Available online at www.sciencedirect.com
SCIENCE DIRECT®
Chemometrics and Intelligent Laboratory Systems 69 (2003) 13–20
Chemometrics and intelligent laboratory systems
www.elsevier.com/locate/chemolab

A comparative study of K-Nearest Neighbour, Support Vector Machine and Multi-Layer Perceptron for Thalassemia screening

S.R. Amendolia[a], G. Cossu[b], M.L. Ganadu[c], B. Golosio[a], G.L. Masala[a,*], G.M. Mura[c]

[a] *Dipartimento di Matematica e Fisica dell'Universita' di Sassari, via Vienna 2, 07100, Sassari, Italy*
[b] *Azienda USL 1, Servizio Immunotrasfusionale, Presidio, Ospedaliero di Sassari, via Monte Grappa 82, 07100, Sassari, Italy*
[c] *Dipartimento di Chimica dell'Università di Sassari, via Vienna 2, 07100, Sassari, Italy*

Received 16 August 2002; received in revised form 19 May 2003; accepted 25 May 2003

Abstract

In this paper, we investigate the feasibility of two typical techniques of Pattern Recognition in the classification for Thalassemia screening. They are the Support Vector Machine (SVM) and the K-Nearest Neighbour (KNN). We compare SVM and KNN with a Multi-Layer Perceptron (MLP) classifier. We propose a two-classifier system based on SVM. The first layer is used to differentiate between pathological and non-pathological cases while the second layer is used to discriminate between two different pathologies (α-thalassemia carrier against β-thalassemia carrier) from the first output layer (pathological cases).

Using the parameters sensitivity (percentage of pathologic cases correctly classified) and specificity (percentage of non-pathologic cases correctly classified), the results obtained with this analysis show that the MLP classifier gives slightly better results than SVM although the amount of data available is limited. Both techniques enable thalassemia carriers to be discriminated from healthy subjects with 95% specificity, although the sensitivity of MLP is 92% while that of SVM is 83%.

Keywords: Thalassemia; K-nearest neighbour (KNN); Multi layer perceptron (MLP); Support vector machine (SVM); Principal component analysis (PCA); Artificial neural networks

1. Introduction

Thalassemias are genetic defects that are commonly found in many parts of the world including Africa, the Far East, and the Mediterranean regions [1]. In order to know how many people suffer from this disease, the heterozygous population should be screened. A 1st-level analysis using hemo-cromocytometric data and a 2nd-level examination (total HbA_2, globin chain synthesis, and genetic analysis) [2,3] should be carried out to identify α and β thalassemia carriers. Although total HbA_2, globin chain synthesis, and genetic analysis can identify carriers accurately, they are time consuming and expensive. Thus, an automated diagnostic support system could be of use when a first classification based only on the hemochromocytometric data is made. There is already a classification system based on MLP [4].

We want to make a comparative study of k-nearest neighbour [5], Support Vector Machine [6–10], and Multi-Layer Perceptron [11,12] for Thalassemia Screening to determine the best performance on thalassemias data reported in Ref. [4]. First, we investigate

* Corresponding author. Tel.: +39-79229586; fax: +39-79229482.
E-mail address: giovanni.masala@ca.infn.it (G.L. Masala).

0169-7439/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/S0169-7439(03)00094-7

the performance of SVM and KNN in the discrimination healthy/sick, then, we propose a multi-layer classifier based on SVM which, after the determination of healthy/sick, differentiates between the two different pathological cases (α-thalassemia against β-thalassemia).

2. Description of the classifiers

2.1. Support Vector Machine

Support Vector Machine (SVM) are fast replacing Neural Networks as tools for solving pattern recognition problems. They are based on some beautifully simple ideas and provide a clear intuition of what learning from examples is all about. More importantly, they are also showing high performances in practical applications [13–16]. In very simple terms, an SVM corresponds to a linear method in a very high dimensional feature space that is nonlinearly related to the input space. Even though we think of it as a linear algorithm in a high dimensional feature space; in practice, it does not involve any computations in that high dimensional space. By the use of kernels, all necessary computations are performed directly in the input space [6–8].

