Evolutionary-Driven Support Vector Machines for Determining the Degree of Liver Fibrosis in Chronic Hepatitis C

Ruxandra Stoean[a,*], Catalin Stoean[a], Monica Lupsor[b], Horia Stefanescu[b], Radu Badea[b]

[a]*Department of Computer Science, University of Craiova, A. I. Cuza Str., No. 13, 200585, Craiova, Romania*
[b]*Department of Ultrasonography, 3rd Medical Clinic, University of Medicine and Pharmacy Cluj - Napoca, Croitorilor Str., No. 19-21, 400162, Cluj-Napoca, Romania*

Summary

Objective: Hepatic fibrosis, the principal pointer to the development of a liver disease within chronic hepatitis C, can be measured through several stages. The correct evaluation of its degree, based on recent different non-invasive procedures, is of current major concern. The latest methodology for assessing it is the Fibroscan and the effect of its employment is impressive. However, the complex interaction between its stiffness indicator and the other biochemical and clinical examinations towards a respective degree of liver fibrosis is hard to be manually discovered. In this respect, the novel, well-performing evolutionary-powered support vector machines are proposed towards an automated learning of the relationship between medical attributes and fibrosis levels. The traditional support vector machines have been an often choice for addressing hepatic fibrosis, while the evolutionary option has been validated on many real-world tasks and proven flexibility and good performance.

Methods and material: The evolutionary approach is simple and direct, resulting from the hybridization of the learning component within support vector machines and the optimization engine of evolutionary algorithms. It discovers the optimal coefficients of surfaces that separate instances of distinct classes. Apart from a detached manner of establishing the fibrosis degree for new cases, a resulting formula also offers insight upon the correspondence between the medical factors and the respective outcome. What is more, a feature selection genetic algorithm can be further embedded into the method structure, in order to dynamically concentrate search only on the most relevant attributes. The data set refers 722 patients with chronic hepatitis C infection and 24 indicators. The five possible degrees of fibrosis range from F0 (no fibrosis) to F4 (cirrhosis).

Results: Since the standard support vector machines are among the most frequently used methods in recent artificial intelligence studies for hepatic fibrosis staging, the evolutionary method is viewed in comparison to the traditional one. The multifaceted discrimination into all five degrees of fibrosis and the slightly less difficult common separation into solely three related stages are both investigated. The resulting performance proves the superiority over the standard support vector classification and the attained formula is helpful in providing an immediate calculation of the liver stage for new cases, while establishing the presence/absence and comprehending the weight of each medical factor with respect to a certain fibrosis level.

Conclusion: The use of the evolutionary technique for fibrosis degree prediction triggers simplicity and offers a direct expression of the influence of dynamically selected indicators on the corresponding stage. Perhaps most importantly, it significantly surpasses the classical support vector machines, which are both widely used and technically sound. All these therefore confirm the promise of the new methodology towards a dependable support within the medical decision-

making.

Keywords: Support vector machines; Evolutionary algorithms; Formula for class prediction; Feature selection; Chronic hepatitis C; Liver fibrosis stage.

1. Introduction

The major indicator of progressive liver disease within chronic hepatitis C – hepatic fibrosis – needs to be most accurately staged for an immediate antiviral therapy in case of a significant level. Recent medical practice for this purpose has moved from the standard liver biopsy, which is both invasive and somewhat unreliable, to either biochemical testing [1] or imaging [2]. The most recent technological advancement for the evaluation of liver fibrosis is the Fibroscan (Echosens, Paris, France), which measures liver stiffness through elastographical means.

However, for the final interpretation of the influence of the liver stiffness and the other complementary medical exams over the corresponding degree of fibrosis, an appropriate learning method from artificial intelligence helps assist the complicated decision-making. Hence, a significant number of papers have arisen in the latest years, exploiting the application of approaches ranging from the classical techniques of naïve Bayes and k-nearest neighbor [3] to the modern neural networks [4].

Remarked for their high predictive power, support vector machines (SVMs) [5] have also been widely used for mining liver fibrosis [1, 2], as well as other medical data [6]. They resolve classification through a geometrical perspective on the problem: Hyperplanes divide data from distinct classes and their optimal formulation must be discovered. Although the accuracy of prediction deriving from a SVM solving of a problem is superior to that resulting from other intelligent systems, the complexity of the traditional mathematical treatment of the inherent optimization task is somewhat uninviting. There are, nevertheless, several software implementations to be straightly utilized, but the inner engine is still a black-box; moreover, they are usually directed towards obtaining the value of an overall accuracy on a test set of data rather than exhibiting a formula for further direct predictions.

It is because of these reasons that the present paper puts forward an alternative methodology by means of a novel hybridization between support vector machines and evolutionary algorithms (EAs) [7]. The evolutionary-driven support vector machines (ESVMs) [8, 9] assume the geometrical principles upon learning/separation of SVMs, but solve the inferred optimization task of determining the optimal hyperplane (coefficients) through evolutionary means.

The motivation of current work is consequently to offer a more flexible methodology provided by the EA, while inheriting the strengths of the classical SVMs, in order to eventually give assistance for liver fibrosis staging. Present approach thus aims towards both an enhanced model and better performance in four main steps:

1. To simplify and exhibit a white-box solving – training of separating hyperplanes – and achieve a hands-on fast testing.
2. To implicitly select the most important indicators for decision making.

*Corresponding author at: Department of Computer Science, University of Craiova, A. I. Cuza, No. 13, 200585, Romania. Tel.: +40 251 413728; fax: +40 251 413728.
E-mail address: ruxandra.stoean@inf.ucv.ro

3. To provide a direct expression (formula) of the influence of the implied medical indicators on the fibrosis level and use it for a direct call to any new patient sample that arrives, without repeating the entire training process.
4. To surpass the accuracy of SVMs in the prediction of the fibrosis degree for new examples.

The paper is structured as follows. Section 2 describes the given problem of predicting stages of liver fibrosis based on stiffness-related and clinical medical indicators: The problem is detailed, the data set is explained and previous studies are discussed. Section 3 outlines the concepts revolving around classification and SVMs – the learning metaphor and training mechanism are reviewed. The proposed ESVM approach is given in Section 4: The aims are defined, the algorithm and components are summarized and an enhancement of the ESVM by exploiting the inner EA in order to additionally achieve the selection of the most important medical indicators is created. The application of the ESVMs on the liver degree establishment is provided in section 5, where three experiments are carried out: two are concerned with the direct solving of the 5- and 3-class formulations of the problem, while the last targets the approach endowed with feature selection. Finally, the performance of prediction of the evolutionary approach in comparison to that of standard SVMs is thoroughly addressed.

2. Materials

The prognosis and management of chronic liver diseases largely depend on the amount and progression of liver fibrosis. In patients with chronic hepatitis C, the precise stage of liver fibrosis is the most important predictor of disease progression and determines the need for antiviral therapy [10]. Until recently, liver biopsy has been the only way to evaluate fibrosis and it has traditionally been considered as the gold standard [11]. Liver fibrosis is evaluated semi-quantitatively according to the METAVIR scoring system as follows: F0 - no fibrosis, F1 - portal fibrosis without septa, F2 - portal fibrosis and few septa, F3 - numerous septa without cirrhosis and F4 - cirrhosis.

However, liver biopsy is an invasive and painful procedure, often with poor patient compliance, also carrying a significant, although small risk of life-threatening complications [12]. An obvious trend in the clinical practice observed in the latest years consists of finding a correct method for liver fibrosis evaluation in a non-invasive way, both by biochemical tests as well as imaging methods, as an alternative to liver biopsy. Virtually, all serological tests developed to date have not entirely met clinician expectations, as they insufficiently predict the stage of liver fibrosis on an individual level [13]. The imaging diagnosis methods have the advantage of being non-invasive and, at the same time, they allow a complete evaluation of the entire organ, with a precise appreciation of the seriousness of the disease when the fibrosis does not affect the liver uniformly. Unfortunately, these methods have been limited for the tracing of cirrhosis and its complications only [14].

