilarity value. Such representations are general and can be derived in many ways, e.g., from raw (sensor) measurements such as images, histograms or spectra or form initial representations by features, strings or graphs. The choice of such representations can also be suggested by an application or data specification. In fact, in all types of problems referring to string-graph, shape-or template–matching, as well as to all kinds of information retrieval or image retrieval, the use of (dis)similarities seems to be the most feasible approach.

The K-Nearest Neighbours classifier is commonly practiced on dissimilarity representations due to its simplicity and good asymptotic behavior (on metric distances). It has, however, three main disadvantages: larges storage requirements, large computational effort for evaluation of new objects and sensitivity to noisy examples. Prototype optimization techniques can diminish these drawbacks, so research efforts have been devoted to this task; see e.g. . From the initial prototypes, such as the objects in the training set, the prototype optimization chooses or constructs a small portion of them such that a high classification performance of the 1-NN rule is achieved. This might be especially of interest when the dissimilarity measure is based on expensive object comparisons.

Although the K-NN rule is mostly applied to metric distances, many non-metric distances are often designed to respond better to practical requirements. They are naturally derived when images or shapes are aligned in a template matching process. For instance, in computer vision, it is known that in the presence of partially occluded objects, non-metric measures are preferred [9]. Other examples are pairwise structural alignments of proteins, variants of the Hausdorff distance and normalized edit-distances. By common-sense reasoning, the principle behind the voting among the nearest neighbours can be applied to non-metric dissimilarities. The K-NN rule can also work well in such cases. It is simply more important that the measure itself is discriminative for the classes than its strict metric properties. However, many traditional prototype optimization methods are not appropriate for non-metric dissimilarities, especially if no accompanying feature-based representation is available, as they often rely on the triangle inequality.

Since all object in the training set can be initially used in training, it suggests to construct classifiers defined as weighted linear (or quadratic) combinations of the dissimilarities to a set of selected prototypes. In such a framework the metric requirements are not essential. In previous experiments it has been found that random selection of prototypes often works well.

B. Modelling the classes

In international literature the problem of massive lesions is normally dealt with as a two class problem where the relevant discrimination is between healthy ROI and pathological ROI. The idea is to increase the number of classes to five, starting from the assumption that the radiologist can distinguished four main typologies of masses. The interest is not to create an expert system able to classify between five types of object (including the healthy case). To distinguish the various types of masses it can be useful to discriminate better the healthy cases from the pathological cases using dissimilarity representation. Preliminary studies involving five classes classification (without dissimilarity representation) do not supply substantial improvements on the classifiers performances.

C. Dissimilarity construction

To construct a decision rule on dissimilarities [9],[22], the interesting set T with n elements and the representation set R with r elements will be used. R consists of prototypes which are representatives of all involved classes. In the learning process, a classifier is built on the n x r dissimilarity matrix D (T,R), relating all training objects to all prototypes. The information on a set S of s new objects is provided in terms of their distances to R, *i.e.* as an s x r matrix D(S,R). In our case the Euclidean distance and a representative set R composed by r = 24 records with m = 16 features (characterizing the ROI) are chosen. The R set is composed by 12 healthy ROIs and 12 pathological ROIs extracted from several good images (with different tissue, type of massive lesions, projection, side, and other tips) which are a good database sampling.

Furthermore, a 5 classes division of dataset is made (using 4 classes to distinguish various types of massive lesions) only to improve the difference between the pathological classes then non pathological class by dissimilarity representation.

The dissimilarity representation and the reduction of the dimensionality is made by the following two steps:

→ Calculation of the distance for each record i of the interesting set T to each record k of the representation set R.

Each record of T and R is a vector with m elements (number of features):

$T_i = (t_{i1}, t_{i2}, .., t_{im})$ $\quad i = 1,..,n$ (1)

with n defined as the number of records (ROIs) of the set T

$R_k = (r_{k1}, r_{k2}, .. r_{km})$ $\quad k = 1,..,r$ (2)

with r = 24 defined as the number of records (ROIs) of the set R

$$d_{ik}{}^{j} = \sqrt{\sum_{m} (t_m - r_m)^2} \quad (3)$$

with m = 1,..,16, k = 1,..,r and j = 0,..,4 the class of the R set

→ For each record i of the set of interest, the class j of each record k of the R set is known to the expert system, while the classes of the T set are unknown.