SVM performs pattern recognition for two-class problems by determining the separating hyperplane with maximum distance to the closest points of the training set [17]. These points are called *support vectors* that we show in Fig. 1.

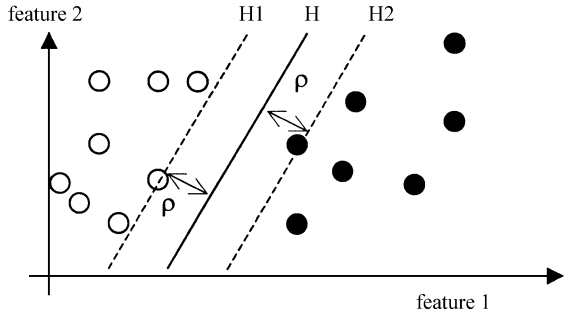

Fig. 1. The maximum margin for separating hyperplane is shown; the margin is called ρ. The hyperplane H is equidistant from the nearest class points (white and black group). These points are called *support vectors*.

The SVM algorithm constructs a separating hypersurface in the input space [7]. It acts as follows:

(i) maps the input space into a higher dimensional feature space through some nonlinear mapping chosen a priori (kernel);
(ii) constructs the Maximal Margin Hyperplane (MMH) in this features space; MMH maximizes the distance of the closest vectors belonging to the different classes to the hyperplane.

Let S be a set of l vectors $\mathbf{x}_i \in \mathbf{R}^n$, $(i = 1,2, \ldots, l)$ in a high (possibly infinite) n-dimensional space $\boldsymbol{H}$ which is called feature space. Each vector $\mathbf{x}_i$ belongs to either of two classes identified by the label $y_i \in \{-1, 1\}$. If the two classes are linearly separable, then there exists a hyperplane, defined by $\mathbf{w}\cdot\mathbf{x} + b = 0$, which divides S leaving all the vectors of the same class on the same side. It can be easily shown that MMH is given by the solution to the problem:

$$\begin{cases} \text{minimize } 1/2 \,\|\mathbf{w}\|^2 \\ \text{with } y_i(\mathbf{w}\cdot\mathbf{x}_i + b) \geq 1 \quad (i = 1, 2 \ldots, l) \end{cases} \tag{1}$$

where $b/||\mathbf{w}||$ is the distance between origin and hyperplane. This is a quadratic programming problem, solved by Karush–Kuhn–Tucker theorem. If we denote with $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_l)$, the l nonnegative Lagrange multipliers associated with the constraints, the solution to the problem is equivalent to determining the solution of the *Wolf dual* problem:

$$\begin{cases} \text{maximize } \sum_i \alpha_i - 1/2 \sum_{i,j} \alpha_i \alpha_j (\mathbf{x}_i \cdot \mathbf{x}_j) y_i y_j \\ \text{with } \sum_i \alpha_i y_i = 0 \qquad \alpha_i \geq 0 \end{cases} \tag{2}$$

The solution for $\mathbf{w}$ reads

$$\mathbf{w} = \sum_i \alpha_i y_i \mathbf{x}_i \tag{3}$$

The only α_i that can be nonzero in Eq. (3) are those for which the constraints of the first problem are satisfied with the equality sign. Since most of the α_i are usually null, the vector $\mathbf{w}$ is a linear combination of an often relatively small percentage of the vectors $\mathbf{x}_i$. These

vectors are termed support vectors and they are the only vectors of S needed to determine the MMH. The problem of classifying a new data vector $\mathbf{x}$ is now simply solved by looking at the sign of $\mathbf{w}\cdot\mathbf{x}+b$ with b obtained from the Karush–Kuhn–Tucker conditions [7].