The latest technological discovery in the evaluation of liver fibrosis is the Fibroscan, a specially adapted ultrasound device using the principle of the one-dimension transient elastography for the assessment of liver stiffness. This gained popularity as a user-friendly, non-invasive technique for measuring liver stiffness, which has been shown to be significantly correlated with liver fibrosis stages in a variety of clinical conditions including chronic hepatitis C [15, 16]. Although such a good binary threshold differentiation among degrees of fibrosis has been obtained from investigating solely stiffness [17, 18] (also on the particular data set involved in this study [19]),

the interaction between itself and the complementary medical examinations could further improve the performance while increasing the complexity of the problem through a simultaneous distinction between more degrees of fibrosis. It is in this respect that the association between the elastographical value, together with the other clinical and biochemical medical attributes, and the corresponding different stages of liver fibrosis must be artificially discovered and intelligently learnt in order to offer a reliable automated assistance within the modern medical decision-making in the field.

2.1. Data set of liver fibrosis degree prediction within chronic hepatitis C

The chronic hepatitis C data set employed in this paper comes from the 3rd Medical Clinic, University of Medicine and Pharmacy, Cluj-Napoca, Romania, and contains 722 samples of 24 indicators, with a number of missing values. The five possible degrees of fibrosis have the following number of representatives:

- F0 – 29 examples;
- F1 – 227 examples;
- F2 – 164 examples;
- F3 – 87 examples;
- F4 – 215 examples.

The medical attributes that are chosen to trigger a certain degree of liver fibrosis are outlined in Table 1. The first one is the stiffness indicator from the Fibroscan. The others represent the usual hematological and biochemical (illustrative of the hepatic function) exams that are required in a patient with chronic hepatitis C. The fibrosis stage is confirmed by the result of the liver biopsy procedure. Summing up, a patient sample refers 24 non-invasive attributes (including the liver stiffness derived from the Fibroscan) and the outcome established by the invasive biopsy.

2.2. Previous artificial intelligence studies of liver fibrosis grade prediction

Earlier SVM approaches for fibrosis forecasting refer the interpretation of other medical procedures: B-mode ultrasound [2] or non-invasive serum markers [1]. The former used SVM classification to distinguish between the possible levels of fibrosis using extracted image features from 20 post-surgical liver samples. The best classification accuracy of 2, 3, 4 and 5 classes considered in turn were of 91%, 85%, 81% and 72%, respectively [2]. The latter employed a SVM-based feature selection to differentiate between the low-grade fibrosis and the therapy indicating significant level. The obtained accuracy for this 2-class problem on 204 patients was of 87%.

On the complementary level of computationally interpreting the stiffness results from the Fibroscan, there are several works that addressed other different artificial intelligence techniques for assisting decision-making in this non-invasive evaluation of liver fibrosis. ROC analysis is mainly used for this purpose and [19] exploits this idea on the exact current medical environment. 324 results of the Fibroscan alone were taken and the technique was used to resolve the binary classification between tasks defined as F0 vs. F1234 (87.74% obtained correct differentiation), F01 vs. F234 (78.62% accurate distinction), F012 vs. F34 (84.91% performance), F0123 vs. F4 (89.94% accuracy of prediction). Further referring merely to the proposed medical structure

of the data, a primary database that included 125 patients and 26 attributes (stiffness from the Fibroscan plus other clinical indicators) was addressed through the application of neural networks, naïve Bayesian classification and the k-nearest neighbor algorithm, while investigating the importance of the stiffness indicator within the discrimination [3]. On the same data set, [4] employed a feature selection approach, based on the analysis of correlations between the medical indicators and the fibrosis degrees, to improve the classification process performed by both the naïve Bayes classifier and the probabilistic neural network model. The best reported resulting accuracy of prediction was of 70%. To the best of our knowledge, no artificial intelligence-based investigation has been yet performed on the specific data set reported in this paper.

As compared to these preceding studies, the use of proposed ESVM on the data set enunciated in subsection 2.1 aims for taking the classification power from the SVM, for the theoretical success, while simplifying its operation and improving its inner mechanics, for a better performance and usefulness on the practical side. Therefore, following the previous efficient application of SVM approaches for the differentiation among liver fibrosis degrees (and in further other medical fields), the target of present paper is to offer a more flexible and easily functional alternative, which is achieved through the hybridization of SVM learning and EA training/optimization.

3. Prerequisites of the model

The intelligent automated classification of chronic hepatitis C patients into grades of liver fibrosis is approached by a novel technique that proceeds from the state-of-the-art learning paradigm of SVMs, while it further derives the detection of discriminating surfaces between outcomes through the means of EAs [8, 9].

3.1. A characterization of classification

A classification problem may be defined by the subsequent components:

- a set of m training pairs of the form $(\mathbf{x}_i, y_i)$, where each couple holds the information related to a data sample and its confirmed outcome;
- every example is described by n attributes: $\mathbf{x}_i \in [a_1, b_1] \times [a_2, b_2] \times ... \times [a_n, b_n]$, where a_i, b_i denote the bounds of definition for every attribute of a sample;
- each corroborated outcome $y_i \in \{0, 1, ..., k-1\}$, where there are k possible classes;
- a set of p test pairs of the type $(\mathbf{x}_i^{'}, y_i^{'})$, where the target is unknown to the learning machine and must be predicted.

Remark: As in the customary notation, throughout present paper, vectors are written in bold.

Learning then pursues the following steps:

- A chosen classifier learns the associations between each training sample and the acknowledged output.
- Either in a black-box manner or explicitly, the obtained classifier takes each test sample and makes a forecast on its probable class, according to what has been learnt.
- The prediction accuracy of the technique is eventually assessed as the number of correctly labeled cases over the total number of test samples.

- The information regarding the disposal of the misclassified examples – the confusion matrix – brings additional insight into the problem and possible limitations of the considered method.

3.2. Support vector machines

A modern and powerful way to tackle classification is provided by the original, though internally complex technique of SVMs. The methodology considers the distinct classes of examples being divided by geometrical surfaces – separating hyperplanes – whose optimal behavior is determined by an extension of the method of Lagrange multipliers.

3.2.1. Classification particularities

To start with, there are a couple of characteristics in their reference to classification. A k-class task is reconsidered through corresponding 2-class reformulations that are separately and inherently addressed and afterwards combined in a voting scheme. The most famous such mechanism is the one-against-all procedure (1aa) [20]. Moreover, as compared to the other known method, the one-against-one (1a1) scheme, 1aa requires less SVMs to be constructed, although the individual binary problems are larger in this case [21]. 1aa builds k classifiers: Every i^{th} SVM considers all training samples labeled with i as positive and all the remaining as negative. Once each separating surface is determined by the SVM learning and optimization engines, the label for a test sample is given by the class that has the maximum value as output from the k obtained decision functions. Therefore, the second feature of classification within SVMs is that no data example remains unclassified – there is always a class to which it is appointed.

3.2.2. Support vector learning and optimization

Training samples may be either linearly or nonlinearly separable and SVMs build a corresponding hyperplane accordingly. In the best case scenario, when the opposite (positive and negative) sets of samples are strongly disconnected from each other, there exists a linear hyperplane of equation (1) which performs the separation.

$$\langle \mathbf{w}, \mathbf{x}_i \rangle - b = 0, i = 1, 2, ..., m \ . \tag{1}$$

In (1), $\mathbf{x}_i$ stands for every training sample in turn, $\mathbf{w} \in \Re^n$ and $b \in \Re$ are the coefficients of the hyperplane and $\langle\rangle$ denotes the scalar product.

As a stronger condition for linear separability, examples of each class are, in fact, positioned behind a supporting hyperplane for that label, which can be mathematically expressed as (2) [22].

$$y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle - b) > 1, i = 1, 2, ..., m \ . \tag{2}$$

Equation (2) outlines the existence of the supporting hyperplanes for both the positive and negative training sets in a concise manner, where it is regarded that $y_i \in \{-1, 1\}$.

In practice, the situation when (2) holds for all given examples with no exception is extremely rare. Errors of separation may be nevertheless added to the formulation with the additional requirement that they are kept to a minimum. By observing the deviation of every data sample from the corresponding supporting hyperplane, which can be measured as $\pm\xi_i/\|\mathbf{w}\|$, $\xi_i \geq 0$, the condition in (2) may be relaxed by introducing such error variables (3).

$$y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle - b) \geq 1 - \xi_i, \xi_i \geq 0, i = 1, 2, ..., m \ . \tag{3}$$

On the other hand, a discriminating surface must be general enough to handle new instances with proper decision ability. This leads to the prerequisite that the two sets of training examples must have a maximal margin of separation in-between. The hyperplane is therefore compelled to a minimal $\|\mathbf{w}\|$.