For each record i of the interesting set T we can build the vector of the minimum distances from all records of R in the class j, so to obtain a features reduction:

$d_i = (d_{min}{}^{0}, d_{min}{}^{1}, .., d_{min}{}^{j})$ (4)

The main steps are shown in the following fig. 2.

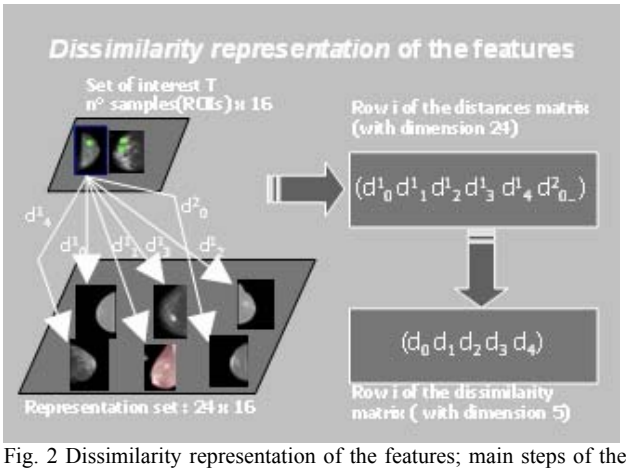

Fig. 2 Dissimilarity representation of the features; main steps of the algorithm are shown

VI. K-FOLDER CROSS VALIDATION

It is important that the results of the classification methods be reproducible independently of chosen samples. A strong model of cross validation with five random samples extraction and in particular a k-folder cross validation (with k=5) supplies mean results unaffected by variation correlated to the choice of the training samples [8]. Such validation uses five training set and a testing set whose size is reported in table I. So the mean of 5x2 results is displayed in tables VI and VII.

VII. RESULTS ON MAMMOGRAPHIC DATASET

Using sensitivity (percentage of pathologic ROIs correctly classified) and specificity (percentage of non pathologic ROIs correctly classified), the results obtained with this analysis are described in terms of the ROC (Receiver Operating Characteristic) curve [1]-[7], which shows the true positive fraction (sensitivity), as a function of the false positive fraction (1-specificity) obtained varying the threshold level of the ROI selection procedure. In this way, the ROC curve produced allows the radiologist to detect massive lesions with predictable performance, so that he can set the desired true-positives fraction value and know the corresponding false-positives fraction value. The overall performance on the two class problem is evaluated in term of the area under the ROC curve obtaining for each classifier the tables below.

In table III we show the performance of the classifiers before the dissimilarity representation:

TABLE III
PERFORMANCES OF THE CLASSIFIERS ON THE FEATURES-SPACE THROUGH 5X2 K-FOLDER CROSS VALIDATION

Classifiers	Area under ROC curves
FF-NN	A_z = (0.80 ± 0.02)
K-NN	A_z = (0.78 ± 0.02)
SVM	A_z = (0.75 ± 0.02)

In table IV we show the results of the classifiers after the dissimilarity representation:

TABLE IV
PERFORMANCE OF THE CLASSIFIERS ON THE DISSIMILARITY-SPACE THROUGH 5X2 K-FOLDER CROSS VALIDATION

Classifiers	Area under the ROC curve
FF-NN	A_z = (0.82 ± 0.03)
K-NN	A_z = (0.87 ± 0.03)
SVM	A_z = (0.78 ± 0.03)

In fig.3 we show the ROC Curves of the classifiers after dissimilarity representation.