In case the set S cannot be separated by any hypersurface due to the partial overlapping of the two classes, the previous analysis can be generalized by introducing l nonnegative *slack* variables $\xi=(\xi_1, \xi_2, \ldots, \xi_l)$ such that

$$y_i(\mathbf{w}\cdot\mathbf{x}_i+b)\geq 1-\xi_i \quad (i=1,2,\ldots,l) \tag{4}$$

The solution to

$$\begin{cases} \text{minimize } 1/2\,\|\mathbf{w}\|^2 + C\sum_i \xi_i & (i=1,2,\ldots,l) \\ \text{with } y_i(\mathbf{w}\cdot\mathbf{x}_i+b)\geq 1-\xi_i & (i=1,2,\ldots,l) \end{cases} \tag{5}$$

is called Soft Margin Separating Hyperplane (SMSH). Once again, the vectors satisfying the constraints of above with the equality sign are termed *support vectors* and are the only vectors needed to determine the decision surface. Similar to the linearly separable case, the dual formulation requires the solutions of a quadratic programming problem with linear constraints:

$$\begin{cases} \text{maximize } \sum_i \alpha_i - 1/2\sum_{i,j}\alpha_i\alpha_j(\mathbf{x}_i\cdot\mathbf{x}_j)y_iy_j \\ \text{with } \sum_i \alpha_i y_i = 0 \qquad 0\leq\alpha_i\leq C \end{cases} \tag{6}$$

Fig. 1 depicted an example of a linearly separable vectors belonging to two classes A_1 and A_2 (black and white circles), the SMSH H which separates them and the *support vectors*.

The entire construction can be extended rather naturally to include nonlinear separating hypersurfaces. Each vector $\mathbf{x}$ in input space is mapped into a vector $z=\phi(\mathbf{x})$ in a higher dimensional feature space. We can then substitute the dot product $<\phi(\mathbf{x}), \phi(\mathbf{y})>$ in feature space with a nonlinear function $K(\mathbf{x},\mathbf{y})$, named *kernel*. Conditions for a function to be a kernel are expressed in a theorem by Mercer [7].

An important family of admissible kernel functions are:

- Gaussian kernel, $K(\mathbf{x},\mathbf{y})=\exp(-||\mathbf{x},-\mathbf{y}||/2\sigma^2)$, where σ is the variance of the Gaussian.
- Polynomial kernel, $K(\mathbf{x},\mathbf{y})=(1+\mathbf{x}\ \mathbf{y})^d$, where d is the degree of the polynomial ($d=0$ for linear kernel).

Since in the dual formulation example vectors are present only in dot products, the performing of point (i) becomes quite simple.

We would like to stress here that SVM in form (5) does suffer from a limitation in two common situations: it is not suitable neither in case of unbalanced distributions, nor whether we need to outweigh misclassified examples of one class (e.g. when one type of misclassification is more serious than another). In order to generalize SVM algorithm to these cases, it is necessary to modify Eq. (5) in the following way [18]:

$$\begin{cases} \text{minimize } 1/2\,\|\mathbf{w}\|^2 + C^-\sum_i \xi_i^- + C^+\sum_i \xi_i^+ \\ \text{with } (\mathbf{w}\cdot\mathbf{x}_i+b)\geq 1-\xi_i^+, \quad (\mathbf{w}\cdot\mathbf{x}_i+b)\leq -1+\xi_i^+ \quad (i=1,2,\ldots,l) \end{cases} \tag{7}$$

where the first sum is for i with labels $y_i=-1$ and the second sum is for i with labels $y_i=+1$ and C^- and C^+ give different costs to false-positive and false-negative errors, respectively.

2.2. K-Nearest Neighbour

For this type of 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 $N^{1/2}$ [19].

3. Database and features normalisation

The initial database that was used to train the computer to classify patients consisted of 304 clinical records based on a thalassemia screening carried out by the Ozieri Hospital on public school's students. Several public schools of Northern Sardinia were used in the

test. Each eighth grade student (14- to 15-year-old boys and girls) took part in the screening. Although the records can be considered a random sample, subjects with iron deficiency were excluded from the test, as in their case, iron levels in the blood must be normalised before thalassemia diagnosis can be made [13].

