Expert Systems With Applications 86 (2017) 224–234

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Towards an autonomous human chromosome classification system using Competitive Support Vector Machines Teams (CSVMT)

Ali Osman Kusakci[a,*], Berk Ayvaz[a], Elif Karakaya[b]

[a] Department of Industrial Engineering, Istanbul Commerce University, Küçükyalı E5 Kavşağı İnönü Cad. No: 4, Küçükyalı 34840, Istanbul, Turkey
[b] Istanbul Medeniyet University Faculty of Engineering and Natural Sciences, Department of Industrial Engineering 34700 Üsküdar, Istanbul, Turkey

ARTICLE INFO

Article history:
Received 27 October 2016
Revised 12 May 2017
Accepted 27 May 2017
Available online 29 May 2017

Keywords:
Support Vector Machines
Karyotyping
Chromosome classification
Committee machines

ABSTRACT

In broad terms, karyotyping is the process of examination and classification of human chromosome images to diagnose genetic diseases and disorders. It requires time consuming manual examination of cell images by a cytogeneticist to distinguish chromosome classes from each other. Thus, a reliable autonomous human chromosome classification system not only saves time and money but also reduces errors due to the inadequate knowledge level of the expert. Human cell contains 23 pairs of chromosome, 22 autosomes and a pair of sex chromosomes. Hence, we face a multi-class classification task which represents a challenging case for any sort of classifier. In this work, to solve this classification problem, we propose a novel methodology consisting two stages: (i) data preparation and training, and (ii) testing. To determine the most informative content of the dataset several preliminary experiments are conducted and a Principal Component Analysis is done. Then, a single Support Vector Machine (SVM_{ij}) is trained to separate a pair of classes, (i,j) where a numerical optimization method Pattern Search (PS), is employed to find the optimal parameters for the SVM_{ij}. Considering 22 pairs of autosomes, 22×22 experts are trained and optimized. The cluster of experts, we obtain is named as Competitive SVM Teams (CSVMTs) where each SVM_{ij} competes with the others to label a new classification instance. The final output of the classifier is determined by majority voteing. The results obtained on Copenhagen dataset proves the merit of the algorithm as correct classification rates (CRR) on train and test samples are 99.55% and 97.84% respectively, which are higher than any accuracy rate achieved so far in the related literature.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

The cytogenetics analysis process starts with karyotyping, a graphical representation and classification of human chromosomes for diagnostic purposes. This process, in broad terms, is named as Human chromosome analysis and important part of medical diagnostic process of inherited syndromes and degenerations due to environmental factors. Also, recent advancements in genetics, such as Human Artificial Chromosomes (HACs), necessitate automated and precise karyotyping which can ease gene therapy applications. Artificial chromosomes are designated as potential non-viral gene therapy or gene transfer vectors. Among conventional gene transfer techniques such as Bacterial Artificial Chromosomes (BACs), and Yeast Artificial Chromosomes (YACs), Human Artificial Chromosomes (HACs) are recent and promising tools that exhibit several advantages in comparison to the other common approaches (Kazuki et al., 2011). HACs, as candidates of ideal gene transfer vectors, have been studied by several researchers from various aspects (Katoh et al., 2004; Liskovykh, Lee, Larionov, & Kouprina, 2016; Mejıá et al., 2002; Mejía, Willmott, Levy, Earnshaw, & Larin, 2001).

The classification process consists of several steps. Firstly, a tissue, mostly white blood cells, is cultured and the cells are kept in metaphase (Gagula-Palalic & Can, 2014). Secondly, a cytogeneticist analyzes the prepared slides for metaphase cells with microscope. When a good cell with least overlapping chromosomes is identified, it is digitized with a camera. Then, the chromosomes are counted automatically by a computer or manually by cytogeneticist. Next, karyotyping is done where the chromosomes are separated into classes (Gagula-Palalic & Can, 2014). This process is usually conducted manually by an expert, cytogeneticist. The product of this process is called karyogram. Finally, analysis of the karyogram is done for diagnosis.

The most well-established visualization techniques in Cytogenetics rely on the idea of producing grey bands by Giemsa staining after pre-treating the chromosomes with trypsin, G-banding (Legrand, Chang, Ong, Neo, & Palanisamy, 2008). This allows the classification of the chromosomes by a unique banding pattern

* Corresponding author.
E-mail addresses: aliosmankusakci@gmail.com, aokusakci@ticaret.edu.tr (A.O. Kusakci), bayvaz@ticaret.edu.tr (B. Ayvaz), elif.karakaya@medeniyet.edu.tr (E. Karakaya).

http://dx.doi.org/10.1016/j.eswa.2017.05.070
0957-4174/© 2017 Elsevier Ltd. All rights reserved.

specific to each type. Other dominant characteristics exploited by almost all research studies are features related to length and centromere of chromosome, a constricted region lying on the chromosome responsible for joining two copies of a replicated chromosome and accurate inheritance of eukaryotic chromosomes during cell division (Kazuki et al., 2011; Legrand et al., 2008; Ritter & Gao, 2008).

In conventional cytogenetics analysis, chromosome classification is usually done by experts of the field at the expense of time consuming and tedious inspections (Lundsteen, Gerdes, & Maahr, 1986). To ease this process, development of reliable classification systems for automatic chromosome karyotyping is of paramount importance (Wang et al., 2005). Such a system requires addressing two issues: (i) defining and extracting the right features to be used for classification (Moradi & Setarehdan, 2006; Ritter & Schreib, 2001) and (ii) building very advanced and complex classification algorithms with high reliability.

Considering that each human cell normally contains 23 pairs of chromosomes, 22 pairs of chromosomes and 1 pair of sex chromosome, chromosome classification is a difficult multi-class problem (MCP) compared to one-against-one (OAO) classification tasks (Subasi, 2012). Thus, development of novel algorithms is a must since classical solution approaches may fail to provide sufficient results for MCPs. One of the recent solution approaches to this kind of problems is building an ensemble of several classifiers. Proven by many studies in machine learning field, a good ensemble can positively affect the performance of a stand-alone classifier (Nanni & Lumini, 2009a).

Focusing to the issues discussed above, this work proposes a novel solution approach to the chromosome classification problem. We can summarize the contributions of this study as follows. First, a comprehensive analysis is conducted with the aim of extraction of the right features as the performance of a classifier can be directly improved by correct problem representation and feature selection. Second, a novel ensemble of competitive classifiers, Competitive Support Vector Machine Teams (CSVMTs), is proposed to handle chromosome classification problem as an MCP. Third, the proposed classifier is trained and tested on a well-known dataset (Piper & Granum, 1989) where the merit of the classifier is proved considering its performance in terms of classification error and robustness.

The remainder of this paper is organized as follows: in Section 2, related work is summarized. Section 3 is devoted to theoretical background and methodology where the basics of Support Vector Machines (SVM), used dataset and preliminary experiments are presented. The details of the proposed model are given in Section 4 while the results of the study will be demonstrated and discussed in Section 5. Lastly, Section 6 concludes the study and summarizes the findings.

2. Related work

Recent advances in machine learning ease the ways medical services are delivered in various application areas ranging from terminology extraction from online community generated medical records (Nie, Wang, et al., 2015; Nie, Zhao, Akbari, Shen, & Chua, 2015) to disease prediction and progression modelling (Åström & Koker, 2011; Nie, Zhang, et al., 2015). Also, considerable amount of research attempts have been dedicated to automated karyotyping problem using various machine learning tools. The large portion of the research addressed classification problem (Kao, Chuang, & Wang, 2008; Oskouei & Shanbehzadeh, 2010) while some studies give emphasis to feature extraction stage (Piper & Granum, 1989; Ritter & Schreib, 2001).

Ruan (1999) proposes a dynamic neuro-fuzzy system. It provides a novel implementation which is a combination of artificial neural network (ANN) and fuzzy logic. In his next study, the author improves the previous classifier to identify and categorize human chromosomes automatically (Ruan, 2000). Delshadpour (2003) also uses ANN for the same purpose. In his article 304-chromosome samples for 24 classes are used and 88% accuracy is obtained.

Sampat, Bovik, Aggarwal, and Castleman (2005) applies three methods which are Maximum likelihood, nearest neighbor and k-nearest neighbor method on the same dataset. After a quite few experimentations, k-nearest neighbor method is chosen as the most accurate classifier and the obtained accuracy rate is 90.52%.

Nanni (2006) points out the ensemble of classifiers in his work where he utilizes a self-acting karyotyping method and then examines the results of real dataset scenarios. In the article of Nanni and Lumini (2009a), many ensemble classifiers are investigated in attempt to find the best ensemble example which can maximize the performance. For this purpose, different combinations of ensembles are carried out. In the same year, they improve a prototype generation method using genetic algorithm principles to reduce the complexity of feature plane classifier (Nanni & Lumini, 2009b).