Hence, the SVM learning scheme specifically leads to the optimization problem (4).

$$\begin{cases} \min_{\mathbf{w},b} \\ \text{such that the objective } \|\mathbf{w}\|^2 + C \sum_{i=1}^{m} \xi_i, C > 0, \text{ holds} \\ \text{subject to constraints } y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle - b) \geq 1 - \xi_i, \xi_i \geq 0, i = 1, 2, ..., m \ . \end{cases} \tag{4}$$

If the samples cannot be delimited in a linear fashion, then a nonlinear surface is obtained by the kernel trick [5]: Data are mapped into a higher dimensional space where a linear surface is able to do the separation. Kernels are commonly taken in either a polynomial expression (with a certain degree) or Gaussian (radial) formulation (of parameter γ).

The resulting optimization statement in (4) may be standardly resolved by relying on a mathematically complex extension of the Lagrange multipliers technique [5]. A dual formulation is derived and the optimal Lagrange multipliers are considered as the solutions of the system resulting by setting the gradient of the new objective function to zero. Once the Lagrange multipliers are found, several underlying conditions may be used to further compute the coefficients of the hyperplane. Existing software programs, however, only output the class of a test case directly after the black-box training; no formula describing the relations among indicators can be visualized, at least not straightforwardly.

3.2.3. Pros and cons of support vector machines for practical reference

SVMs are popular for real-world classification due to several advantages over classical paradigms:

- From the performance perspective, they are one of the most powerful classifiers.
- Their working is independent of the number of features involved in a decision problem [23]; the "curse of dimensionality" is clearly handled.
- From the computational point of view, SVMs provide a fast training.
- They are very useful for a fast insight into the capability of a robust learning technique, when facing an explicit problem to be solved.

However, there are also several disadvantages that they bring along to their modeler and user:

- Although learning is fast, present implementations do not provide a formula for parameters-decision interaction, which is extremely useful for gaining inside information into the practical phenomenon.
- These decision equations could be nevertheless computed, but this requires further inspection of the available software programs and the design of a second application that could store them and be called whenever a new case appears.
- A complete SVM algorithm may be implemented and expanded for the particular requirements of a current problem, but this will prove to be a relatively demanding task.

- The learning process is rather black-box and the testing phase must necessarily include the training, which slows down the effective application, especially if the number of known cases is large.

4. Methodology

In what follows, the proposed evolutionary approach is described with respect to its components and algorithm.

4.1. Motivation and aim

The recently developed method of evolutionary-driven support vector learning (ESVM) [8] has been thoroughly investigated and validated on several benchmark real-world data sets [9]. Experiments have demonstrated both the absolute performance of the hybridized classifier and its relative competitiveness, in simplicity and prediction ability, in direct comparison to the standard mechanism of SVMs.

The prior application of the SVMs for the liver fibrosis staging and the previous findings and results of ESVMs have provided the motivation to consider employing the new learning strategy towards the discovery of the relationship between the clinical, biochemical, elastographical indicators and the corresponding degree of liver fibrosis. The specific features and implementation of the ESVM open the possibility not only for a strong support in the decision-making process, but also for an explicit mathematical view on the interaction of each medical parameter on the outcome.

In terms of comparison to the existing SVM solution to the problem, it is expected that, besides their inherent straightforwardness and transparency, ESVMs will manage to also outperform the standard SVM methodology as accuracy of the trained classifier is concerned.

The proposed policy unfolds in the following manner. A training data set, comprised of samples with a given outcome for the liver fibrosis grade, is given to the ESVM for learning the underlying example-class correspondence. The ESVM regards the separation into categories in the same geometrical manner as the original SVMs. Once an optimization formulation is reached, the evolutionary component generates the coefficients of each surface of decision between the different degrees. When the coefficients are determined, an equation differentiating each grade of fibrosis from the others is obtained, stored and ready to be utilized, at no further computational cost, whenever a new case arrives. The accuracy of the method is reported as a result of testing a set of previously unseen samples.

4.2. Evolutionary-powered optimization within support vector machines

The optimization task in (4) may be alternatively plainly addressed by the adaptable and general algorithms of evolutionary computation. Based on principles of evolution and hereditary [7], EAs are able to inherently and unrestrainedly determine the coefficients that lead to an optimal separation into classes. EAs maintain a population of potential solutions: Each such individual encodes a candidate array of $\mathbf{w}$ and b and the whole set interacts towards the creation of enhanced solutions and the survival of the fittest against the training examples. After a number of generations, the EA converges to an (almost) optimal solution, which represents the best decision hyperplane that separates samples from divergent classes and is both accurate and general enough. Since the coefficients $\mathbf{w}$ and b are encoded in the structure of individuals, the equation of the evolving separating hyperplane is available at all times and especially useful at the end of the

process. The evolved mathematical combination of chronic hepatitis C indicators may prove to be helpful to understand the weight and interaction of each medical attribute on the liver fibrosis grade. The accuracy of prediction of the model is computed using the available test samples, while the formula is stored for future reference and employment when a new case appears.

4.2.1. The evolutionary algorithm for support vector optimization

The specific evolutionary steps to reach the optimal coefficients of a SVM decision surface are outlined in pseudo-code in Algorithm 1 and are detailed in section 4.2.2.

Algorithm 1 The EA to establish the optimal equation of a SVM decision hyperplane

Require: The optimization problem in (4)
Ensure: The fittest separating hyperplane with high training accuracy and generalization ability
begin
t ← 0
initialize population P(t) of potential hyperplanes of coefficients **w** and b
evaluate P(t) in terms of training accuracy and maximal separation between decision classes
while maximum number of evolution generations not reached **do**
t ← t + 1
select parents in P(t) from P(t-1)
recombination on P(t)
mutation on P(t)
select survivors in P(t)
evaluate P(t)
end while
return fittest hyperplane coefficients
end

4.2.2. The evolutionary components

The considered EA features the following elements and behavior:

Individual encoding An individual c is represented as an array of the coefficients of the hyperplane, **w** and b (5).

$$c = (w_1, ..., w_n, b). \tag{5}$$

Initial population Individuals are randomly generated, such that $w_i \in [-1, 1], i = 1, 2, ..., n$ (recall that n is the number of features of a sample), and $b \in [-1, 1]$.

Fitness assignment The fitness expression derives from the objective function of the optimization problem (4) and is subject to minimization. Constraints are handled by penalizing the infeasible individuals through appointing a function $t : R \rightarrow R$ which returns the value of the argument, if negative, and zero otherwise. Consequently, the expression of the fitness function for determining the optimal coefficients of the decision hyperplane is defined as in (6).

$$f(\mathbf{w}, b) = \|\mathbf{w}\|^2 + C \sum_{i=1}^{m} \xi_i + \sum_{i=1}^{m} [t(y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle - b) - 1 + \xi_i)]^2. \tag{6}$$

The scalar product is employed in the form (7):

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^T \mathbf{v}. \tag{7}$$

Indicators for errors Additionally, all deviations ξ_i, $i = 1, 2, ..., m$ (m is the number of samples), are computed [22] in order to be referred in the fitness function. The current individual (separating hyperplane) is taken and matching supporting hyperplanes are appointed. A minimum and maximum, respectively, of the value for scalar products between $\mathbf{w}$ and the positive and negative samples, correspondingly, are calculated as in (8).

$$\begin{cases} m_1 = min\{\langle \mathbf{w}, \mathbf{x}_i \rangle | y_i = +1\}. \\ m_2 = max\{\langle \mathbf{w}, \mathbf{x}_i \rangle | y_i = -1\}. \end{cases} \tag{8}$$

The coefficients of the supporting hyperplanes are then accordingly derived in (9).

$$\begin{cases} p = |m_1 - m_2|. \\ \mathbf{w}' = \frac{2}{p}\mathbf{w}. \\ b' = \frac{1}{p}(m_1 + m_2). \end{cases} \tag{9}$$

Next, for every training sample $\mathbf{x}_i$, the deviation to its corresponding supporting hyperplane is obtained from (10).

$$\delta(\mathbf{x}_i) = \begin{cases} \langle \mathbf{w}', \mathbf{x}_i \rangle - b' - 1, y_i = +1, \\ \langle \mathbf{w}', \mathbf{x}_i \rangle - b' + 1, y_i = -1, \\ i = 1, 2, ..., m. \end{cases} \tag{10}$$

Finally, if the sign of the deviation equals that of the class, the corresponding $\xi_i = 0$; else, the (normalized) absolute deviation is returned as the indicator for error. Normalization may be necessary as the sum of deviations is added to the expression of the fitness function. As a consequence, in the early generations, when the generated coefficients lead to high deviations, their sum, considered from 1 to the number of training samples, takes over the whole fitness value and the evolutionary process is driven off the course to the optimum.