Hemochromocytometric data, HbA_2, and genetic determination of the main thalassemia defects ($\alpha^{3.7}$ and α^{Nco} variants) were used to make the medical diagnoses. HbA_2 was determined to identify β carriers. Twenty-seven subjects had a HbA_2 of $\geq 4\%$, while HbA_2 was $\leq 3\%$ in the other 277 cases. The first group were diagnosed as being β carriers by medical analysis. Genetic analysis was used to diagnose α carriers.

Various attempts were made to normalise the values of the feature but none demonstrated particular advantages. Analysis of the principal components (PCA) [5,20] reduced the number of relevant features (described below) but the following application of the classifiers after this transformation does not bring improvements 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 RBC, Hb, Ht, and MCV (note: RBC: red blood cell count (10^6/μl). Hb: hemoglobin (g/dl). Ht: hematocrit (%). MCV: mean corpuscular volume (fl)) (without normalization). The data from the 304 clinical records were divided in a testing set (with 55 normal cases, 44 α and 9 β) made of 108 records, and in a training set made up of 196 clinical records (141 normal cases, 37 α and 18 β). The data were divided in the same testing and training set utilized in our previous work on the MLP classification of thalassemic pathologies [4].

4. Methods

4.1. System used for data analysis

All calculations were carried out by using an AMD K6 III 400 MHz computer running Windows 98 operating system. We used version 7.2 of Red Hat (LINUX) operating system only for SVM_light.

4.2. PCA

We carried out linear PCA with version 6 of the Simulink MATLAB program. In the approach (not successful), we used all the features (without G6PD and white blood corpuscles, which were not found by normal blood analysis) obtaining a transformation matrix of 196 (training pattern) × 8 (features). Other linear PCA approach did not provide significative results. In the best approach, we did not use PCA but we used only four features (chosen empirically) in the input (RBC, Hb, Ht, and MCV) for the classifiers.

4.3. SVM

SVM classification was performed by using version 4 of SVM_light software [21].

We tried using defaults parameter, and to estimate the quality of the model, we use the leave-one-out (LOO) options. LOO approach was used to gain maximum information from the data; in this method, one pattern is removed from the data set and the model is calibrated with the remaining patterns [10,22,23]. The pattern taken out is used for the evaluation of mean square error (MSE). LOO Q^2 statistic is just the predicted mean square error divided by the sample variance [24,25]. With all the options delineated before, we repeated the classification using a linear, polynomial, and rbf functions for kernel. For each case, we tried many *C* value parameters until the best relative accuracy was obtained, as shown in Fig. 2.

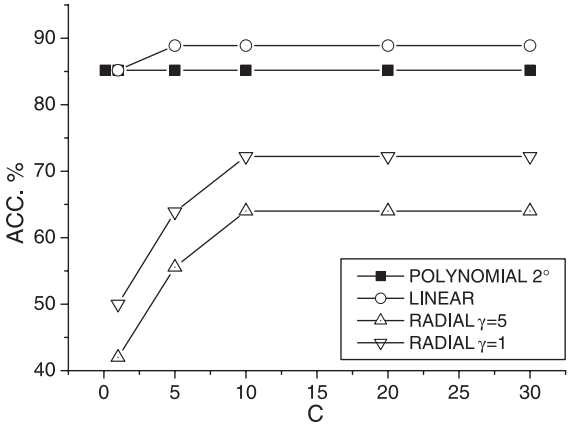

Fig. 2. Accuracy of SVM then *C* variation, in the discrimination between healthy and sick patients. For each kernel type is shown the best performance.

Table 1
Applying KNN for $K=23$ in discriminating between healthy and sick patients

	H	S	OK	BAD	TOT	ACC.
H	92.73^{+5}_{-11}	7.27^{+11}_{-5}	51	4	55	92.73
S	22.64^{+10}_{-14}	77.36^{+14}_{-10}	41	12	53	77.36
			92	16	108	85.19

Rows represent the proportion % of the corresponding row that is classified by classifier in the category of the column. Lower and upper limits of 95% confidence intervals are evaluated through a method described by Wilson with correction for continuity.