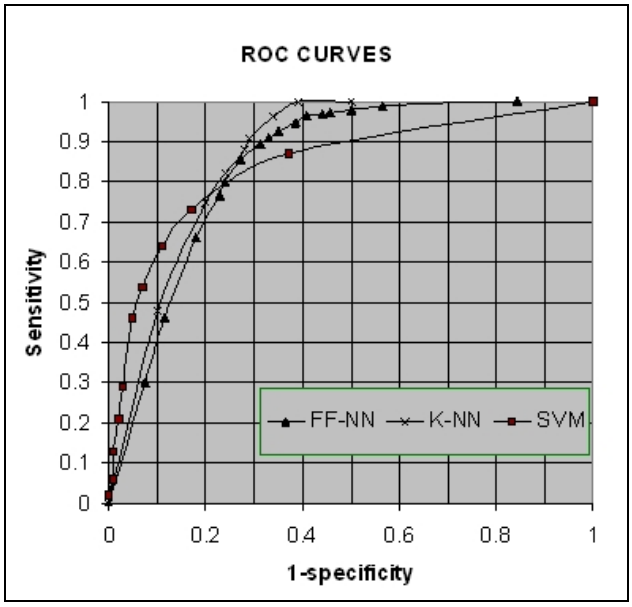

Fig. 3 Roc curves for the classifiers after dissimilarity representation

In figure 4 the performances of the K-NN with and without dissimilarity representation are shown.

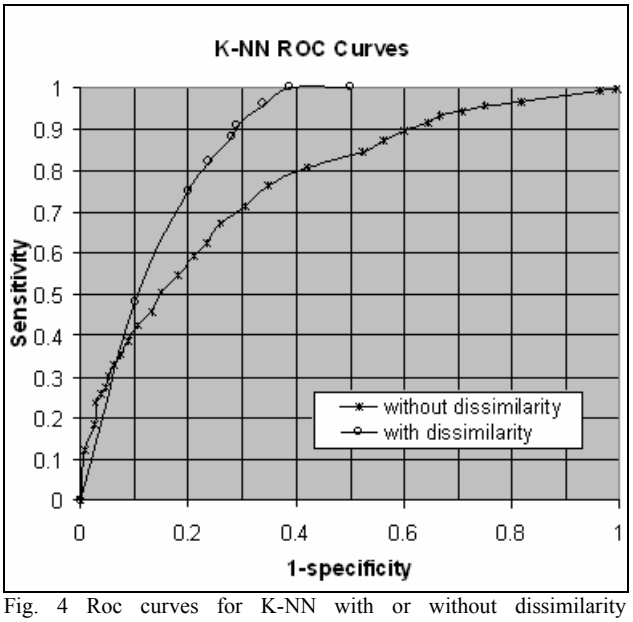

Fig. 4 Roc curves for K-NN with or without dissimilarity representation

VIII. Discussion

The discriminating performance of the algorithm was checked by an evolution of a linear method as SVM, a statistic method as K-NN and a non-algorithm method as FF-NN. The best results in terms of area under the ROC curve and sensitivity was best obtained by the K-NN than by the other classifiers. The low results obtained by SVM, indicate that all linear methods are not suitable for this medical problem.

Tests inter crossed on several of the parameters in order to obtain features from the co-occurrence matrices have demonstrated that the parameters considered in table II with the representation features-based, are those that supply the best performances. In the same way the parameters used in this work, are optimal ones. The performances in terms of classification on the base of the matrices of co-occurrence with the two representation of data are varied: it is obvious that the best classifier K-NN trained on representation with dissimilarities of the features, with an area of 0.87 ± 0.03 improves the performances of the same classifier, having an area 0.78 ± 0.02 optimized on the representation features-based. Furthermore K-NN on dissimilarity representation is better then the best FF-NN having an area of 0.80 ± 0.02 in the features-space and 0.82 ± 0.03 in the dissimilarity representation.