Parent selection Perhaps the most simple and practical selection method is tournament selection [7] that is based on a relative fitness comparison and consequently is decided to be used in the ESVM. For the number of times required to fill the parents pool, two (or more) individuals are randomly chosen and the best one of them, based on their evaluations, is selected for reproduction.

Variation operators The types of recombination and mutation operators [7] that commonly act within canonical real encoded EAs are also chosen to be employed within the ESVMs.

Within recombination, for every individual in the current population, a random number in the [0, 1] interval is generated; if it is smaller than the given recombination probability, then the current individual is chosen for recombination. If the number of chosen individuals is even, then they recombine as pairs which are randomly formed. Otherwise, one individual is deleted or another one is added from the parents pool, a decision which is also randomly taken. The typically used recombination for real encoded problems is the intermediate scheme, which presumes that the value of each gene of the offspring is a convex combination of the corresponding values of the parents (11). The usual number of offspring is one or two.

$$o_i = \alpha \cdot p_{1_i} + (1 - \alpha) \cdot p_{2_i}, i = 1, 2, ..., n. \tag{11}$$

In present implementation, only one offspring is presumed and α is taken equal to 0.5.

Mutation takes every individual in the current population in turn and, for each gene of that individual, a random number in the [0, 1] interval is generated. If the mutation probability is higher than the generated number, then that gene suffers mutation. For the continuous instances, mutation customarily performs a small perturbation in the value of the selected gene, which is induced by a parameter called mutation strength multiplied by a number that is randomly generated, usually following an either uniform or normal distribution (12).

$$o_i = o_i + N(0, 1) \cdot ms, i = 1, 2, ..., n. \tag{12}$$

Mutation strength is denoted by ms and $N(0, 1)$ stands for the standard normal distribution.

Survivor selection The population of the next generation is decided to be formed of the newly obtained individuals plus the individuals that were not selected for reproduction; this means that the offspring resulting after either recombination or mutation automatically replace their parents. In the current case of recombination, where one offspring results from two parents, it is the less fit of them that is substituted by the descendant.

Stop condition It is opted for the modality when the EA stops after a predefined number of generations. The fittest individual (hyperplane) from the entire evolutionary process is taken as the solution to the classification problem.

Test step As the fittest coefficients of the separating hyperplane, $\mathbf{w}^{opt}$ and b^{opt}, are found, the target for a new, unseen test data instance $\mathbf{x}_i^{'}$ can be determined directly following the function in (13).

$$f(\mathbf{x}) = \langle \mathbf{w}^{opt}, \mathbf{x}_i^{'} \rangle - b^{opt}, i = 1, 2, ..., p. \tag{13}$$

If the classification task is intrinsically binary, then the class is determined from the sign of the function as positive or negative. If the problem has several ($k > 2$) classes, then a voting mechanism, based on the values of the current sample as a parameter of the different resulting decision functions, is applied. In the 1aa mechanism that is chosen to act on the present prediction challenge, a distinct ESVM is created to separate one fibrosis grade from the united others. Each ESVM algorithm determines the optimal expression of a corresponding

decision function (13). In order to establish the label for a test sample, the values of the k functions for that example are compared and the highest estimate is taken to point to the class that triggered that ESVM.

The accuracy of the ESVM classifier is calculated by counting the percent of correctly appointed test cases.

4.3. GA-based feature selection within ESVM

Although SVMs have been demonstrated to circumvent the curse of dimensionality for highly multidimensional data sets [23], it may prove important to concentrate learning only on the most influential features from all given indicators. The ESVM could also surely benefit from some form of prior feature selection. Therefore, one of the commonly used feature extraction mechanisms, Principal Component Analysis (PCA), has been employed in this respect. This led to a sizeable reduction of the data dimensionality from 24 to only 6 assembled attributes. Unfortunately, on the newly obtained data, both the ESVM and SVM performed comparative to the previously achieved accuracies, but slightly worse. There was however a major improvement in runtime.

Nevertheless, since EAs power the currently proposed technique, we can once more benefit from their flexibility in order to implement an original approach for choosing features leading to improvement in the performance of the ESVM. The presented method is thus further endowed with a mechanism for dynamic feature selection provided by a genetic algorithm (GA).

The novel GA-ESVM is consequently constituted of a population of individuals on two levels. The higher level of an individual selects optimal indicators, whose presence/absence is encoded by binary chromosomes, while the lower one generates weights for the chosen attributes.

The fitness assignment triggers the evaluation on both levels, i.e. what are the selected features and what are their weights. The individuals to produce the next generation are selected as before, always retaining both encodings.

Each level of an individual in the population is nevertheless modified by its own genetic operators: the low one as before, while the high GA parts of a chromosome undergo one-point recombination and bit-flip mutation [7]. This type of recombination swaps parts of the parents after a randomly generated cut point in order to obtain two offspring that therefore borrow particularities from both. Bit-flip mutation changes the value of a stochastically chosen gene from 1 to 0 and viceversa. Survivor selection is also performed differently at the high-level, since offspring (based on the symmetric evaluation value between the two encodings) replace their parents only if better.

The EA for the double encoding follows the evolutionary cycle in Algorithm 2.

4.4. Theoretical implications of an evolutionary support vector solution to fibrosis differentiation

The newly elaborated ESVMs thus preserve the originality of SVM learning but simplify the solving of the optimization derived assignment through the flexible nature of EAs.

As a result, in contrast to the SVMs, the ESVMs exhibit a number of qualities, both globally and in particular for the given task of fibrosis degree prediction:

- The evolutionary technique resolves the complexity of support vector training, traditionally performed by the mathematically intricate and constrained extension of the Lagrange multipliers method.

Algorithm 2 The GA-ESVM to establish both the optimal features and weights of a SVM decision hyperplane

Require: The optimization problem in (4)
Ensure: The fittest separating hyperplane with high training accuracy and generalization ability
begin
t ← 0
initialize population P(t) of potential hyperplanes of coefficients **w** and b, where each individual is made of two levels: a binary one when the powerful attributes are selected and a real-valued level encoding where weights for these chosen features are generated
evaluate P(t) in terms of training accuracy and maximal separation between decision classes, taking into account both the selected indicators and the weights in each individual
while maximum number of evolution generations not reached **do**
t ← t + 1
select parents in P(t) from P(t-1)
recombination on high-level P(t)
recombination on low-level P(t)
mutation on high-level P(t)
mutation on low-level P(t)
select survivors in P(t)
evaluate P(t)
end while
return fittest hyperplane coefficients along with its selected features
end

- The implementation is thought to output the coefficients of the separating hyperplane, so that the test phase actually takes place separately (from the training module) and instantly as a new case is given as input.
- Because the ESVMs specifically yield the coefficients of the hyperplanes, the decision equations are additionally available and these are highly important to physicians for the study regarding the impact of the indicators on the outcome.
- From the algorithmic point of view, the methodology involves active understanding and participation from the modeler that connects the artificial learning with the real-world medical situation.

There are however also some possible limitations that may arise in the experimental stage:

- EAs make use of many parameters which are nevertheless relatively easy to be set in practical applications.
- Because of the reference to all samples at the evaluation of every individual, ESVMs have a reasonably slower training period of time than the corresponding procedure within SVMs, which makes only one call to the data set.

5. Results and interpretation

Apart from the theoretical arguments that the ESVM surpasses the classical SVM in terms of simplicity and operability, the assumption that the evolutionary alternative also outperforms

the standard engine as regards accuracy on the practical fibrosis prediction task is further on investigated. Several experiments and enhancements are performed, the involved parametrization issues are addressed and an objective comparison to the original formulation is thoroughly accomplished.

5.1. Pre-validation on different medical data

The proposed approach has demonstrated its efficiency in its application to several test cases from different real-world domains [8, 9]. Nevertheless, it may be important to review some prevalidation of its performance on another data set also from the medical domain with a population of patients different from the one in this paper. In this respect, we analyzed data on diabetes mellitus diagnosis from the Johns Hopkins University. All patients in the data set are females of at least 21 years old, of Pima Indian heritage, living near Phoenix, Arizona, USA. There are 8 attributes (either discrete or continuous) referring personal data (age, number of pregnancies) and medical data (blood pressure, body mass index, result of glucose tolerance test etc).