4.4. KNN

KNN analysis was performed by using a custom program coded in *C*. The best K ($K'=23$) was found by testing all K values from 1 to 60 on training set. So we used this fixed K' value in the test classification.

5. Results

All the results obtained are shown in terms of sensitivity (percentage of pathologic cases correctly classified), specificity (percentage of non-pathologic cases correctly classified), and with the accuracy (percentage of all cases correctly classified).

5.1. Use of KNN as first layer

Using KNN classifier to differentiate between pathological cases and non-pathological cases allowed us to obtain 85% of accuracy with 93% of specificity (percentage of non-pathological cases correctly classified) and 77% of sensitivity (percentage of pathological cases correctly classified) with $K=23$. It should be noted that KNN is not efficient in discriminating between healthy and sick patients. The results are reported in Table 1 with the confidence intervals evaluated by Wilson method with correction for continuity [26].

Table 2
Applying SVM with linear kernel for $C=10$ and leave-one-out algorithm in discriminating between healthy and sick patients

	H	S	OK	BAD	TOT	ACC.
H	94.55^{+4}_{-11}	5.45^{+11}_{-4}	52	3	55	94.55
S	16.98^{+13}_{-8}	83.02^{+8}_{-13}	44	9	53	83.02
			96	12	108	88.89

Rows represent the proportion % of the corresponding row that is classified by classifier in the category of the column. Lower and upper limits of 95% confidence intervals are evaluated through a method described by Wilson with correction for continuity.

Table 3
Applying SVM with linear kernel for $C=1$ and leave-one-out algorithm in discriminating between healthy and sick patients

	H	S	OK	BAD	TOT	ACC.
H	92.73^{+5}_{-11}	7.27^{+11}_{-5}	51	4	55	92.73
S	15.09^{+13}_{-8}	84.91^{+8}_{-13}	45	8	53	84.91
			96	12	108	88.89

Rows represent the proportion % of the corresponding row that is classified by classifier in the category of the column. Lower and upper limits of 95% confidence intervals are evaluated through a method described by Wilson with correction for continuity.

5.2. Use of SVM as first layer

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 details are shown in Table 2 with the confidence intervals.

We obtained the same results in terms of accuracy using a linear kernel and $C=1$: in this case, sensitivity and specificity were different, as shown in Table 3.

5.3. Use of SVM and KNN as second layer

The discrimination between α and β pathologies is obtained by using a second classification layer specialised on these patterns, which receives, as input, the

Table 4
Application, as second classifier, of the SVM for a linear kernel with $C=0.001$, or equivalently (with the same results) of KNN with $K=23$, in discriminating sick patients (primary SVM with $C=1$) carrying both α and β thalassemia

	α	β	OK	BAD	TOT	ACC.
α	92.5^{+6}_{-14}	7.5^{+14}_{-6}	37	3	40	92.5
β	11.11^{+38}_{-11}	88.89^{+11}_{-38}	8	1	9	88.89
			45	4	49	91.84

Rows represent the proportion % of the corresponding row that is classified by classifier in the category of the column. Lower and upper limits of 95% confidence intervals are evaluated through a method described by Wilson with correction for continuity.

Table 5
Application, as second classifier, of the SVM for a linear kernel with $C=0.001$, or equivalently (with the same results) of KNN with $K=23$, in discriminating sick patients (primary SVM with $C=10$) carrying both α and β thalassemia

	α	β	OK	BAD	TOT	ACC.
α	92.11^{+6}_{-15}	7.89^{+15}_{-6}	35	3	38	92.11
β	11.11^{+38}_{-11}	88.89^{+11}_{-38}	8	1	9	88.89
			43	4	47	91.49

Rows represent the proportion % of the corresponding row that is classified by classifier in the category of the column. Lower and upper limits of 95% confidence intervals are evaluated through a method described by Wilson with correction for continuity.

first classifier output which divides the cases into healthy and sick. As is shown in Tables 1 and 2, the best classification between healthy and sick was with SVM, and so this classifier was used as first classifier. Tests for discriminating α pathologies from β pathologies gave the same results for KNN and SVM classifiers. These are shown in Tables 4 and 5. It should be noted that the small number of clinical records of healthy patients, which were erroneously passed to classifier 2 were considered to be correctly classified when they were associated with type α thalassemia.