IX. Thalassemia Dataset

The initial database [22]-[23] that was used to train the computer to classify patients, consisted of 304 clinical records based on Thalassemia screening performed by the Ozieri Hospital on Public School's students. 8th grade students (14 to 15-year-olds) from several Public Schools in Northern Sardinia took part in the screening. Although the records can be considered a random sample, subjects with an iron deficiency were excluded from the test because blood iron levels must be normalised before Thalassemia diagnosis can be made. Through hemacytometeric data, HbA2 and genetic determination and the main Thalassemia defects (α3.7 and αNco variants), the medical diagnoses were made where HbA2 was determined to identify β carriers. It was determined that 27 subjects had HbA2 of $\geq$ 4%, while the other 277 cases had HbA2 $\leq$ 3%. The first group was diagnosed as being β carriers by medical analysis while genetic analysis was used to diagnose α carriers. Various attempts were made to normalise the values of the feature but none demonstrated particular advantages. Principal components analysis (PCA), reduced the number of relevant features (described below) but the following application of the classifiers after this transformation did not improve with respect to the case in which all the features were used. The features which were considered relevant for the classification were only the values of red blood cell count (RBC), haemoglobin (Hb), hematocrit (Ht) and mean corpuscular volume (MCV). The dataset composition is shown in table VI below.

X. Thalassemia Classification

TABLE V
Composition of the Thalassemia Dataset

	Normal cases	Pathological cases
Training set	196	141 normal, 37 α, 18 β
Testing set	108	55 normal, 44 α, 9 β

We wished to make a comparative analysis [23] of the performance of SVM and K-NN, by analysing their performance versus a combination of 3 specialised neural networks on the three class problem.

Such specialized FF-NN [22] are made each with 1 output neuron (4 input and 1 hidden neurons) for the discrimination of one class vs the other two and were trained-validated (with the back propagation algorithm) on cases pertaining to the respective output categories; a code to represent the single output allows the combination of the FF-NN. Furthermore a single three output FF-NN is tested but the results are inferior then previous system.

We propose a two layers classifier [23] system using SVM or K-NN. First dedicated layer needs for discrimination between healthy individuals or those with affected by two types of pathologies. The discrimination between α and β pathologies is obtained by using a second classification layer specialised on these patterns which receives, as input, the first classifier output which divides the cases into healthy and sick. Also a single three output classifiers are used but the performances are lower then the two layers system.

XI. Results on Thalassemia Dataset

The main results are reported in terms of accuracy (percentage of total cases correctly classified), specificity (percentage of non pathological cases correctly classified) and sensitivity (percentage of pathological cases correctly classified).

Using KNN classifier to differentiate between pathological cases and non-pathological cases allowed us to obtain 85% of accuracy with 93% of specificity and 77% of sensitivity, with K=23. It should be noted that KNN is not efficient in discriminating between healthy and sick patients. Using SVM, for the same type of discrimination between pathological and non pathological cases, the best result in terms of accuracy is approximately 89%, with 95% of specificity and 83% of sensitivity, obtained with a linear kernel, parameter C = 10 and leave-one-out (LOO) algorithm for estimating model quality.

The best classification between healthy and sick was with SVM, and so this classifier was used always as first classifier. Tests for discriminating α pathologies from β pathologies gave the same results for KNN and SVM classifiers. So these results are reported together specialized FF-NN (3 nets with 4 input, 1 hidden and 1 output neurons) performances in the tables.

In particular in table VI and fig. 5 are shown the results of the dedicated classifiers on the discrimination in the two classes problem between healthy and sick patients (α and β cases).

TABLE VI
COMPARATIVE STUDY OF THE DEDICATED CLASSIFIERS ON 2 CLASSES (DISCRIMINATION BETWEEN HEALTHY AND PATHOLOGICAL CASES)

Classifiers	accuracy	specificity	sensitivity
FF-NN	94%	95%	92%
SVM	89%	95%	93%
K-NN	85%	93%	77%

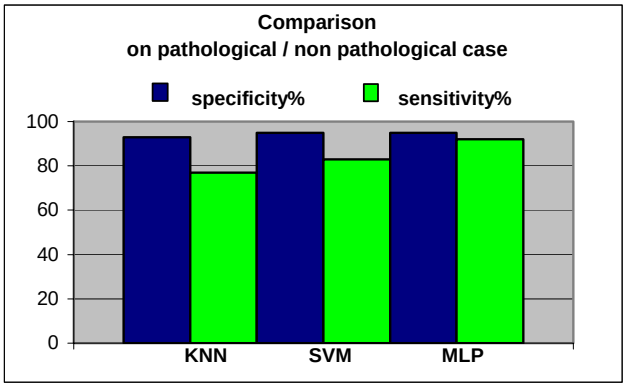

Fig. 5 Comparison of the dedicated classifiers on the 2 classes

In table VII are shown the results of the dedicated classifiers on three classes recognition. As evidenced in the table VII, in the second layer there is the same value with SVM or with K-NN. No results are showed in the table about a single three outputs K-NN, SVM (with *one-vs-all algorithm*) or FF-NN because the performances are lower then the other system.