To carry out chromosome classification, some new classifier examples can be seen in the literature as well. One of them is adapted dynamic time warping (DTW) (Legrand et al., 2008). They proof the efficiency of the original classifier by attaining the accuracy rate of 88% whereas the classification accuracy rate of Bayesian Classifier is about 81%.

Wang et al. (2009) use both a two-layer classification platform and ANN. 92.9% classification accuracy is obtained in the first layer, whereas for the second layer, classification accuracy is ranging from 67.5% to 97.5%. In this paper, they highlight two articles as a framework in which different group of wavelet transforms are applied (Wang, 1999) (Wang & Lee, 1998).

Ventura, Khmelinskii, and Sanches (2010) carried out many experimentations using various pairing set of chromosomes. After applying structural band pattern techniques based on Bayesian method, the obtained average pairing rate is 92.8%.

Poletti, Ruggeri, and Grisan (2011) improve the functionality of chromosome classification when facing the missing data problem. To figure out this difficulty, Monte Carlo methodology is used. In their next study, Poletti, Grisan, and Ruggeri (2012), a new modular framework is proposed and with the help of this method 94% of accuracy rate is achieved.

Feng, Cong, Zhu, and Du (2012) present a computer aided pairing of metaphase chromosomes by utilizing gradient profile and similarity matching algorithm. After performing the combined method, accuracy rate of males is 91.3% while the same ratio is 100% for female chromosomes.

Gagula-Palalic and Can (2012) draw our attention to automatic segmentation which is an essential step towards an autonomous karyotyping machine. In this paper, many arbitrary points are generated to be used as search centers to initiate the segmentation process. Gagula-Palalic and Can (2014) present a method which is an ensemble of ANN and nearest neighbor classifiers to human chromosome classification. With this new method, the obtained classification rate is about 95%.

Another significant study using ANN is presented by Oskouei and Shanbehzadeh (2010). They propose the wavelet neural networks (WNN) which is able to perform both manually and automatically. According to their analyses, WNN has shown better performance than standard ANN with an overall accuracy rate of 93%.

Sjahputera and Keller (1999) deal with fuzzy rule-base evolution in order to optimize performance of chromosome classifier. Dynamic image segmentation is investigated by Halder and Pramanik (2012) by using Fuzzy Hopfield Neural Network (Mao & Liu, 2004; Takahashi, 1997). In this paper, Genetic Algorithm is utilized to optimize the structural parameters of the classifier.

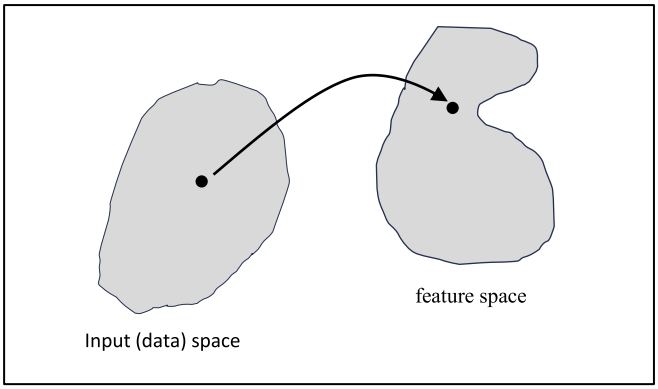


Fig. 1. Step 1: Mapping of input space to feature space (Haykin, 1999).

Considering success rates reported above, there is still some space for improvement in the field of autonomous karyotyping. Additionally, as medical diagnosis is a challenging task which requires employing very precise classification algorithms, the mentioned accuracy values are not truly satisfying. Thus, designing more powerful algorithms is extremely essential so that an autonomous karyotyping can be put into practice with minimum bearable classification error.

3. Background and methodology

3.1. Support vector machines

As an efficient classification tool, Support Vector Machines (SVMs) have gained remarkable importance in the late decade due to its specific properties favoring them. SVM is pioneered by Vapnik (2000) and based on statistical learning theory. Since its first introduction, it has been more frequently applied for pattern classification and nonlinear regression purposes (Nie, Zhao, et al., 2015; Subasi, 2013).

In the case of fully separable patterns, the main idea of SVM is to construct a hyperplane representing the decision surface such that the margin of separator, also called decision boundary, between the two classes is maximized. Practically, this boundary is a strip with a maximum width between the classes. The width of the strip is determined based on the samples lying close to the boundary of the two classes to be separated. These samples are called support vectors where the name of the algorithm is derived from. Hence, support vectors are those data points placed closest to the boundary and, therefore, they are hardly separable (Haykin, 1999). SVM performs two main mathematical operations on the dataset while separating the classes.

1. Nonlinear mapping of the sample of input space into a high-dimensional hidden feature space in which the data can be separated linearly, see Fig. 1.
2. Construction of a hyperplane for separating data points projected another feature space obtained in step 1 (Haykin, 1999). This stage is depicted in Fig. 2 for a linearly separable problem. The shape of the hyperplane is determined by solving the dual form of a quadratic programming problem which will be formulated later in this section.

We briefly state mathematical background of SVM in the remainder of this section.

Given a training set for a SVM, $P = \{(x_i,\ y_i)|x_i \in \mathbb{R}^k, y_i \in \{-1, 1\}\}_{i=1}^{N}$, where x_i stands for the n-dimensional input vector and y_i is output vector. Remembering that SVM tries to build a decision boundary with the widest margin between two classes, for linearly non-separable samples, a transfer to a feature space with higher dimension is conducted as follows: $\phi(x) : x_i \in \mathbb{R}^k \rightarrow \mathbb{R}^m,\ k \ll m$.

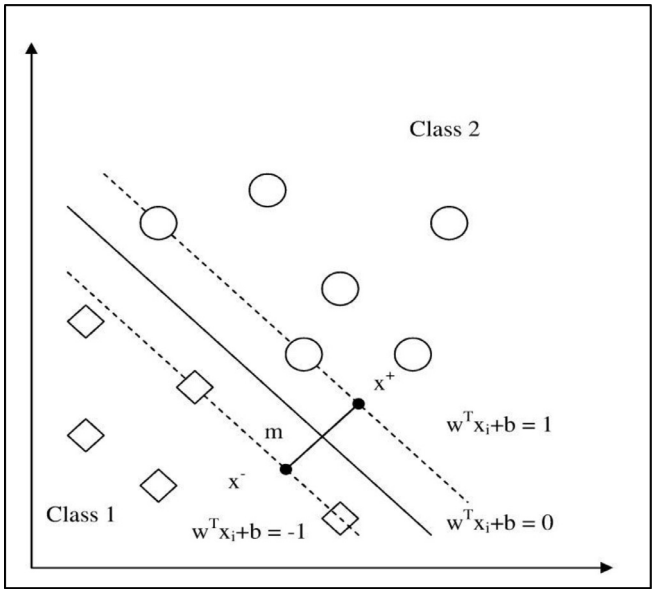


Fig. 2. Linear support vector classifier (Subasi & Gursoy, 2010).

The transformation process is realized with the help of so called transfer functions. Vapnik (2000) suggested that the transfer function should be selected from a family of universal approximation functions with linearly separable character. By this way, a linear decision surface is constructed by the transfer function in the feature space. However, the boundary in the original input space may or may not be separable. Mathematically, in case of separation of two classes, we can formulate this as follows:

$$\begin{aligned} f(x_i) &= w^T\phi(x_i) + b \geq 1 \quad \forall\ i : y_i = 1 \\ f(x_i) &= w^T\phi(x_i) + b \leq -1 \quad \forall\ i : y_i = -1 \end{aligned} \tag{1}$$

Here, given the training set P, the aim is to find the optimal values of w and b to construct a hyperplane, maximizing the margin of separation. The formula in Eq. (1) is illustrated in Fig. 2.

The maximization of the margin is realized by minimizing the Euclidian norm of the weight vector w. For complex problems of imperfect separable character, the constraints in Eq. (1) can be replaced with a soft margin using slack variables ε_i. More precisely, for non-separable case, this formula allows that some data points may lie inside the margin area and the sum of constraint violations is incorporated into the objective function. In mathematical terms, the optimization problem can be formulated as:

$$\begin{aligned} &\text{Min} \\ &\psi(w, \varepsilon) = 1/2 w^T w + C \textstyle\sum_{i=1}^{N} \varepsilon_i \end{aligned} \tag{2}$$

Subject to:

$$\begin{aligned} y_i\left(w^T x_i + b\right) &\geq 1 - \varepsilon_i \text{ for } i = 1, \ldots, N \\ \varepsilon_i &\geq 0 \text{ for all } i \end{aligned} \tag{3}$$

where C is a user defined penalty parameter controlling the trade-off between the complexity of the classification machine and the number of non-separable points (Haykin, 1999).