5.4. Comparation with neural network

We wished to make a comparative analysis of the performance of SVM and KNN, by analysing their performance versus a combination of three specialised neural networks (MLP) [4]. Such specialised neural network are made each with one output neuron for the discrimination of one class versus the other two and were trained-validated with the cases pertaining to the

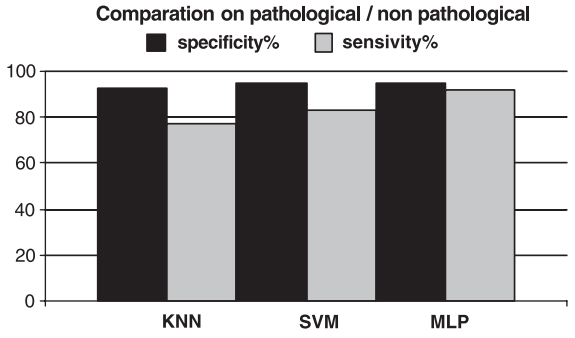

Fig. 3. Comparison between KNN, SVM, MLP on pathological/non pathological case.

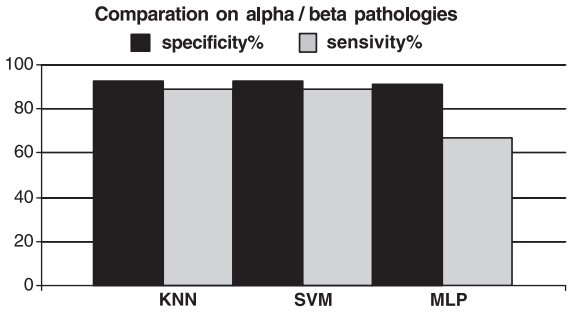

Fig. 4. Comparison between KNN, SVM, MLP on alpha/beta pathologies.

respective output categories; a code to represent the single output allows the combination of the neural networks. The results are reported comparated in Figs. 3 and 4.

In Fig. 3, the discrimination is between healthy and sick patients. The specificity percentages are identical while the sensitivity percentages are different, with the specialised neural network performing better than SVM. Basically, while waiting to compare the results with a larger data set than that available to us, one can provisionally state that the two classifiers are almost equivalent in their ability to discriminate sick from healthy patients.

As far as their ability to recognise type α from type β thalassemia is concerned, the following observations may be made. It can be seen in Fig. 4 that the

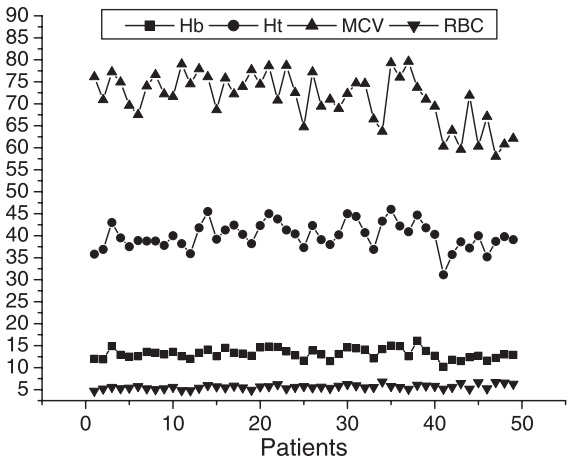

Fig. 5. Values of the RBC, Hb, Ht, MCV parameters for patients classified sick by the primary classifier (SVM). False-positive (for example patient 17) are hidden in the average of the feature values.