TABLE VII
COMPARISON OF THE DEDICATED CLASSIFIERS ON THE 3 CLASSES

Classifiers	normal	β	α
FF-NN	95%	67%	91%
SVM/SVM	89%	89%	93%
K-NN/K-NN	85%	89%	93%

In the following fig. 6 are shown the details about the comparison on alpha / beta pathologies.

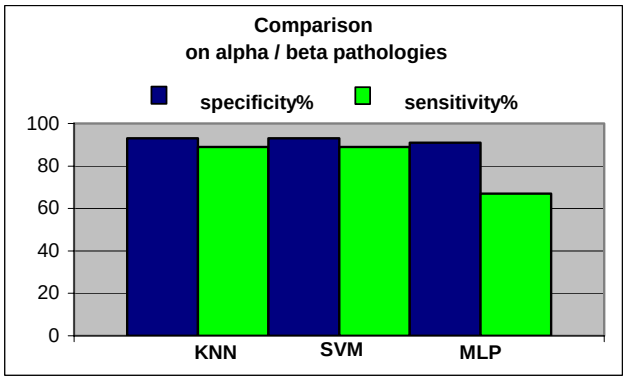

Fig. 6 Comparison of the dedicated classifiers on the discrimination of pathological classes

XII. Discussion

The proposed SVM two layer method is as efficient as compared to specialised neural networks. The FF-NN system gives slightly better results than SVM method although the amount of data available is limited. Both techniques enable Thalassemia carriers to be discriminated from healthy subjects with the same specificity, although the sensitivity of FF-NN is better then SVM. In the ability to recognise type α from type β Thalassemia the SVM classification performs similarly to the specialised FF-NN in terms of specificity and is more accurate in sensitivity then FF-NN. So in fig. 7 we propose a final multi-layer system where we use first a FF-NN in the classification between healthy / pathological cases; the division between α and β cases is made through a second layer using a SVM or K-NN.

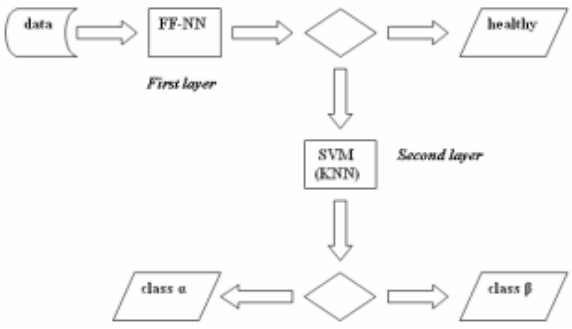

Fig. 7 Diagram for the final two layers classification systems

XIII. Conclusions

The objective of this work was to show some important aspect of the pattern recognition through two examples of efficient classification systems on biomedical data. The first case was the development of a Computer Aided Detection (CADe) system for the recognition of massive lesions from digital radiology images of analogical mammography.

An important part of the classifiers comparison was to use the same dataset and objective criteria for the analysis of the performances through the ROC curves.

The importance of the representation of data has been demonstrated. In fact, starting from the presupposed idea that not always the original space of the features is the ideal space where to make the analysis - the transformation for dissimilarity of the data has been operated. In particular, new carriers of distances and representatives of the samples in respect to a set of prototypes were realized. With the realization of this new approach, a new representation of mammographic dataset was on which the classifiers were tested. The test of the classifiers on the new space of distances supplied an area under the ROC curve better then the results obtained in the original space with the same method. Therefore, the new representation of the same data appears to improve the performances of the classifiers.

Moreover, these results show that there is not an optimal classifier that always functions better then others but for every data analysis based on their representation, it is always necessary to verify which system of classification is more suitable.