The inner product of two vectors induced in the feature space, $\phi(x)$ and $\phi(x_i)$, defines a kernel function, as formulated in Eq. (4), which constructs a decision surface that is nonlinear in the input space, but its image is linear in the feature space (Haykin, 1999).

$$K(x, x_i) = \phi(x)^T \phi(x_i) \tag{4}$$

Gaussian, Radial Basis, and Polynomial kernels are most frequently preferred kernel functions in the literature.

Instead of solving the optimization problem stated above, practically, it is simpler to optimize its dual form. With a specific kernel function $K(.,.)$, the optimization problem in the dual form can be formulated as follows:

Given the training set $\{(x_i, y_i)\}_{i=1}^{N}$, find the Lagrange multipliers $\{\alpha_i\}_{i=1}^{N}$ that maximize the objective function

$$\text{Max} \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j K(x_i, x_j), \quad (5)$$

Subject to the constraints:

$$\sum_{i=1}^{N} \alpha_i d_i = 0 \quad 0 \le \alpha_i \le C \text{ for } i = 1, \ldots, N. \quad (6)$$

When a radial basis function (RBF) type kernels are used, the number and the center of the RBFs are determined by the number and the values of the support vectors. However, the optimal value of the width of the RBFs, σ^2, has to be specified by the user. Hence, while designing an efficient SVM with RBFs as kernels, two parameters are of paramount importance, (C, σ^2), the penalty parameter and the width of the RFB, respectively (Subasi, 2013). In the formulation given in Eq. (6), C is also referred as box constraint parameter imposing an explicit upper bound to the constraints.

Basically, SVM classifiers are employed for one-against-one (OAO) classification problems where two classes are to be separated from each other. This sort of classification is also known as binary classification problems (Kou, Ji, & Zhang, 2002). Thus, SVMs are not originally suitable tools for multi-class problems as in the case of the chromosome sorting. However, two main strategies are common to overcome this burden (Kou et al., 2002):

- One to others: train a team of SVM experts where each is responsible to distinguish one chromosome type from the others,
- Pairwise: train a team of SVM experts where each of them is responsible to separate two classes from each other.

3.2. Dataset

Researchers have proposed several algorithms tested on various chromosome classification test sets, such as Copenhagen, Edinburgh and Philadelphia, which have been constructed based on human cell images published over many years by Clinical Centers and Hospitals (Piper & Granum, 1989). The Copenhagen data base was commonly preferred for experimental purposes in the related literature. Hence, it became a suitable benchmark for performance comparisons. However, we should note that variations on the experimental setup, such as grouping the chromosomes in less number of classes or focusing a group of chromosomes with certain characteristics, may not allow making a sound comparison between scientific works although they reportedly use the same dataset.

Due to its wide use, the proposed methodology in this work is applied also to Copenhagen dataset with 4400 samples, namely, 200 samples for each chromosome type.

In the original dataset, following features are provided (Lundsteen, Philip, & Granum, 2008):

- A unique chromosome identifier (N).
- A secondary index indicating the metaphase that the sample came from (M).
- The third dimension refers to the chromosome type (CT).
- Length of chromosome (L) – expressed as the number of pixels on the longitudinal axis of a chromosome.
- Length of the short arm (p-arm) (PA): It is expressed as the number of pixels between the top of chromosome and the centromere on the longitudinal axis, depicted in Fig. 3.

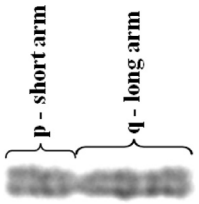


Fig. 3. G-banded grey image of a human chromosome.

Table 1
Statistical ranges of the 200 samples of each chromosome type.

Type	Min (L)	Max (L)	Mean (L)	Min (CI)	Max (CI)	Mean (CI)
1	70,0	106,0	85,6	33,0	52,0	41,7
2	70,0	98,0	82,2	28,0	40,0	32,5
3	59,0	82,0	68,9	20,0	39,0	32,3
4	57,0	74,0	65,0	14,0	23,0	19,3
5	56,0	71,0	63,1	15,0	22,0	18,8
6	53,0	72,0	61,3	18,0	30,0	24,1
7	50,0	64,0	56,1	18,0	26,0	22,0
8	46,0	58,0	51,4	15,0	22,0	18,2
9	44,0	56,0	49,3	12,0	24,0	17,6
10	44,0	55,0	49,0	13,0	21,0	16,8
11	44,0	55,0	49,2	16,0	24,0	19,5
12	44,0	54,0	48,9	10,0	19,0	14,8
13	38,0	46,0	41,6	4,0	13,0	8,0
14	37,0	45,0	40,6	2,0	13,0	9,0
15	36,0	43,0	39,4	5,0	14,0	9,6
16	33,0	38,0	35,5	12,0	17,0	14,9
17	32,0	38,0	35,0	9,0	14,0	12,0
18	30,0	36,0	32,7	7,0	13,0	10,3
19	26,0	32,0	28,8	9,0	15,0	12,5
20	26,0	32,0	28,5	8,0	15,0	12,5
21	21,0	26,0	23,4	4,0	11,0	7,9
22	23,0	28,0	25,6	4,0	13,0	8,5

- The band pattern features: The samples are extracted from G-banded grey images. For a G-banded chromosome, a band is defined as an interval on a chromosome clearly distinguishable from neighboring regions by the intensity of staining (Granum & Thomason, 1990). To produce the required visual material, usually, a light microscope is used to photograph cell nuclei during metaphase stage at which the chromatin is condensed inside the chromosomes. This enables us to easily detect their bands with a light microscope. Given a raw greyscale image of a chromosome, the band pattern profile is extracted illustrating the grey level of the bands on the chromosome. There are two main approaches for expressing the band profiles. One way is to use strings 1, 2, 3, 4, 5, 6 for each incremental band, called as a string representation. Alternatively, difference string representation may be employed where the transition relation between two adjacent bands is coded with a letter alphabet (Lundsteen et al., 2008).

In this study, for practical purposes, we will work with string representation using an alphabet of six-numbers {1, 2, 3, 4, 5, 6} where each of them shows directly incremental intensity of a band. For the longest chromosome in the database, the density profile of grey scale images is represented with 105 band regions; whereas shorter samples are digitized with the least number of band pattern features and the remaining positions are padded with zeros so that we have all instances of the same length.

Some descriptive statistics for the 200 samples of each chromosome type are given in Table 1.

3.3. Preliminary experiment

As mentioned earlier, two features extracted from G-banded images are almost universal in all chromosome classification

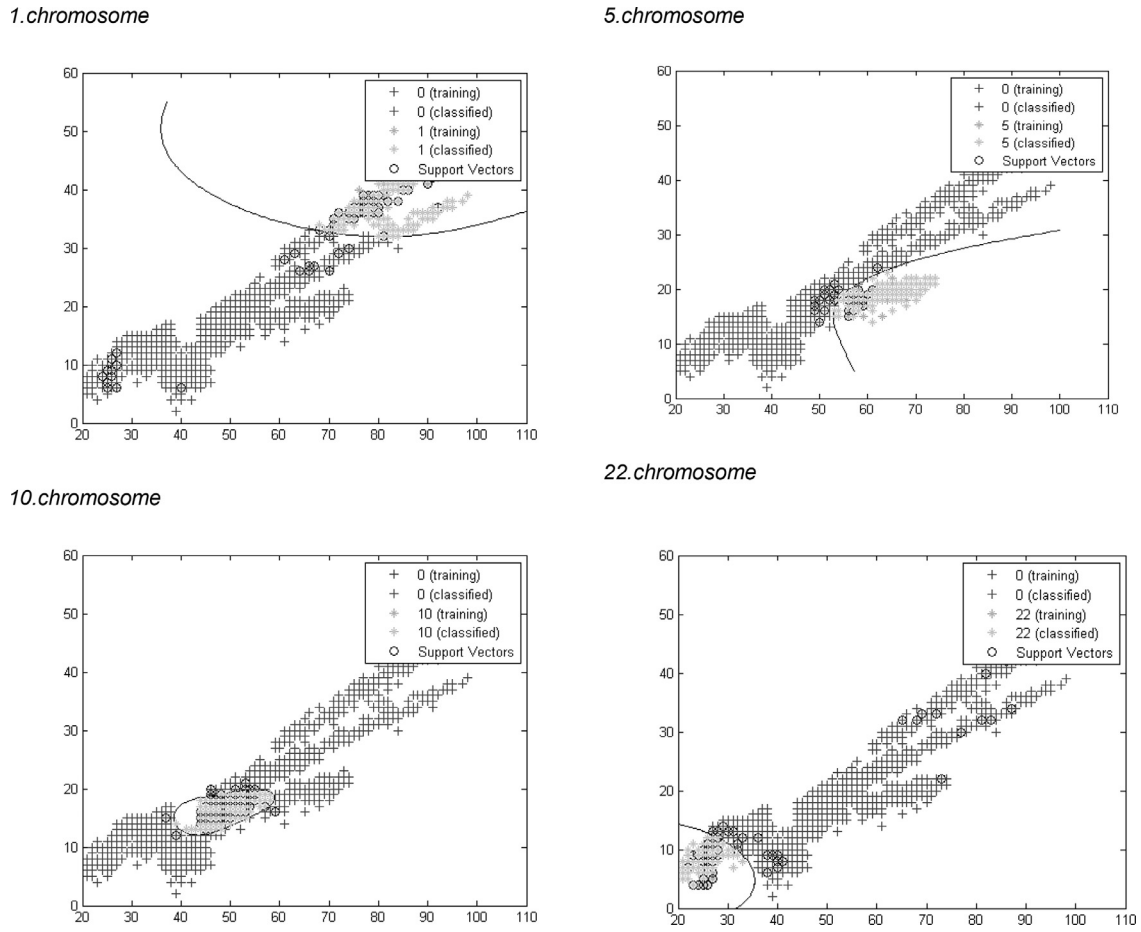


Fig. 4. Separation margins obtained in preliminary experiments for some chromosome classes.