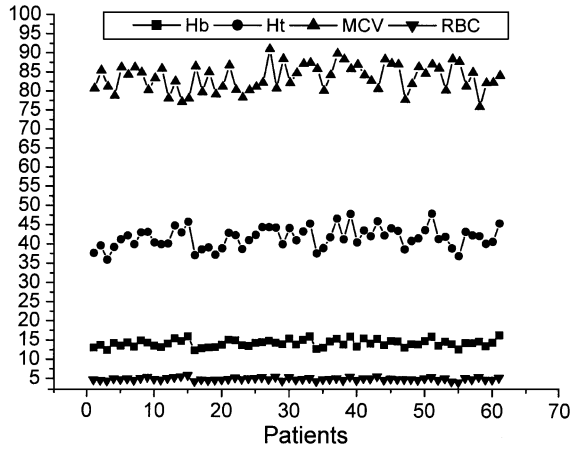

Fig. 6. Values of the RBC, Hb, Ht, MCV parameters for patients classified healthy by the primary classifier SVM. False-negative (for example patient 41) are hidden in the average of the feature values.

SVM classifier performs similarity[1] to the specialised neural network classifier in terms of specificity and is more accurate in the sensitivity then MLP. Once again, one must bear in mind that the results come from a very small data set and must be verified by analysis of more extensive data.

5.5. Complementarity of classifiers

The ability of the classifiers to complement one another in discriminating between sick and healthy subjects was checked to see how much they overlapped. Generally speaking, the SVM classifier and the MLP classifier erred in the same cases. This shows that the two classifiers work in the same field and thus, from the evidence presented above, are not complementary. No multi-classification criteria [27], such as combining the two classifiers, would improve the accuracy of the method.

5.6. A study of the false-positive and false-negative signals

A study on cases where patients where erroneously diagnosed positive has shown that these subjects had the RBC, Hb, Ht, and MCV values in the same range as the pathological cases. In Fig. 5, we show such values for patient classified sick.

The same is true in cases where patients were erroneously diagnosed as negative. Such situation is shown in Fig. 6.

It is difficult to distinguish between thalassemia α cases and healthy subjects using the above-mentioned parameters, while talassemia β cases can be easily distinguished from healthy subjects when these methods are used.

6. Conclusions

The proposed SVM method is as efficient as, and can be compared to, specialised neural networks. The MLP classifier gives slightly better results than SVM although the amount of data available is limited. Both techniques enable thalassemia carriers to be discriminated from healthy subjects with 95% specificity, although the sensitivity of MLP is 92% while that one of SVM is 83%. In addition, the analysis of the results obtained using SVM and MLP networks shows that their specific accuracy and sensitivity cannot be improved using the features studied here.

In the ability to recognise type α from type β thalassemia, the SVM classifier performs similarity to the specialised neural network classifier in terms of specificity and is more accurate in sensitivity than MLP.

Furthermore, it is shown that no multi-classification criteria, such as combining the two classifiers (MLP and SVM), would improve the accuracy of the method.

The results obtained are rather accurate, they should be confirmed by further tests with a larger statistical sample.

Acknowledgements

The author would like to thank the MIUR (Italy) for financial support.

References

[1] F.H. Bunn, B.G. Forget, H.M. Ranney, Hemoglobinopathies, WB Saunders, Philadelphia, PA, 1997.

[1] We can only say that they were similar because, due to the different approaches used in the classification, the percentage in this case refers to a different number of α samples.