The last attribute is a discrete one and outlines the diagnosis, which is either 0 (negative) or 1 (positive). 34.9% of the patients in the data set are assigned diabetes positive. The total number of cases is 768 and the data is complete.

When averaging 30 random trials with a rate of 75%-25% training-test samples, ESVMs outperform SVMs with a result of 77.31% accuracy as opposed to 76.82% of the latter.

5.2. Setup and pre-experimentation

The 722 patient examples available for fibrosis degree determination are split into 542 training and 180 test samples, in the proportion 75-25% that is customary in machine learning experimentation. The common method of mean substitution is chosen for the imputation of missing values; therefore, the mean values – computed for every class in turn – replace the missing records for instances of the corresponding categories. Data are normalized, and so are errors of classification ξ_i, $i = 1, 2, ..., m$, as these choices led to better performance in the pre-experimentation stage.

The ESVM is run 30 times on randomly generated training and test sets of the established percentage, for reasons of both statistical validity and mandatory analysis within EA trials. The reported performance is computed as the average of the accuracy results in 30 applications of the algorithm on the test samples.

The values for the parameters to be tuned are generated by Latin hypercube sampling (LHS) in order to automatically obtain a set (as large as necessary) of different configurations. This statistical method is employed to generate a space-filling (fair) sample of the algorithms parameters. For small sample sizes, it is well-known to generate more even distributions than random sampling [24]. The application of this tool requires an a priori setting of parameter intervals. The upper bound of the population size is restricted to 200 and that for the number of generations to 500. The mutation and recombination probabilities are selected in the range [0, 1]. The upper limit for the mutation strength parameter is set to 2. 30 designs are subsequently generated and the best one in resulting accuracy gives the appropriate values for the parameters.

As the choice of the kernel is concerned, the debate as to which type would be more appropriate for the current problem does not have a straightforward answer. It is generally confirmed that the Gaussian kernel outperforms the linear form, however it is theoretically acclaimed that one situation when the latter is more suitable than the former is when the number of features is very large (as here, where the fibrosis differentiation assignment has 24 attributes). In this case, the data do not need to be mapped into a higher dimensional space, as a nonlinear kernel does

not improve the performance and is more difficult to parameterize at the same time [25, 21]. It finally remains for the automatic parameter setting procedure to establish the proper kernel when confronted with the actual task.

For the multi-class treatment, 1aa is preferred to 1a1 in the undertaken experiments, as a small number of ESVMs to be built is cheaper for the particular methodology than a relatively fewer number of samples for each class that would still not considerably reduce runtime for the evolutionary solving. What is more, 1aa permits a more accurate discrimination between the particularities of each individual class and that of the rest. Finally, the early experiments demonstrated a better prediction with the 1aa mechanism.

5.3. Experiment 1: Demarcation between five possible grades of fibrosis

The most difficult task within the establishment of the fibrosis grade on the base of several medical factors is to distinguish between every single such stage. Therefore, the aim is to design an appropriate technique to discover the individual conditions for each degree from F0–F4.

The first experiment thus targets to obtain the ESVM model for the separation of the chronic hepatitis C patients into the 5 distinct classes of fibrosis and attempts to provide the formula of indicators interaction for each outcome as opposed to the others.

The best obtained design of the LHS holds values for the evolutionary and SVM parameters as depicted in the corresponding column (ESVM 5c.) of Table 2. However, the LHS obtained many different well performing configurations for the involved parameters, which brings another quality of the proposed solution: The ease in establishing values for the variables supports both the practical implications of its utilization and the viability of the choice of an evolutionary multi-parameterized methodology.

The average over the 30 performed trials of random cross-validation resulted in a test accuracy of 62.03% (Table 9, ESVM column). The distribution of correctly predicted versus misclassified cases per class can be visualized in the confusion matrix of Table 3, in an example run that obtained 65% accuracy.

The solution provided by the ESVM is also important for the capability to output formulas for the differentiation of each class from all the others, as for every degree it gives the coefficients for each indicator. When applied to the values a new patient has for the considered indicators, the coefficients are multiplied with the corresponding patient values and then summed as in (13). As a consequence, each such coefficient value represents a weight for that given attribute: be that the weight is closer to zero, then the corresponding feature has a smaller importance; if it has a higher value, be that it is positive or negative, it will more powerfully influence the final outcome. Figure 1 illustrates sample coefficients found by ESVM for each of the 24 attributes and for each of the 5 classes. So, the coefficients for the $F0$ class, for example, can be obtained from the first bar in each group. It has to be underlined that a smaller negative value or a higher positive one does not mean that the value for the corresponding attribute has to be low or high: it just means that the attribute has high significance.

The figure brings evidence of the high significance of the Fibroscan-derived stiffness for classes $F0$, $F1$ and $F4$, while the usual medical exams exhibit different smaller influences as the three degrees are concerned. These classes are nevertheless controlled in a significant degree also by aspartate aminotransferase (w_8). $F2$ and $F3$ are however influenced in a higher extent by the total bilirubin indicator (w_{11}). It is worth noticing that most of the weights change in an almost linear manner from $F0$ to $F4$: starting from negative values for the first and reaching to relatively high positive values for the latter or viceversa.

Once the expressions of the relationships between the medical indicators and each fibrosis degree are found as in this example, the labeling of any new case can be instantly conducted: Its values for every attribute ($A1$-$A24$) are introduced in each stored formula and the maximum obtained number indicates the predicted grade from whose expression it was computed.

5.4. Comparison to standard support vector machines on the initial 5-class problem

Out of the available programs for SVM modeling, it was decided to use two packages within the R software and environment, i.e. kernlab and e1071, to implement a SVM for an objective comparison. The reasons for employing these very packages were twofold: kernlab is a flexible SVM implementation and the svm() function in e1071 is a robust interface.

The best performing design of the LHS gives the values for the inner parameters (type of kernel, parameter degree or γ of the kernel and the penalty for errors C) that are found in the corresponding column (SVM 5c.) of Table 2. The chosen implementation presumes an implicit 1a1 separation.

30 random trials were also executed by the SVM and the approach obtained a test accuracy of 61.5% (Table 9, column SVM). The resulting confusion matrix is given in Table 4 on a particular run that obtained 64.44% accuracy. It must be noted that the two confusion matrices should not be compared in absolute values, but rather in a relative manner, as the training and test configurations are randomly chosen for the two examples.

In order to verify the significance of the difference between the accuracy results of ESVMs and SVMs, a t-test for independent samples is used to compare the difference in means between the two approaches. An alternative given by the Wilcoxon rank-sum test is also employed, since in some instances it gives greater power to reject the null hypothesis than the t-test. The obtained p-values were of 0.54 and 0.33, respectively, therefore the difference in performances is not significant (Table 10).

5.5. Experiment 2: Discrimination among mild, clinically significant and severe fibrosis

The ESVM provided reliable results and also sustained the functional purpose of supporting the decision-making process of the physician. However, it is important to continue the examination of the problem and apply the approach to a classification task restricted to less classes, i.e. by uniting the very similar categories of fibrosis: F0 and F1, on the one hand, and F2 and F3, on the other. This is both medically reasonable and practically essential by observing the misclassified cases among classes in the attained results. Joining such similar fibrosis grades together is also reported by [1].

As regards the purpose of this paper, it will be investigated if the data, now distributed to only three classes, is less ambiguous; this is anticipated to sustain the ESVMs get significantly more accurate than the SVMs.

The values for the parameters as provided by the best performing LHS design points are given in the matching column (ESVM 3c.) of Table 2, although many valuable configurations resulted.

By applying the ESVM on the 3-class fibrosis problem, an accuracy on the test set of 73.6% was now achieved (Table 9, ESVM column), breaking the barrier induced by the large similarities between values of the indicators for grades that are now in the same category. The confusion matrix on a sample run that exhibited 75.6% accuracy of predicting the classes of test cases is depicted in Table 5.

The relationship between the medical indicators and the fibrosis grades $F0$-$F1$, $F2$-$F3$ and $F4$ are found once more by the ESVM and detailed in Figure 2 in the same example run as referred

in Table 5. Again, the stiffness seems to have the highest importance for classes $F0 - F1$ and $F4$. Alkaline phosphatase and thrombocytes (w_{12} and w_{23}) appear to also make a very straight demarcation between the 3 outcomes.