Instead, an ulterior improvement of the ratio signal on noise of the data can be given just with the description and modelling of the same classes: although it can serve to discriminate problems of *n* classes, it is necessary that the *n* classes are homogenous to their inside and sufficiently distinguished from the other. As in the case of massive lesions or cases of pathologies that contain objects structurally different between themselves which are more difficult to train the classifier in generalizing the training for the distinction between healthy cases.

Finally, with a good representation of data it is always a good practice to try classifiers of various principles in order to find what it maximizes their performances. Instead of completely changing the dataset and working on data coming from Thalassemia screening, it is shown how various classifiers can be competitive in various points of the same problem. In fact the proposed SVM method is as efficient as, and can be compared to, specialised neural networks on the same data. Both techniques enable Thalassemia carriers to be discriminated from healthy subjects with the same specificity, although the sensitivity of FF-NN is upper then SVM. In the ability to recognise type α from type β Thalassemia, the SVM classifier performs similarly to the specialised neural network classifier in terms of specificity and is more accurate in sensitivity then FF-NN. Therefore, also in this case it is demonstrated that the search for the best performances occurs by the comparison of different classification systems because the performances can often alternatively increased.

Finally it can be concluded that the importance of the intelligent systems goes beyond the single problem of classification in examination since it is applicable to every type of data. In fact, once the features extractor produces numerical vectors, the problem obtains an abstract representation on which it is possible to apply whichever experience matured in the Pattern Recognition.

Acknowledgment

G. L. Masala thanks for their part and role: the medical doctors and researchers of the project Magic-5 and all the authors of the work about Thalassemia classification in the ref. [10]-[11].

References

[1] O. Duda, P. E. Hart, D. G. Stark, "Pattern Classification", second edition, A Wiley-Interscience Publication John Wiley & Sons, 2001.
[2] S. Haykin "Neural Networks – A comprehensive foundation", second edition, Prentice Hall, 1999.
[3] T. D. Sanger , "Optimal Unsupervised Learning in a Single-Layer Linear Feedforward Neural Network", Neural Networks, vol. 2, pp. 459-473, 1989.
[4] M. A. Kramer, "Nonlinear Principal Component Analysis Using Autoassociative Neural Networks", AiCh Journal, vol. 37, No. 2, 1991.
[5] Aapo Hyv¨arinen, Erkki Oja, "Independent Component Analysis: Algorithms and Applications", Neural Networks Research Centre, Helsinki University of Technology, Finland, "Neural Networks", 13 (4-5):411-430, 2000.
[6] H. Gupta, A. K Agrawal, T. Pruthi, C. Shekhar, R. Chellappa, "An Experimental Evaluation of Linear and Kernel-Based Methods for Face Recognition," wacv, p. 13, Sixth IEEE Workshop on Applications of Computer Vision, 2002.
[7] S. Serpico, G. Vernazza, "Teorie e tecniche del riconoscimento", CUSL "Il gabbiano", 1997.
[8] Massimo Buscema & Semeion Group, "Reti Neurali artificiali e sistemi sociali complessi", volume 1 Teoria e modelli 1409.1, Franco Angeli, 1999.
[9] E. Pekalska, R.P.W. Duin, R.P.W. and P.Paclik, "Prototype Selection for Dissimilarity-based Classifiers", Pattern Recognition, vol. 39, no. 2, pp. 189-208, February 2006.
[10] V. N. Vapnik. "Statistical Learning Theory. Wiley", New York , 1998.
[11] M. Pontil, A. Verri "Properties of Support Vector Machines", Neural Computation, Vol. 10, pp 955-974, 1998.
[12] S. J. Russel, P.Norvig, "Artificial Intelligence. A modern approach", UTET, 1998.
[13] V. N. Vapnik. "Statistical Learning Theory. Wiley", New York , 1998.
[14] M. Pontil, A. Verri "Properties of Support Vector Machines", Neural Computation, Vol. 10, pp 955-974, 1998.
[15] N. Cristianini, J. Shave-Taylor. "An Introduction to Support Vector Machine (and other kernel-based learning methods)". Cambridge University Press, 2000.
[16] SVM_light software is available in the following location : ftp://ftp-ai.cs.unidortmund.de/pub/Users/thorsten/svm_light/current/ svm_light.tar.gz
[17] T. Joachims, Text Categorization with Support Vector Machines: Learning with Many Relevant Features, Proc. 10th European Conf. Machine Learning (ECML), Springer-Verlag, 1998.
[18] T. Mitchell "Machine Learning" , McGraw-Hill, 1997.
[19] Bottigli et al, Search of Microcalcification clusters with the CALMA CAD station. The International Society for Optical Engineering (SPIE) 4684: 1301-1310, 2002
[20] F. Fauci, S. Bagnasco, R. Bellotti, D. Cascio, S. C. Cheran, F. De Carlo, G. De Nunzio, M. E. Fantacci, G. Forni, A. Lauria, E.Lopez Torres, R. Magro, G. L. Masala, P. Oliva, M. Quarta, G. Raso, A. Retico, S.Tangaro, Mammogram Segmentation by Contour Searching and