manuscripts in the literature, the length (L) and the length of the short arm (PA) of chromosome (Granum & Thomason, 1990). The relevance of these two variables is investigated with some preliminary experiments we conducted. Initially, for the purpose of simplicity, we design a one to others classifier where 22 distinct SVM experts are trained to separate one chromosome class from the others. The dataset is divided into two parts where randomly selected 80% is used for training while the remaining 20% is put aside for testing purposes.

Fig. 4 illustrates the results of the preliminary experiments using L and PA as input variables, where separating margin and support vectors are shown in chromosome classes 1, 5, 10, and 22. More details are provided in Table 2 where the correct classification rate (CCR), error rate (ER), Sensitivity, Specificity, true negative, false negative, true positive, and false positive values are shown (Kusakci, 2012). According to the table, overall correct classification is 91%. Especially, true positive rate of 98.91% is noteworthy. Namely, classifier is very successful in identifying a chromosome's real class. On the other hand, the SVM predicts some "zero" labeled chromosomes are predicted as "m" indicating that these chromosomes belong to the considered chromosome type m. Thus, we obtained lower sensitivity rates when compared with specificity rates.

Considering the reported values, we may conclude that L and PA are highly distinctive indicators and the proposed SVM team is very promising classifier for chromosome classification problem. Thus, we will devote more efforts on (i) adding more input features, such as use of band pattern features, (ii) designing more complex classifiers exploiting the strengths which were implied by the results of the first experiment, and (iii) hybridizing SVM with an effective optimization method so that design parameters of SVM are optimized. In the remainder of this work, we investigate these issues and accordingly design a classifier with increased performance.

4. The proposed algorithm

Considering SVM's accuracy on binary classification problems, it can be trained to distinguish any pair of classes. Namely, we may construct an ensemble of intelligent machines or experts having different expertise area with pairwise training. Then, the knowledge acquired by the expert teams can be fused to derive a universal decision (Kusakci, 2012).

However, this approach requires a well-structured strategy while combining outputs of individual experts and, finally, providing a single reliable answer. This section summarizes the proposed methodology where the most attention will be devoted to the main steps of the algorithm and to the combination strategy of responses of several SVM experts trained on different chromosome pairs.

Let us define a Support Vector Machine, $SVM_{i,j}$, where $i, j = 1, 2, \ldots, 21, 22$, and $i \neq j$. More precisely, $SVM_{i,j}$ is assigned to distinguish classes of chromosome type i and type j, creating the output i, if new data is of type i, and creates the output j, if a new sample is of type j. Following the same idea for each pair of distinct chromosome type, the overall structure of the team classifiers can be represented as a 22×22 matrix as seen in Fig. 5. Clearly, no SVM is assigned on the diagonal axis where $i = j$.

Table 2
The results of the preliminary experiment.

i	CCR	ER	Sensitivity	Specificity	True neg.	False neg.	False pos.	True pos.
1	98.05	1.95	0.98	1	2057	43	0	100
2	98.68	1.32	0.99	1	2071	29	0	100
3	96.23	3.77	0.96	0.98	2019	81	2	98
4	94.41	5.59	0.94	1	1977	123	0	100
5	93.55	6.45	0.93	1	1958	142	0	100
6	90.77	9.23	0.9	0.98	1899	201	2	98
7	88	12	0.87	1	1836	264	0	100
8	81.59	18.41	0.81	0.99	1696	404	1	99
9	79.86	20.14	0.79	0.99	1658	442	1	99
10	83.32	16.68	0.83	0.98	1735	365	2	98
11	88.05	11.96	0.88	0.98	1839	261	2	98
12	85.82	14.18	0.85	0.98	1790	310	2	98
13	92.18	7.82	0.92	0.98	1930	170	2	98
14	89.55	10.46	0.89	0.99	1871	229	1	99
15	89.73	10.27	0.89	0.99	1875	225	1	99
16	95.5	4.5	0.95	0.99	2002	98	1	99
17	91.14	8.86	0.91	0.98	1907	193	2	98
18	93.64	6.36	0.93	1	1960	140	0	100
19	85.73	14.27	0.85	0.97	1789	311	3	97
20	94.64	5.36	0.94	0.98	1984	116	2	98
21	96.36	3.64	0.96	1	2020	80	0	100
22	93.55	6.45	0.93	1	1958	142	0	100
Mean	91	9	0.91	0.99	1901.41	198.59	1.09	98.91

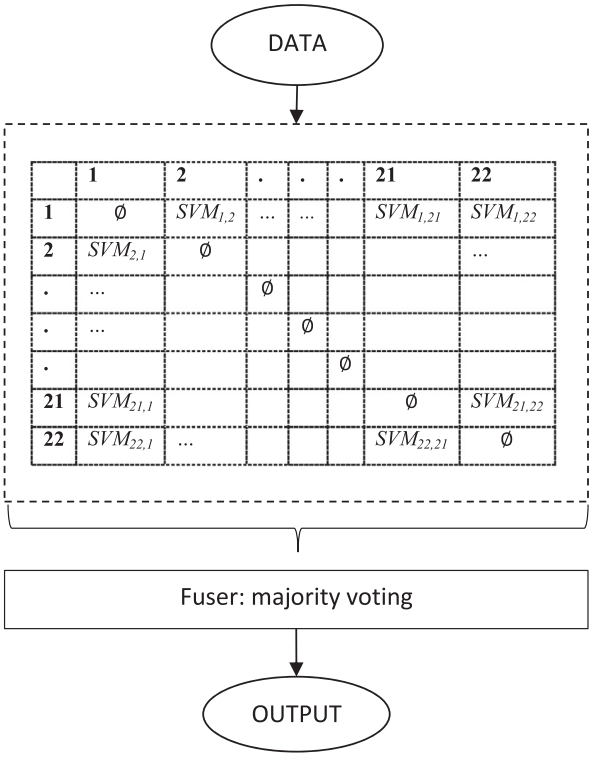


Fig. 5. Overall structure of the proposed competitive SVM teams.

After the training stage, assume a new assignment has to be made for a chromosome of unknown type. Theoretically, if the new data belong to a chromosome of type *k*, then, it is expected that all SVMs in the *m*th column, $SVM_{j,m}$ $j = 1, 2, \ldots, 21, 22$ will give an output of *m*. On the other hand, we expect that all SVMs in the *m*th row $SVM_{m,i}$, $i = 1, 2, \ldots, 21, 22$ provide an output of *m*. The entries in this matrix represent the votes of each SVM assigned to each pair of chromosome types.

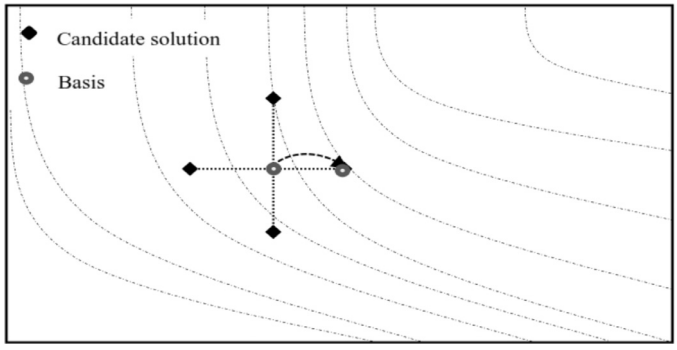


Fig. 6. Pattern search in 2-D.

4.1. Optimizing parameters of SVM: pattern search

For each $SVM_{i,j}$, the optimal values of the couple, (C, σ^2), have to be determined. For this purpose, a numerical optimization method, called pattern search (PS) is applied. PS is a direction-based deterministic search algorithm which does not require any derivative information of the objective function surface. The main steps of all PS algorithms consist of the following; (i) construction of mesh of points around the current solution, and (ii) refining the mesh and repeating the process if the current solution remained unimproved (Kajee-Bagdadi, 2007).