5.6. Validation against standard support vector machines on the 3-class formulation

The SVM model was applied to the fibrosis data set and provided a test accuracy of 71% (Table 9, SVM column). The correctly classified versus the misclassified cases are shown in the confusion matrix in Table 6 of a run that obtained 75% accuracy. The generated LHS configuration for best performance is outlined in Table 2 (SVM 3c.).

The difference in resulting accuracies between ESVMs and SVMs was also further subject to statistical testing. The t- and Wilcoxon rank-sum tests (Table 10) now show a significant difference in the results of the two techniques, with p-values of 0.002 given by both tests, and thus validate the hypothesis that the ESVMs perform notably better than the SVMs for the given 3-class task.

5.7. Experiment 3: Feature selection within proposed method

The current experiment investigates whether the GA-based feature selection manages to enhance the performance of the original method.

Values for parameters for both the 5 and 3 classes formulations are again derived from LHS. The best corresponding designs are depicted in Table 2 (GA-ESVM 5c. and 3c.).

When applied to the 5-classes fibrosis instance, the GA-ESVM algorithm obtained an accuracy of 64.07%, which is a large improvement as opposed to the previous performance without feature selection. The 3-classes case was resolved with a correct classification rate of 74.76%, surpassing the earlier ESVM algorithm alone again.

The confusion matrices, once more relative to the training/test formations, are outlined in Tables 7 (a run of 66.66% accuracy for 5 classes) and 8 (one of 77.77% performance for 3 outcomes).

The supremacy, in terms of classification accuracy, of the GA-ESVM over the SVM and even the former ESVM (Table 9, GA-ESVM column) is statistically confirmed by both the t- and Wilcoxon tests in Table 10.

The GA-ESVM additionally outputs the information regarding the most significant attributes that reach a certain fibrosis level. For the differentiation of each class, the GA-ESVM dynamically selects those indicators that distinguish it from the other degrees. Starting from randomly chosen attributes, the algorithm cycle pursues an active evolution and an increase in the importance of certain features in the absence of others. In the end, we find that in the formula for each fibrosis level it is at most 2 attributes that are missing, one often chosen being the *age* ($A2$). Figures 3 and 4 illustrate the attributes weights for the 5- and 3-classes instances. When an attribute is removed by the GA, its corresponding weight is set to zero in the figures. Beside the Fibroscan stiffness that is again clearly the most important attribute, other attributes seem to have important roles in the differentiation between the classes for both the 5 and 3 outcomes: aspartate aminotransferase (w_8), prothrombin index (w_{13}), medium erytrocity volume (w_{20}), thrombocytes (w_{23}). The GA-ESVM approach has a good advantage in the fact that an attribute is not removed for all outcomes at once; for instance, it might be the case that an attribute has no meaning for the healthy people, but it might have a meaning for the ones that possess a certain degree of liver fibrosis.

5.8. Practical assistance of the evolutionary support vector algorithm

Regarding the functionality of proposed algorithm for the complex task of a non-invasive prediction of the correct stage of liver fibrosis, there are several observations to be drawn:

- ESVMs perform very well and stable.
- The automatic tuning of the involved variables illustrates that ESVMs are very robust and flexible, with several possible good working configurations.
- ESVMs, as well as GA-ESVMs, provide degrees of importance for each involved indicator.
- ESVMs are able to provide a formula of the relationship between the indicators and the fibrosis degree, which is both informative and prompt with the arrival of new cases.
- The enhanced GA-ESVM is moreover able to also inherently give the most important indicators and, based on these, a significantly increased performance as opposed not only to SVMs but also to the prior version of ESVMs.

6. Conclusions and prospective work

An evolutionary-propelled support vector learning methodology as opposed to the traditional paradigm is proposed for the prediction of stages for liver fibrosis within chronic hepatitis C, as resulting from standard medical examination and the new technological non-invasive approach of Fibroscan. Motivated by the previous successes of SVMs for this hepatological problem and the former validated superiority of the ESVM over the SVM application, the current paper investigates the ESVM solution to the task in terms of simplicity, accuracy and usefulness, taking into consideration both a simultaneous 5-class discrimination and a coupled 3-degree formulation.

The design and functionality of the application of ESVMs for the discrimination among stages of liver fibrosis attained its planned goals. The solution is straightforward, advantageous and handy in several aspects:

- The whole training/optimization module, provided by means of EAs, is transparent and adaptable.
- Aside from the delivery of the most appropriate fibrosis grade for a given configuration of medical attributes, ESVMs also output the formula of interaction between the latter and the former. This is highly important for an immediate testing of new cases and for understanding relationships between medical variables that may be difficult to grasp even for experienced physicians.
- The importance of each indicator is highlighted by the ESVMs by observing the weights that are found for the considered attributes.
- Most notably, the performance superiority of the alternative improved GA-ESVM is statistically significant in comparison to the traditional SVM.
- Moreover, the GA-ESVM is able to also dynamically determine what are the medical variables that influence the fibrosis degree by an intrinsic feature selection mechanism.

- Parametrization is easily manageable even with the increase in the number of variables that is inherently triggered by the use of an EA.

Although the expected behavior of the evolutionary was confirmed by the experimental evaluation, there is one small restraint to be overcome in forthcoming investigations. Training is somewhat slow as opposed to that within standard SVMs, due to repeated reference to all samples. This can be counterattacked through the use of a chunking procedure that selects only parts of the training set instead of all data to be learnt.

Acknowledgements

Present work was supported by the research grant No. 41071/2007, entitled "SONOFIBROCAST" (Diagnostic algorithm for the stage and prediction of evolution of hepatic fibrosis using non-invasive ultrasound techniques, optimized by stochastic analysis and image analysis), financed by the Romanian Ministry of Education and Research and the National Authority for Scientific Research (MEdC & ANCS).

The authors would like to thank the editor and the reviewers for their helpful comments and suggestions that brought insight and improvement to the paper during the revisions.

References

[1] Z. Jiang and K. Yamauchi and K. Yoshioka and K. Aoki and S. Kuroyanagi and A. Iwata et al, Support Vector Machine-Based Feature Selection for Classification of Liver Fibrosis Grade in Chronic Hepatitis C, *J Med Syst* 20 (2006) 389–394.

[2] W. Yeh and S.W. Huang and P.C. Li, Liver fibrosis grade classification with B-mode ultrasound. *Ultrasound Med. Biol.* 29(9) (2003) 1229–1235.

[3] M. Gorunescu and S. Belciug and A.-B. Salem and M. Lupsor and R. Badea and H. Stefanescu, A machine learning-based diagnosis for liver diseases using the Fibroscan medical ultrasound technique, *4rd ACM International Conference on Intelligent Computing and Information Systems ICICIS09*, M.E. Khalifa and A.-B. Salem, eds., Police Press, Cairo (2009) 243–248.

[4] S. Belciug and M. Lupsor and R. Badea, Features selection approach for non-invasive evaluation of liver fibrosis, *Annals of the University of Craiova, Mathematics and Computer Science Series* 35 (2008) 15–20.

[5] V. Vapnik, *The Nature of Statistical Learning Theory* (Springer-Verlag, Berlin, Heildelberg, New York, 1995).

[6] S.-H. Chiu and C.-C. Chen and T.-H. Lin, Using Support Vector Regression to Model the Correlation between the Clinical Metastases Time and Gene Expression Profile for Breast Cancer, *Artif. Intell. Med.* 44 (2008) 221–231.

[7] A.E. Eiben and J.E. Smith, *Introduction to Evolutionary Computing* (Springer-Verlag, Berlin, Heildelberg, New York, 2003).

[8] R. Stoean and M. Preuss and C. Stoean and E. El-Darzi and D. Dumitrescu, Support Vector Machine Learning with an Evolutionary Engine, *J Oper Res Soc*, 60(8) (2009) 1116–1122.

[9] R. Stoean and M. Preuss and C. Stoean and E. El-Darzi and D. Dumitrescu, An Evolutionary Approximation for the Coefficients of Decision Functions within a Support Vector Machine Learning Strategy, in: A. E. Hassanien and A. Abraham, eds., *Foundations on Computational Intelligence*, Studies in Computational Intelligence (Springer-Verlag, Berlin, Heildelberg, New York, 2009) 83–114.

[10] NIH Consensus Statement on Management of Hepatitis C, *NIH Consens State Sci Statements 2002* 19 (2002) 1–46.