[2] Thalassemia Working Party of the British Committee for Standards in Haematology Task Force, Guidelines for investigations of the alpha and beta thalassemia traits, J. Clin. Pathol. 47 (1994) 289–295.
[3] British Committee for standards in Haematology, Guideline: the laboratory diagnosis of haemaoglobinopathies, Br. J. Haematol. 101 (1998) 783–792.
[4] 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 (2002) 18–26.
[5] J. Kittler, F. Roli, Multiple Classifier System, Second International Workshop, MCS 2001 Cambridge, UK, July 2–4, 2001, Proceedings, Lecture Notes in Computer Science, vol. 2096, Springer, Berlin, Heidelberg, New York, 2001, ISBN 3-540-42284-6.
[6] N. Cristianini, J. Shave-Taylor, An Introduction to Support Vector Machine (and Other Kernel-Based Learning Methods), Cambridge Univ. Press, Cambridge, UK, 2000.
[7] V.N. Vapnik, Statistical Learning Theory, Wiley, New York, 1998.
[8] M. Pontil, A. Verri, Properties of support vector machines, Neural Comput. 10 (1998) 955–974.
[9] C.J.C. Burges, A tutorial on support vector machines for pattern recognition, Data Mining and Knowledge Discovery 2 (2) (1998).
[10] T. Joachims, in: B. Scholkopf, C. Burges, A. Smola (Eds.), Making Large-Scale SVM Learning Practical Advances in Kernel Methods—Support Vector Learning, MIT-Press, Cambridge, MA, 1999.
[11] U. Bottigli, Neural networks in medical physics, Phys. Med. 9 (1993) 75–81.
[12] T. Mitchell, Machine Learning, McGraw-Hill, New York, 1997.
[13] T. Joachims, Text categorization with support vector machines: learning with many relevant features, Proc. 10th European Conf. Machine Learning (ECML), Springer-Verlag, Berlin, Heidelberg, New York, 1998.
[14] S. Dumais, et al., Inductive learning algorithms and representations for text categorization, Proc. Conf. Information and Knowledge Management, 1998, submitted for publication.
[15] M.A. Hearst, B. Schölkopf, S. Dumais, E. Osuna, J. Platt, Trends and controversies—support vector machines, IEEE Intell. Syst. 13 (4) (1998) 18–28 (http://www.guppy.mpe.nus.edu.sg/~mpessk/svm/x4018.pdf).
[16] Y. Yao, P. Frasconi, M. Pontil, Fingerprint classification with combinations of support vector machine, 3rd International Conference of Audio- and Video-Based Person Authentication AVBPA (Halmstad, Sweden, June), 2001, pp. 253–259.
[17] B. Schölkopf, C. Burges, A. Smola, Advances in Kernel Methods: Support Vector Machines, MIT Press, Cambridge, MA, 1998 December, ISBN 0-262-19416-3; http://www.guppy.mpe.nus.edu.sg/~mpessk/svm/book_intro.ps.
[18] K. Morik, P. Brokhausen, T. Joachims, Combining statistical learning with a knowledge-based approach—a case study in intensive care monitoring, Proceedings 16th International Conference on Machine Learning, 1999.
[19] K. Fukunaga, Introduction to Statistical Pattern Recognition, 2nd ed., Academic Press, Boston, MA, 1990.
[20] O. Duda, P.E. Hart, Pattern Classification and Scene Analysis, Wiley-Interscience Publication, 1973.
[21] 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.
[22] Mj. Kearns, D. Ron, Algorithmic stability and sanity-check bounds for leave-one-out cross-validation, Proceedings of the Tenth Annual ACM Workshop on Computational Learning Theory, Nashville, Tennessee, 1997 Jul. 6–9, ACM Press, New York, 1997, pp. 152–162.
[23] S.B. Holden, PAC-like upper bounds for the sample complexity of leave-one-out cross validation, Paper Presented at: Ninth Annual ACM Workshop on Computational Learning Theory; Jun. 28; Desenzano del Garda, Italy, 1996.
[24] M. Momma, K.P. Bennett, A pattern search method for model selection of support vector regression, Proc. of SIAM Conference on Data Mining, 2002.
[25] S.S. Keerthi, C.J. Ong, M.M.S. Lee, Two Efficient Methods for Computing Leave-One-Out Error in SVM Algorithms, Technical Report, 2000.
[26] E.B. Wilson, Probable inference, the law of succession, and statistical inference, J. Am. Stat. Assoc. 22 (1927) 209–212.
[27] S. Raudys, Combining the expert networks: a review, in: R. Sadykhov (Ed.), Proc. of Int. Conference Neural Networks and Artificial Intelligence, 2001 2–5 October, pp. 81–91, Minsk, Belarus.