More precisely, for an optimization problem of n-dimensions, the algorithm maintains the current best solution as a basis and generates 2n candidate solutions around this basis point. If a better solution in terms of the objective function value is obtained, a greedy replacement scheme applies and the basis is updated. Additionally, the step size in each direction is expanded by a certain factor. If the search is not successful, that is, no better solution could be found, the step size is decreased, and another set of $2n$ points are produced (Torczon, 1997) as depicted in Fig. 6.

4.2. Adding band pattern features

As stated before, we randomly selected 80% of the data for training and the rest is kept for testing. Additional to L and PA of a chromosome, we exploit the band pattern features obtained

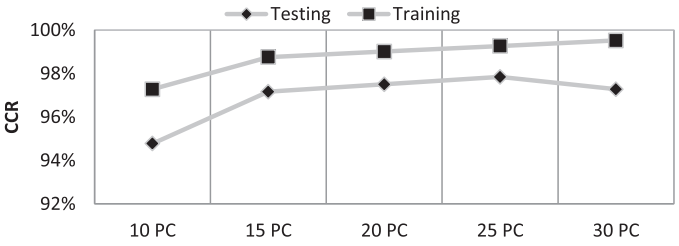


Fig. 7. Effect of number principal components on correct classification rate (CCR).

from digitized images. Although SVMs are known with their robust performance on high dimensional problems (Kou et al., 2002), the curse of dimensionality rule applies also for SVMs as our preliminary experiments have shown that adding all 105 band pattern features increases training time enormously. Thus, the dimension is reduced with Principal Component Analysis (PCA) by conducting several trials where number of principal components has been gradually increased from 10 to 30. Another dimension contributing to the success of the algorithm is observed as the ratio of the length of the short arm to the other arm, so the called Centromere Index (CI). The centromere holds together the couple of chromatids before their separation during the mitotic cell division (Legrand et al., 2008). It is one of the most common features reported as a good characterization metric in karyotyping.

The results of the PCA are shown in Fig. 7 which indicates a slight decrease in testing performance if the algorithm is trained with 30 PCA. Accordingly, we decided to employ 28 input features in total, more precisely L, PA, CI, and 25 principal components obtained through PCA of band pattern features.

4.3. Competitive SVM teams: CSVMTs

The overall steps of the proposed algorithm is summarized below and illustrated in Fig. 8. At the training phase, four successive stages should be completed:

1. **Preprocessing the data:** Initially, the dataset is divided into two parts, training and testing sets. Next, PCA is applied on both parts so that the number of input features is reduced and the most informative content is extracted from the dataset.
2. **Optimizing (C, σ^2):** At this stage, the best (C, σ^2) pair for each $SVM_{i,j}$ dedicated to distinguish the classes $\{i, j\}$ are determined by pattern search (PS). A 10-fold cross validation is used and the (C, σ^2), which minimizes the average classification error over 10-fold cross validation, is taken as the best combination of two parameters specified for each $SVM_{i,j}$ expert. By default, PS doubles the mesh size after each successful poll and halves it after each unsuccessful poll.
3. **Learning the training data:** The optimized values are then employed for constructing the SVM expert, and the trained machine is labeled as $SVM_{i,j}$ which is assigned to the task of separating the chromosome i from j.
4. **Building Competitive SVM Teams:** Following the same approach given in step 3, we train in total $22*21 = 462$ SVM experts responsible for binary classification of all possible couples of $\{i, j\}$. Here, we should note that there is no expert on the diagonal axis.

The testing phase consists of two consecutive stages:

1. **Feeding a test sample:** An unsorted test sample is fed to the CSVMTs classifier where each expert produces either i or j as output. Next, the obtained results are recorded.
2. **Majority voting:** The responses of all experts are combined by majority voting. More precisely, assuming that we have a test sample originally extracted from a chromosome type m, each expert evaluates the newly introduced data point based on the separating hyperplane learnt during the training stage and assigns accordingly a chromosome class to the sample. Ideally, we expect that the perfect classifiers $SVM_{m,j}$ and $SVM_{i,m}$ estimate correctly the type of the chromosome as m for any i and j value. Table 3 visualizes the response matrix of SVM experts responsible for differentiating chrome type m from the others in this ideal case. Obviously, remembering there is no clue that the sample data is of type m beforehand, other experts also provide their own responses as we simultaneously request an answer from all 462 classifiers in the team. However, $SVM_{i,j}$, $i \neq m$ and $j \neq m$ which are responsible to distinguish other types of chromosomes, are expected to demonstrate an inferior performance. Thus, an aggregated output of CSVMTs classifier is generated based on the majority voting approach.

Table 3
The ideal response matrix.

	1	2	.	.	m-1	m	m+1	.	22
1	-					m			
2		-				m			
.				-		m			
.						m			
m-1					-	m			
m	m	m	m	m	m	-	m	m	m
m+1						m	-		
.						m		-	
22						m			-

In brief, a team of competing classifiers is implemented with SVMs in this work. That is, a competition will take place between the teams of SVMs trained for a target class. Furthermore, without losing generality, there are several studies from various scientific fields, which have reportedly proved the merit of SVMs when dealing with binary classification problems (Bhasin & Raghava, 2004; Chen, Hsu, & Shen, 2005). The relevance of this statement is especially high if the parameters of SVM are carefully tuned. For this purpose, a heuristic optimization method is hybridized with SVM in this study.

5. Results and discussion

As stated before, out of 4400 data samples 3520 (80%) instances are used for training, whereas the rest are kept for testing purposes. During the training stage a 10-fold cross validation strategy along with Pattern Search algorithm for optimization of design parameters is implemented.

To demonstrate the robustness of the algorithm under random effects, such as random portioning of the dataset for training and testing, 30 trials have been run. Before we introduce the performance measures used, we note that by the best trial, we mean the trial with the maximum average correct classification over 22 classes, whereas the worst trial indicates the trial with minimum overall correct classification ratio.

The following indicators are employed:

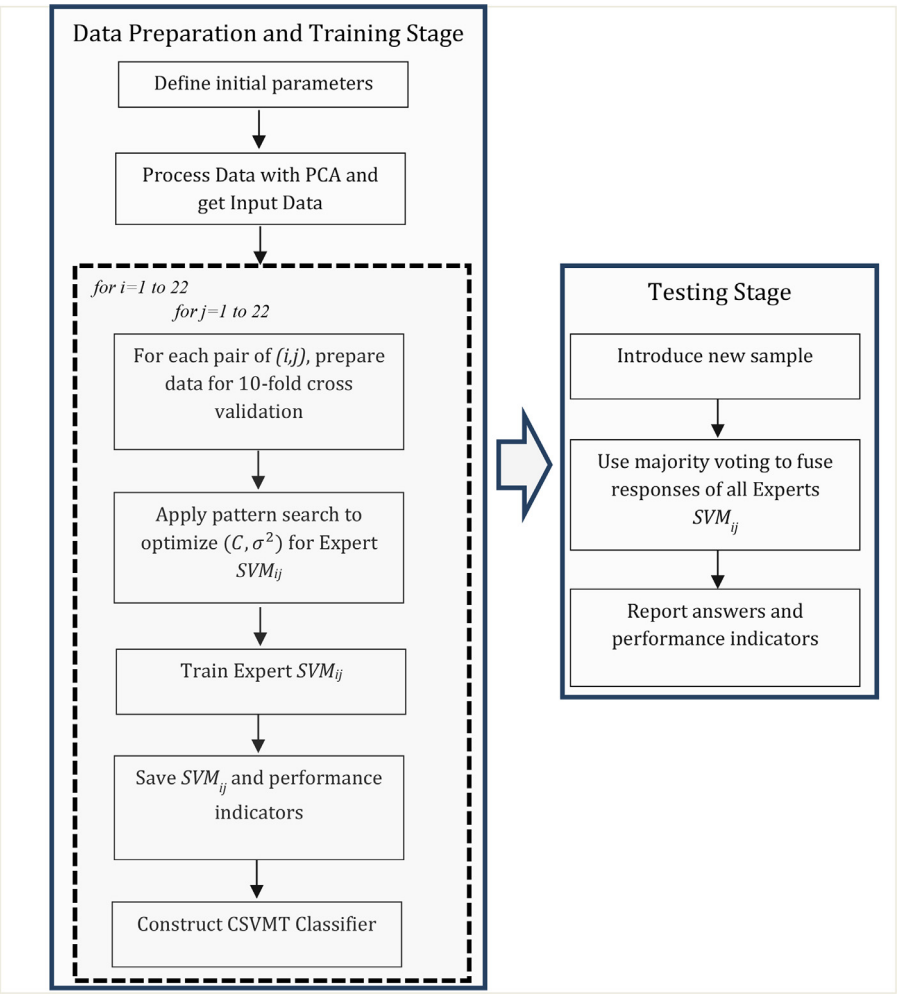


Fig. 8. Main steps of the proposed algorithm.