[11] A.A.Bravo and S.G. Sheth and S. Chopra, Liver biopsy *N Engl J Med* 344 (2001) 495–500.

[12] J.F. Cadranel and P. Rufat and F. Degos, Practices of liver biopsy in France: results of a prospective nationwide survey. For the Group of Epidemiology of the French Association for the Study of the Liver (AFEF) *Hepatology* 32 (2000) 477–481.

[13] N.H. Afdhal and M. Curry, Technology evaluation: a critical step in the clinical utilization of novel diagnostic tests for liver fibrosis *J Hepatol* 46 (2007) 543–545.

[14] M. Pinzani and K. Rombouts and S. Colagrande, Fibrosis in chronic liver diseases: diagnosis and management *J Hepatol* 42 (2005) S22-S36.

[15] L. Sandrin and B. Fourquet and J.M. Hasquenoph and S. Yon and C. Fournier and F. Mal et al, Transient elastography: a new noninvasive method for assessment of hepatic fibrosis, *Ultrasound Med Biol* 29 (2003) 1705–1713.

[16] L. Castera and J. Vergniol and J. Foucher, Prospective comparison of transient elastography, Fibrotest, APRI, and liver biopsy for the assessment of fibrosis in chronic hepatitis C *Gastroenterology* 128 (2005) 343–350.
[17] M. Ziol and A. Handra-Luca and A. Kettaneh, Noninvasive assessment of liver fibrosis by measurement of stiffness in patients with chronic hepatitis C *Hepatology* 41 (2005) 48–54.
[18] J. Foucher and E. Chanteloup and J. Vergniol and L. Castera and B. Le Bail and X. Adhoute et al, Diagnosis of cirrhosis by transient elastography (FibroScan): a prospective study *Gut* 55 (2006) 403–408.
[19] M. Lupsor and R. Badea and H. Stefnescu and M. Grigorescu and Z. Sparchez and A. Serban et al, Analysis of histopathological changes that influence liver stiffness in chronic hepatitis C. Results from a cohort of 324 patients *J Gastrointestin Liver Dis* 17(2) (2008) 155–163.
[20] C.-W. Hsu and C.-J. Lin, A comparison of methods for multi-class support vector machines, *IEEE Trans. Neur. Net.* 13 (2004) 415–425
[21] A. Karatzoglou and D. Meyer and K. Hornik, Support Vector Machines in R, *J Stat Soft* 15(9) (2006) 1–28
[22] R.A. Bosch and J.A. Smith, Separating hyperplanes and the authorship of the disputed federalist papers, *Amer Math Mon* 105(7) (1998) 601-608.
[23] T. Joachims, Text categorization with support vector machines: Learning with many relevant features, *10th European Conference on Machine Learning (ECML-98)*, C. Nedellec and C. Rouveirol, eds., Springer-Verlag, Berlin, Heildelberg, New York (1998) 137–142.
[24] D.C. Montgomery, *Design and Analysis of Experiments* (5th ed.) (Wiley, NY, 2001).
[25] C-W. Hsu and C-C. Chang and and C-J. Lin, A Practical Guide to Support Vector Classification http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf (Accessed: 10 January 2010) (2003).

Table 1: Description of fibrosis data: tissue stiffness, clinical and biological parameters and their values in mean and standard deviation.

Indicator	F0	F1	F2	F3	F4	Total
Stiffness ($A1$)	4.28 ± 0.88	6.06 ± 2.02	7.8 ± 3.6	11.92 ± 6.93	32.28 ± 18	14.9 ± 15.42
Sex ($A2$)	Male: 8	Male: 78	Male: 55	Male: 32	Male: 118	Male: 291
BMI (body mass index) ($A3$)	25.23 ± 5	23.76 ± 8.05	24.94 ± 6.27	24.31 ± 8.2	24 ± 9.6	24.27 ± 8.12
Glycemia ($A4$)	110 ± 79.83	83.4 ± 41.39	80.45 ± 47.42	99 ± 44.62	101.93 ± 49.3	91.2 ± 48.5
Triglycerides ($A5$)	125.66 ± 90	92.62 ± 79.96	87.48 ± 64.66	98.85 ± 64	106.6 ± 74.92	97.7 ± 74.25
Cholesterol ($A6$)	207.48 ± 73.92	154.44 ± 94.38	150 ± 86.23	164 ± 73.8	147.27 ± 74.5	154.58 ± 84.42
HDL Cholesterol ($A7$)	39.21 ± 28.49	33.42 ± 32.45	37.4 ± 32.93	33.97 ± 32.6	27.9 ± 29.46	33 ± 31.7
Aspartate aminotransferase ($A8$)	32.41 ± 20.42	39 ± 29	47.47 ± 36.5	68.56 ± 54.37	77.15 ± 54.9	55.58 ± 45.95
Alanin aminotransferase ($A9$)	53.72 ± 37.85	63.77 ± 51.24	75.62 ± 60.63	102.75 ± 94.93	77.2 ± 59.82	74.75 ± 63.24
Gama glutamyl transpeptidase ($A10$)	63.59 ± 63.52	49.19 ± 46.74	62.45 ± 71	78.34 ± 84.3	137.4 ± 208.44	82.56 ± 130.7
Total bilirubin ($A11$)	0.57 ± 0.31	0.56 ± 0.4	0.58 ± 0.46	0.7 ± 0.4	2.15 ± 7.68	1.05 ± 4.26
Alkaline phosphatase ($A12$)	188.83 ± 146.31	148.17 ± 97.01	150.33 ± 110.45	178.97 ± 106.77	248.95 ± 163.3	184 ± 133
Prothrombin index ($A13$)	87 ± 42.82	76.64 ± 47.5	63.84 ± 50.45	70.15 ± 44.28	69.4 ± 36.8	71.21 ± 45
TQS (Quiq Time) ($A14$)	12.36 ± 5.93	10.78 ± 10.37	9.06 ± 7.76	10.6 ± 7.9	16 ± 7	12 ± 8.86
INR (prothrombin time ratio) ($A15$)	0.84 ± 0.36	0.75 ± 0.43	0.68 ± 0.48	0.84 ± 0.46	1.09 ± 0.48	0.85 ± 0.48
Prolonged activated partial thromboplastin time ($A16$)	24.8 ± 12.26	20.43 ± 14.23	17.85 ± 15.08	21 ± 14.42	28.26 ± 11.66	22.42 ± 14.23
Haematids (erythrocytes) ($A17$)	4.05 ± 1.72	3.87 ± 2.08	29.97 ± 336.66	4.16 ± 1.85	6.26 ± 29.13	10.55 ± 161.2
Hemoglobin ($A18$)	11.89 ± 5.07	11.5 ± 5.95	11.05 ± 6.36	12.62 ± 5.37	12.85 ± 4.32	11.95 ± 5.54
Hematocrit ($A19$)	35.18 ± 14.86	31 ± 19.68	25.65 ± 21.44	30.9 ± 19.87	36.06 ± 14.38	31.45 ± 18.88
Medium erytrocity volume ($A20$)	69 ± 36.23	57.7 ± 41.63	48.92 ± 44.11	58.8 ± 42	72.82 ± 35.73	60.8 ± 41.3
Avg. erytrocitary hemoglobin ($A21$)	22.16 ± 12.83	18.7 ± 14.38	16.43 ± 15.04	19.2 ± 14.62	26.3 ± 15.8	20.65 ± 15.4
Avg. concentration of hemoglobin in a red blood cell ($A22$)	16.31 ± 17.2	17.26 ± 17	15.17 ± 16.8	18.6 ± 17.05	21.68 ± 17.22	18.23 ± 17.16
Thrombocytes ($A23$)	209.69 ± 73.35	184.6 ± 105.54	170.85 ± 105.3	175.84 ± 85.19	116.33 ± 76.3	161 ± 98.56
Sideraemia ($A24$)	56.59 ± 56	69.38 ± 58.45	73.31 ± 62.69	86 ± 74.2	85.25 ± 77.72	76.5 ± 67.82

Table 2: Values for parameters of ESVM, SVM and GA-ESVM for the 5- and 3-classes tasks, respectively, as obtained from best design points of the LHS.