Massive Lesion Classification with Neural Network, Proc. IEEE Medical Imaging Conference, October 16-22 2004, Rome, Italy; M2-373/1-5, 2004.

[21] U. Bottigli, B. Golosio, G. L. Masala, P. Oliva, S. Stumbo, D. Cascio, F. Fauci, R. Magro, G. Raso, R. Bellotti, F. De Carlo, S.Tangaro, I. De Mitri, G. De Nunzio, M. Quarta, A. Preite Martinez, P. Cerello, S. C. Cheran, E.Lopez Torres "Dissimilarity Application for Medical Imaging Classification" on proceedings of The 9th World Multi-Conference on Systemics, Cybernetics and Informatics WMSCI 2005, Orlando 10-13 July 2005, vol III pag 258-262, 2005.

[22] G. Masala, B. Golosio, D. Cascio, F. Fauci, S. Tangaro, M. Quarta, S. C Cheran, E. L. Torres, "Classifiers trained on dissimilarity representation of medical pattern: a comparative study" on Nuovo Cimento C, Vol 028, Issue 06, pp 905-912 , 2005.

[23] S.R. Amendolia, G. Cossu, M. L. Ganadu, B. Golosio, G.L. Masala, G.M. Mura "A Comparative study of K-Nearest Neighbour, Support Vector Machine and Multi-Layer Perceptron for Thalassemia Screening" on "Chemometrics and intelligent laboratory system" ;69:13-20, 2003.

[24] S.R Amendolia , A. Brunetti, P.Carta, G. Cossu, M.L. Ganadu, B. Golosio, G.M. Mura, M.G. Pirastru, A Real-Time Classification System of Thalassemic Pathologies Based on Artificial Neural Networks Medical Decision Making; 22:18-26, 2002.

[25] Timp S., Karssemeijer N., A new 2D segmentation method based on dynamic programming applied to computer aided detection in mammography, Medical Physics: 31; 958-971, 2004.

[26] Baydush A.H., Catarious D.M., Abbey C.K., Floyd C.E., Computer aided detection of masses in mammography using subregion Hotelling observers, Medical Physics: 30; 1781-1787, 2003.

[27] Tourassi G.D., Vargas-Voracek R., Catarious D.M. Jr, Floyd C.E. Jr, Computer-assisted detection of mammographic masses: A template matching scheme based on mutual information, Medical Physics: 30 (8); 2123-2130, 2003.

[28] Antonie M.L., Zaiane O.R., Coman A., Application of data mining techniques for medical image classification, Proc. of II Int. Work. On Multimedia Data Mining, USA, 2001.

[29] Vyborny CJ., Giger ML., Computer vision and artificial intelligence in mammography, AJR: 162; 699-708, 1994.

[30] Lai S., Li X., Bischof W., On techniques for detecting circumscribed masses in mammograms", IEEE Transaction on Medical Imaging: 8(4); 377-386, 1989.

[31] Hanley JA, McNeil B, The meaning and use of the area under a receiver operating characteristic (ROC) curve, Radiology: 143; 29-36, 1982.

[32] Hanley JA, McNeil B, A method of comparing the areas under receiver operating characteristic curves derived from the same cases, Radiology: 148; 839-843, 1983.