- $\bar{P}$: Overall correct classification ratio over 22 chromosome types
- $\overline{P_i}$: Average correct classification ratio of chromosome class i
- s_i: Standard deviation of correct classification ratio of chromosome class i
- $\max(P_i)$: Maximum correct classification ratio of chromosome class i.
- $\min(P_i)$: Minimum correct classification ratio of chromosome class i.
- P_i^{best}: Correct classification ratio of chromosome class i at the best trial.
- P_i^{worst}: Correct classification ratio of chromosome class i at the worst trial.

The performance measures obtained after 30 trials on training and testing data are summarized in Tables 4 and 5. The maximum average CCR that the classifier could reach is 99.55% on training data while the algorithm was able to identify 97.84% of the test samples correctly. Also, the average CCR values at the worst trials in training and testing stages are highly competitive, 98.86% and 96.93% respectively. Over 30 trials, the average CCRs we obtained are 99.25% and 97.59% in training and testing. Another remarkable result, we need to highlight is the low standard deviation values, 0.16% and 0.28% on average on training and testing data respectively. This fact is also observed in Figs. 9 and 10 where average CCRs plus/minus one standard deviation ranges of each chromosome class are illustrated. Considering each distinct class, average CCRs over 30 trials are almost perfect for all classes except 19th, 20th, and 22nd. Despite that the values given for these three classes are slightly worse than the average accuracy rates for others, the obtained results are mostly superior to the ones reported in other state-of-the-art studies we have referred. Furthermore, these chromosome classes are also evaluated as difficult-to-identify chromosome types in the literature due to their relatively small sizes (Legrand et al., 2008).

To give some insights which classes are mostly confused with each other, the confusion matrix obtained at the best trial on test data illustrates the error pattern in Table 6 where light grey cells on the diagonal axis indicate correct classification and the darker ones refer to sporadic mistakes which count only for 2.16% of all test samples. A closer look at the matrix reveals that the proposed algorithm mostly confuses the neighboring classes. Obviously, this outcome is not very surprising and in line with the findings commonly stated in the literature (Lundsteen et al., 1986). The most of confused cases arise between class number 19 and 20 which are grouped together in Denver classification system of human chromosome since they resemble with regard to the size and arrangement of band patterns (Kothari, Metha, & Roychoudhury, 2009).

At this point, it would be interesting to compare proposed methodology with a recent work which utilizes an ensemble of ANNs instead of SVMs (Gagula-Palalic, 2013). The method is named as Competitive Neural Network Teams (CNNT). Since both approaches have been tested on the same dataset, a comparison seems to be reasonable. Accuracy rates that the two algorithms

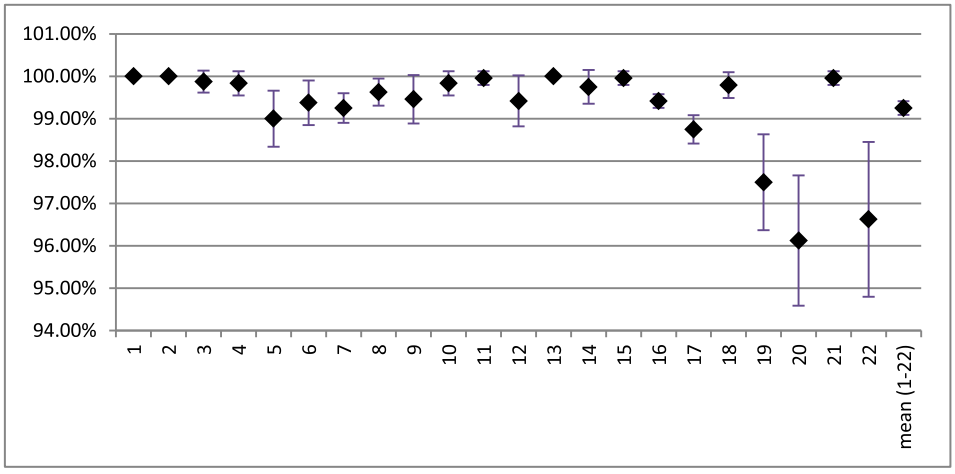


Fig. 9. Average CCRs and ± one standard deviation ranges for each chromosome class in training phase.

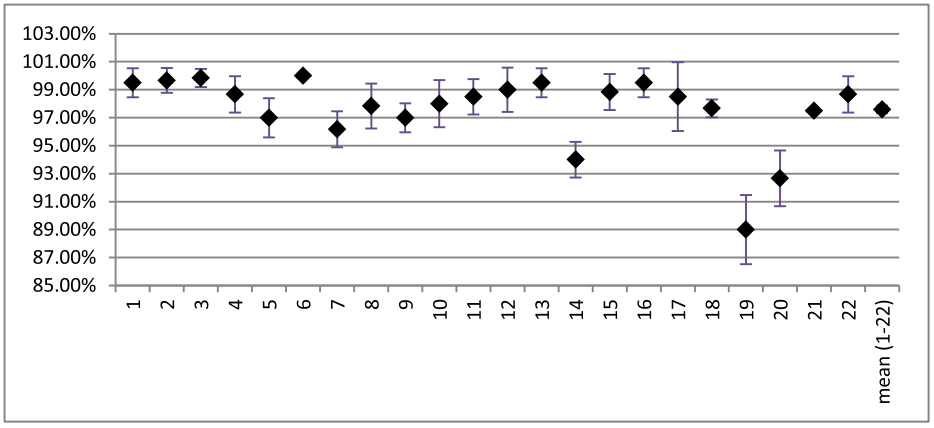


Fig. 10. Average CCRs and ± one standard deviation ranges for each chromosome class in testing phase.

Table 4
Performance measures in training phase obtained by CSVMT.

i	P_i^{best}	P_i^{worst}	Max(P_i)	Min(P_i)	$\overline{P_i}$	s_i
1	100.00	100.00	100.00	100.00	100.00	0.0000
2	100.00	100.00	100.00	100.00	100.00	0.0000
3	100.00	100.00	100.00	99.38	99.88	0.0026
4	100.00	100.00	100.00	99.38	99.83	0.0029
5	100.00	97.50	100.00	97.50	99.00	0.0066
6	98.75	98.75	100.00	98.75	99.38	0.0053
7	99.38	99.38	100.00	98.75	99.25	0.0035
8	100.00	100.00	100.00	99.38	99.63	0.0032
9	100.00	98.75	100.00	98.13	99.46	0.0057
10	100.00	100.00	100.00	99.38	99.83	0.0029
11	100.00	100.00	100.00	99.38	99.96	0.0016
12	99.38	99.38	100.00	98.13	99.42	0.0060
13	100.00	100.00	100.00	100.00	100.00	0.0000
14	99.38	100.00	100.00	98.75	99.75	0.0040
15	100.00	100.00	100.00	99.38	99.96	0.0016
16	100.00	99.38	100.00	99.38	99.42	0.0016
17	98.75	99.38	99.38	98.13	98.75	0.0033
18	100.00	100.00	100.00	99.38	99.79	0.0030
19	98.75	96.88	98.75	95.00	97.50	0.0113
20	98.13	93.13	98.13	93.13	96.13	0.0153
21	100.00	100.00	100.00	99.38	99.96	0.0016
22	97.50	92.50	97.50	91.88	96.63	0.0183
$\bar{P}$	99.55	98.86	99.72	98.30	99.25	0.0016

Table 5
Performance measures in testing phase obtained by CSVMT.

i	P_i^{best}	P_i^{worst}	Max(P_i)	Min(P_i)	$\overline{P_i}$	s_i
1	100.00	100.00	100.00	97.50	99.50	0.0104
2	97.50	97.50	100.00	97.50	99.67	0.0088
3	100.00	100.00	100.00	97.50	99.83	0.0065
4	100.00	97.50	100.00	97.50	98.67	0.0129
5	97.50	95.00	100.00	95.00	97.00	0.0140
6	100.00	100.00	100.00	100.00	100.00	0.0000
7	97.50	95.00	97.50	95.00	96.17	0.0129
8	97.50	97.50	100.00	95.00	97.83	0.0160
9	97.50	97.50	97.50	95.00	97.00	0.0104
10	97.50	97.50	100.00	95.00	98.00	0.0169
11	100.00	97.50	100.00	97.50	98.50	0.0127
12	97.50	97.50	100.00	95.00	99.00	0.0158
13	100.00	100.00	100.00	97.50	99.50	0.0104
14	95.00	92.50	95.00	92.50	94.00	0.0127
15	100.00	97.50	100.00	97.50	98.83	0.0129
16	100.00	97.50	100.00	97.50	99.50	0.0104
17	100.00	95.00	100.00	92.50	98.50	0.0246
18	97.50	97.50	100.00	97.50	97.67	0.0065
19	90.00	90.00	92.50	85.00	89.00	0.0246
20	90.00	92.50	95.00	90.00	92.67	0.0200
21	97.50	97.50	97.50	97.50	97.50	0.0000
22	100.00	100.00	100.00	97.50	98.67	0.0129
$\bar{P}$	97.84	96.93	98.86	95.57	97.59	0.0028

could reach at their best trials during training and testing stages are given in Table 7 where bold-faced values indicate that an equal or a better accuracy is obtained by the corresponding algorithm. According to the reported values, CSVMT outperforms the other method both on training and testing data in most of the chromosome classes and with respect to average accuracy rates.

Table 6
Confusion matrix obtained at the best trial.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	40																					
2	1	39																				
3			40																			
4				40																		
5				1	39																	
6						40																
7			1				39															
8								39	1													
9								1	39													
10							1			39												
11											40											
12									1			39										
13													40									
14														38	2							
15															40							
16																40						
17																	40					
18																1		39				
19																1		1	36	2		
20																			4	36		
21																			1		39	
22																						40

Table 7
Comparison of the results obtained with the proposed algorithm and CNNT.

i	CNNT P_i^{best} (Training)	P_i^{best} (Testing)	CSVMT P_i^{best} (Training)	P_i^{best} (Testing)
1	**100.00**	99.00	**100.00**	**100.00**
2	97.00	**100.00**	**100.00**	97.50
3	96.00	99.00	**100.00**	**100.00**
4	95.00	91.00	**100.00**	**100.00**
5	93.00	89.00	**100.00**	**97.50**
6	96.00	99.00	**98.75**	**100.00**
7	93.00	94.00	**99.38**	**97.50**
8	94.00	89.00	**100.00**	**97.50**
9	93.00	90.00	**100.00**	**97.50**
10	95.00	92.00	**100.00**	**97.50**
11	98.00	96.00	**100.00**	**100.00**
12	91.00	97.00	**99.38**	**97.50**
13	97.00	99.00	**100.00**	**100.00**
14	89.00	**98.00**	**99.38**	95.00
15	99.00	95.00	**100.00**	**100.00**
16	99.00	**100.00**	**100.00**	**100.00**
17	**99.00**	99.00	98.75	**100.00**
18	94.00	**99.00**	**100.00**	97.50
19	89.00	**97.00**	**98.75**	90.00
20	90.00	90.00	**98.13**	**90.00**
21	97.00	**99.00**	100.00	97.50
22	96.00	98.00	**97.50**	**100.00**
$\bar{P}$	95.00	95.86	**99.55**	**97.84**

6. Conclusion

Karyotyping is a clinical technique utilized to classify human chromosomes into 23 distinct classes for medical diagnostic purposes. This task is a time-consuming process and usually accomplished by well-trained clinical cytogeneticists. Thus, developing automated or semi-automated karyotyping systems may help to reduce the efforts devoted to labor-intensive investigations (Wang et al., 2005).

There are two aspects of automated karyotyping: (i) feature extraction and selection, and (ii) development of the classifier. Considering the large number of classes to be distinguished, the classification problem is multi-class classification problem. The complex nature of the problem does not allow to achieve bearable error rates, although there have been several scientific studies aiming to realize such a system. Obviously, the definition of the term 'bearable error rates' is very strict for classification algorithms employed for medical purposes. However, the literature we referred shows that there is still a need for developing better algorithms with minimum misclassification ratios.

In this work, we present a novel approach which consists of two stages: (i) feature selection and extraction with preliminary experiments and PCA, and (ii) a cluster of competing SVMs hybridized with a simple numerical optimization method, *PS*, to distinguish the samples. The algorithm is named as Competitive SVM Teams (*CSVMTs*), and trained and tested on a well-established benchmark dataset. The experimental results we reported here confirm two points. First, the proposed algorithm, *CSVMT*s, is highly reliable with less than 1% and 2.5% average classification error on train and test samples, respectively. Nearly perfect classification performance of the proposed algorithm is clearly visible. Second, it demonstrates a very robust performance with low standard deviations due to random effects which each separate trial possesses. Thus, we can conclude that the higher accuracy rates achieved by the algorithm are not coincidental but due to its robustness.

Despite these highly competitive results, future research effort may be devoted to addressing some issues unaddressed in this work. Firstly, alternative voting strategies other than majority voting may be applied while fusing the responses of SVM experts. The voting scheme may definitely improve the classification performance (Nanni & Lumini, 2009a). Secondly, the performance of the algorithm should be investigated on abnormal metaphase chromosomes. Thirdly, the proposed approach may be exploited in the domain of Human Artificial Chromosomes (HACs) characterization (Katoh et al., 2004; Mejía et al., 2001; Mejiá et al., 2002; Okada et al., 2007).

References

Åström, F., & Koker, R. (2011). A parallel neural network approach to prediction of Parkinson's Disease. *Expert Systems with Applications, 38*(10), 12470–12474. http://doi.org/10.1016/j.eswa.2011.04.028.

Bhasin, M., & Raghava, G. P. S. (2004). Prediction of CTL epitopes using QM, SVM and ANN techniques. *Vaccine, 22*(23–24), 3195–3204. http://doi.org/10.1016/j.vaccine.2004.02.005.

Chen, W.-H., Hsu, S.-H., & Shen, H.-P. (2005). Application of SVM and ANN for intrusion detection. *Computers & Operations Research, 32*(10), 2617–2634. http://doi.org/10.1016/j.cor.2004.03.019.

Delshadpour, S. (2003). Improved MLP neural network as chromosome classifier. In *IEEE EMBS Asian-Pacific Conference on Biomedical Engineering* (pp. 324–325). http://doi.org/10.1109/APBME.2003.1302715.

Feng, X., Cong, P., Zhu, Z., & Du, X. (2012). Automated pairing of human chromosomes applying gradient profile and similarity matching algorithm. *Chemometrics and Intelligent Laboratory Systems, 111*(1), 46–52. http://doi.org/10.1016/j.chemolab.2011.11.009.

Gagula-Palalic, S. (2013). An organized committee of artificial neural networks in the classification of human chromosomes. *International Journal of Computer Applications, 80*(8), 38–41.

Gagula-Palalic, S., & Can, M. (2012). Automatic segmentation of human chromosomes. *Southeast Europe Journal of Soft Computing, 1*(2), 80–83.

Gagula-Palalic, S., & Can, M. (2014). Human chromosome classification using Competitive Neural Network Teams (CNNT) and nearest neighbor. In *IEEE-EMBS international conference on biomedical and health informatics (BHI)* (pp. 626–629). http://doi.org/10.1109/BHI.2014.6864442.

Granum, E., & Thomason, M. G. (1990). Automatically inferred Markov network models for classification of chromosomal band pattern structures. *Cytometry, 11*(1), 26–39. http://doi.org/10.1002/cyto.990110105.

Halder, A., & Pramanik, S. (2012). An Unsupervised Dynamic Image Segmentation using Fuzzy Hopfield Neural Network based Genetic Algorithm. *IJCSI International Journal of Computer Science Issues, 9*(2), 525–532.

Haykin, S. (1999). *Neural networks: a comprehensive foundation*. Upper Saddle River, New Jersey: Prentice Hall International, Inc.

Kajee-Bagdadi, Z. (2007). *Differential evolution algorithms for constrained global optimization*. University of the Witwatersrand.

Kao, J., Chuang, J., & Wang, T. (2008). Chromosome classification based on the band profile similarity along approximate medial axis. *Pattern Recognition, 41*(1), 77–89. http://doi.org/10.1016/j.patcog.2007.05.016.

Katoh, M., Ayabe, F., Norikane, S., Okada, T., Masumoto, H., Horike, S., & Oshimura, M. (2004). Construction of a novel human artificial chromosome vector for gene delivery. *Biochemical and Biophysical Research Communications, 321*(2), 280–290. http://doi.org/10.1016/j.bbrc.2004.06.145.

Kazuki, Y., Hoshiya, H., Takiguchi, M., Abe, S., Iida, Y., Osaki, M., & Oshimura, M. (2011). Refined human artificial chromosome vectors for gene therapy and animal transgenesis. *Gene Therapy, 18*(4), 384–393. http://doi.org/10.1038/gt.2010.147.

Kothari, M. L., Metha, L. A., & Roychoudhury, S. S. (2009). *Essentials of human genetics*. Universities Press.

Kou, Z., Ji, L., & Zhang, X. (2002). Karyotyping of comparative genomic hybridization human metaphases by using support vector machines. *Cytometry, 47*(1), 17–23. http://doi.org/10.1002/cyto.10027.

Kusakci, A. O. (2012). Authorship attribution using committee machines with k-nearest neighbors rated voting. In *11th symposium on neural network applications in electrical engineering* (pp. 161–166).

Legrand, B., Chang, C. S., Ong, S. H., Neo, S., & Palanisamy, N. (2008). Chromosome classification using dynamic time warping. *Pattern Recognition Letters, 29*(3), 215–222. http://doi.org/10.1016/j.patrec.2007.09.017.

Liskovykh, M., Lee, N. C., Larionov, V., & Kouprina, N. (2016). Moving toward a higher efficiency of microcell-mediated chromosome transfer. *Molecular Therapy – Methods & Clinical Development, 3*, 16043. http://doi.org/10.1038/mtm.2016.43.

Lundsteen, C., Gerdes, T., & Maahr, J. (1986). Automatic classification of chromosomes as part of a routine system for clinical analysis. *Cytometry, 7*(1), 1–7.

Lundsteen, C., Philip, J., & Granum, E. (2008). Quantitative analysis of 6985 digitized trypsin G-banded human metaphase chromosomes. *Clinical Genetics, 18*(5), 355–370. http://doi.org/10.1111/j.1399-0004.1980.tb02296.x.

Mao, C., & Liu, S.-H. (2004). Classification of multispectral images through a rough-fuzzy neural network. *Optical Engineering, 43*(1), 103. http://doi.org/10.1117/1.1629685.

Mejiá, J. E., Alazami, A., Willmott, A., Marschall, P., Levy, E., Earnshaw, W. C., & Larin, Z. (2002). Efficiency of de novo centromere formation in human artificial chromosomes. *Genomics, 79*(3), 297–304. http://doi.org/10.1006/geno.2002.6704.

Mejía, J. E., Willmott, A., Levy, E., Earnshaw, W. C., & Larin, Z. (2001). Functional complementation of a genetic deficiency with human artificial chromosomes. *The American Journal of Human Genetics, 69*(2), 315–326. http://doi.org/10.1086/321977.

Moradi, M., & Setarehdan, S. K. (2006). New features for automatic classification of human chromosomes: A feasibility study. *Pattern Recognition Letters, 27*(1), 19–28. http://doi.org/10.1016/j.patrec.2005.06.011.

Nanni, L. (2006). A reliable method for designing an automatic karyotyping system. *Neurocomputing, 69*(13–15), 1739–1742. http://doi.org/10.1016/j.neucom.2006.01.005.

Nanni, L., & Lumini, A. (2009a). A genetic encoding approach for learning methods for combining classifiers. *Expert Systems with Applications, 36*(4), 7510–7514. http://doi.org/10.1016/j.eswa.2008.09.029.

Nanni, L., & Lumini, A. (2009b). Genetic nearest feature plane. *Expert Systems with Applications, 36*(1), 838–843. http://doi.org/10.1016/j.eswa.2007.10.009.

Nie, L., Wang, M., Zhang, L., Yan, S., Zhang, B., & Chua, T.-S. (2015). Disease inference from health-related questions via sparse deep learning. *IEEE Transactions on Knowledge and Data Engineering, 27*(8), 2107–2119. http://doi.org/10.1109/TKDE.2015.2399298.

Nie, L., Zhang, L., Yang, Y., Wang, M., Hong, R., & Chua, T.-S. (2015). Beyond doctors: Future health prediction from multimedia and multimodal observations. In *Proceedings of the 23rd ACM international conference on Multimedia – MM'15* (pp. 591–600). New York, New York: ACM Press.

Nie, L., Zhao, Y.-L., Akbari, M., Shen, J., & Chua, T.-S. (2015). Bridging the vocabulary gap between health seekers and healthcare knowledge. *IEEE Transactions on Knowledge and Data Engineering, 27*(2), 396–409. http://doi.org/10.1109/TKDE.2014.2330813.

Okada, T., Ohzeki, J., Nakano, M., Yoda, K., Brinkley, W. R., & Larionov, V. (2007). CENP-B controls centromere formation depending on the chromatin context. *Cell, 131*(7), 1287–1300. http://doi.org/10.1016/j.cell.2007.10.045.

Oskouei, B. C., & Shanbehzadeh, J. (2010). Chromosome classification based on wavelet neural network. In *2010 international conference on digital image computing: Techniques and applications (DICTA)* (pp. 605–610). http://doi.org/10.1109/DICTA.2010.107.

Piper, J., & Granum, E. (1989). On fully automatic feature measurement for banded chromosome classification. *Cytometry, 10*, 242–255. http://doi.org/10.1002/cyto.990100303.

Poletti, E., Grisan, E., & Ruggeri, A. (2012). A modular framework for the automatic classification of chromosomes in Q-band images. *Computer Methods and Programs in Biomedicine, 105*(2), 120–130. http://doi.org/10.1016/j.cmpb.2011.07.013.

Poletti, E., Ruggeri, A., & Grisan, E. (2011). An improved classification scheme for chromosomes with missing data. In *Proceedings of the annual international conference of the IEEE engineering in medicine and biology society, EMBS* (pp. 5072–5075). http://doi.org/10.1109/IEMBS.2011.6091256.

Ritter, G., & Gao, L. (2008). Automatic segmentation of metaphase cells based on global context and variant analysis. *Pattern Recognition, 41*(1), 38–55. http://doi.org/10.1016/j.patcog.2007.05.013.

Ritter, G., & Schreib, G. (2001). Using dominant points and variants for profile extraction from chromosomes. *Pattern Recognition, 34*(4), 923–938. http://doi.org/10.1016/S0031-3203(00)00035-2.

Ruan, X. (1999). A dynamic neuro-fuzzy system configuration, stability, and fuzzy operational function. *Fuzzy Sets and System, 101*, 315–321.

Ruan, X. (2000). A classifier with fuzzy Hopfield network for human chromosomes. In *Proceedings of the 3rd world congress on intelligent control and automation*.

Sampat, M. P., Bovik, A. C., Aggarwal, J. K., & Castleman, K. R. (2005). Supervised parametric and non-parametric classification of chromosome images. *Pattern Recognition, 38*(8), 1209–1223. http://doi.org/10.1016/j.patcog.2004.09.010.

Sjahputera, O., & Keller, J. M. (1999). Evolution of a fuzzy rule-based system for automatic chromosome recognition. In *FUZZ-IEEE'99. 1999 IEEE international fuzzy systems. conference proceedings (Cat. no. 99CH36315): 1* (pp. 129–134). http://doi.org/10.1109/FUZZY.1999.793219.

Subasi, A. (2012). Medical decision support system for diagnosis of neuromuscular disorders using DWT and fuzzy support vector machines. *Computers in Biology and Medicine, 42*(8), 806–815. http://doi.org/10.1016/j.compbiomed.2012.06.004.

Subasi, A. (2013). Classification of EMG signals using PSO optimized SVM for diagnosis of neuromuscular disorders. *Computers in Biology and Medicine, 43*(5), 576–586. http://doi.org/10.1016/j.compbiomed.2013.01.020.

Subasi, A., & Gursoy, M. I. (2010). Expert Systems with Applications EEG signal classification using PCA, ICA, LDA and support vector machines. *Expert Systems With Applications, 37*(12), 8659–8666. http://doi.org/10.1016/j.eswa.2010.06.065.

Takahashi, Y. (1997). Mathematical improvement of the Hopfield model for TSP feasible solutions by synapse dynamical systems. *Neurocomputing, 15*(1), 15–43. http://doi.org/10.1016/S0925-2312(96)00044-6.

Torczon, V. (1997). On the convergence of pattern search algorithms. *SIAM Journal on Optimization, 7*(1), 1–25.

Vapnik, V. (2000). *The nature of statistical learning theory* (2nd ed.). New York: Springer-Verlag.

Ventura, R., Khmelinskii, A., & Sanches, J. M. (2010). Classifier-assisted metric for chromosome pairing. In *2010 annual international conference of the IEEE engineering in medicine and biology society, EMBC'10* (pp. 6729–6732). http://doi.org/10.1109/IEMBS.2010.5626237.

Wang, X., Zheng, B., Li, S., Mulvihill, J. J., Wood, M. C., & Liu, H. (2009). Automated classification of metaphase chromosomes: Optimization of an adaptive computerized scheme. *Journal of Biomedical Informatics, 42*(1), 22–31. http://doi.org/10.1016/j.jbi.2008.05.004.

Wang, X., Zheng, B., Wood, M., Li, S., Chen, W., & Liu, H. (2005). Development and evaluation of automated systems for detection and classification of banded chromosomes: Current status and future perspectives. *Journal of Physics D: Applied Physics, 38*(15), 2536. Retrieved from http://stacks.iop.org/0022-3727/38/i=15/a=003.

Wang, Y. P. (1999). Image representations using multiscale differential operators. *IEEE Transactions on Image Processing, 8*(12), 1757–1771. http://doi.org/10.1109/83.806621.

Wang, Y. P., & Lee, S. L. (1998). Scale-space derived from B-splines. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*(10), 1040–1055. http://doi.org/10.1109/34.722612.