	Method					
	ESVM		SVM		GA-ESVM	
Parameter	5c.	3c.	5c.	3c.	5c.	3c.
Population size	146	105	-	-	120	90
Number of generations	292	210	-	-	230	192
Binary recombination probability	-	-	-	-	0.07	0.05
Real recombination probability	0.23	0.12	-	-	0.5	0.6
Binary mutation probability	-	-	-	-	0.05	0.04
Real mutation probability	0.21	0.78	-	-	0.5	0.6
Mutation strength	0.98	0.54	-	-	0.5	0.5
Kernel type	pol.	pol.	rad.	pol.	pol.	pol.
Degree of polynomial kernel	1	1	-	1	1	1
γ of radial kernel	-	-	0.1	-	-	-
C	1	1	1	1	1	1

Table 3: Example of a resulting confusion matrix of the ESVM for the 5-degree fibrosis differentiation: Rows exhibit predicted outcomes, columns show actual classes. The diagonal outlines the amount of correctly classified samples, while each intersection of coordinates signals the number of misclassified cases.

	True				
Predicted	$F0$	$F1$	$F2$	$F3$	$F4$
$F0$	0	0	0	0	0
$F1$	7	63	32	9	1
$F2$	1	1	2	1	0
$F3$	0	0	4	4	3
$F4$	0	0	2	2	48

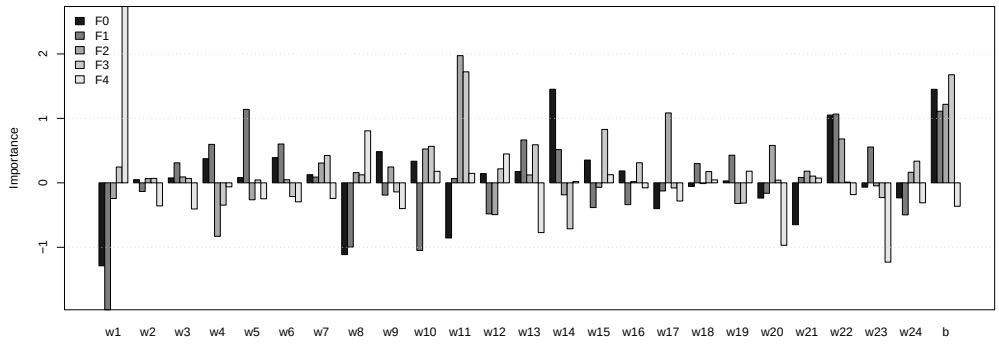

Figure 1: The coefficients discovered by the ESVM for each attribute and for each of the 5 classes. Longer bars (both positive and negative) correspond to more important attributes. Each group of 5 bars relates to the weights of one attribute.

Table 4: Example of an obtained confusion matrix of the SVM for the 5-degree fibrosis differentiation: Predicted outcomes on the rows, actual classes on the columns, number of correct predictions on the diagonal, misclassifications at intersections.

	True				
Predicted	$F0$	$F1$	$F2$	$F3$	$F4$
$F0$	0	0	0	0	
$F1$	7	47	12	10	1
$F2$	1	10	18	8	1
$F3$	0	0	0	2	0
$F4$	0	3	4	7	49

Table 5: Example of a resulting confusion matrix of the ESVM for the 3-degree fibrosis differentiation: Rows exhibit predicted outcomes, columns show actual classes. The diagonal outlines the amount of correctly classified samples, while each intersection of coordinates signals the number of misclassified cases.

	True		
Predicted	$F0-F1$	$F2-F3$	$F4$
$F0-F1$	50	31	1
$F2-F3$	9	38	2
$F4$	0	1	48

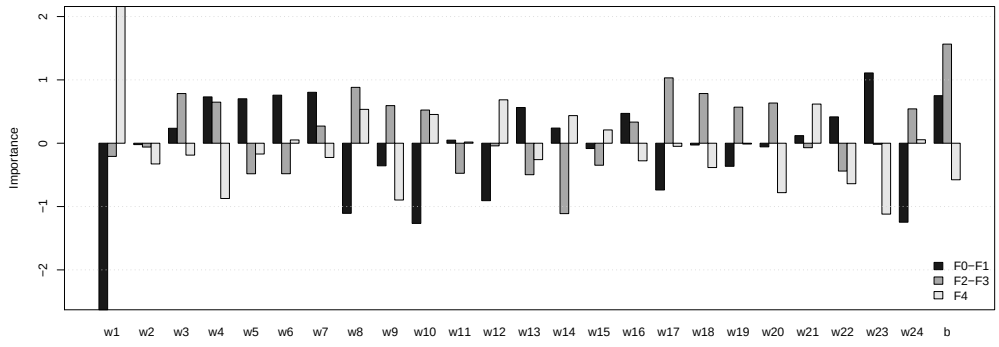

Figure 2: The coefficients discovered by the ESVM for each attribute and for each of the 3 classes. Longer bars (both positive and negative) correspond to more important attributes. Each group of 3 bars relates to the weights of one attribute.

Table 6: Example of an obtained confusion matrix of the SVM for the 3-degree fibrosis differentiation: Predicted outcomes on the rows, actual classes on the columns, number of correct predictions on the diagonal, misclassifications at intersections.

	True		
Predicted	$F0-F1$	$F2-F3$	$F4$
$F0-F1$	46	14	1
$F2-F3$	20	34	3
$F4$	4	3	55

Table 7: Example of an obtained confusion matrix of the GA-ESVM for the 5-degree fibrosis differentiation: Predicted outcomes on the rows, actual classes on the columns, number of correct predictions on the diagonal, misclassifications at intersections.

	True				
Predicted	$F0$	$F1$	$F2$	$F3$	$F4$
$F0$	0	0	0	0	0
$F1$	8	56	39	5	0
$F2$	0	1	3	3	1
$F3$	0	0	0	2	0
$F4$	0	0	0	3	59

Table 8: Example of a resulting confusion matrix of the GA-ESVM for the 3-degree fibrosis differentiation: Rows exhibit predicted outcomes, columns show actual classes. The diagonal outlines the amount of correctly classified samples, while each intersection of coordinates signals the number of misclassified cases.

	True		
Predicted	$F0-F1$	$F2-F3$	$F4$
$F0-F1$	52	29	0
$F2-F3$	8	34	3
$F4$	0	0	54

Table 9: Average accuracies and standard deviations in 30 trials of random cross-validation for the GA-ESVM, ESVM and SVM.

	Method		
Problem	GA-ESVM	ESVM	SVM
5-class	**64.07**±2.95	62.03±3.5	61.5±3.02
3-class	**74.76**±1.71	73.6±2.8	71±3.3

Table 10: When the ESVM is compared to the SVM, p-values from a t- and a Wilcoxon rank-sum test show a significant difference only in the case of the differentiation among 3 degrees of fibrosis, with the advantage on the ESVM side. The GA-ESVM is however significantly better than both the SVM and the ESVM for either 5 or 3 classes, as indicated by the same two tests.

	Tests for every two methods					
	ESVM/SVM		GA-ESVM/SVM		GA-ESVM/ESVM	
Problem	t-	Wilcox.	t-	Wilcox.	t-	Wilcox.
5-class	0.54	0.33	**0.001**	**0.001**	**0.01**	**0.05**
3-class	**0.002**	**0.002**	**1.72e-06**	**4.61e-06**	**0.05**	**0.05**

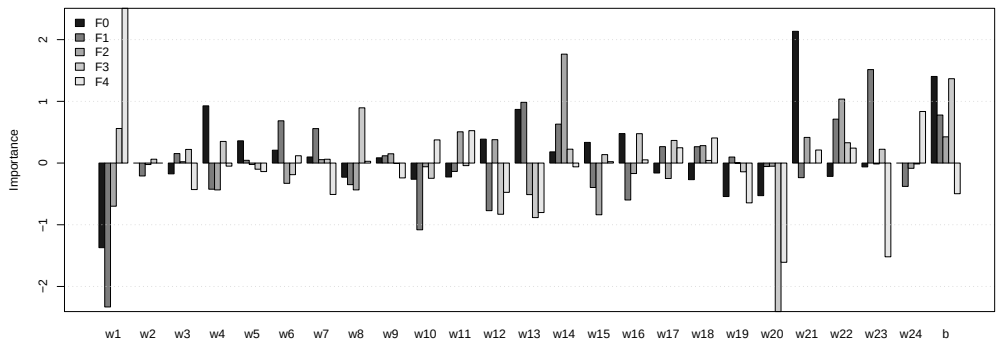

Figure 3: The coefficients discovered by the GA-ESVM for each attribute and for each of the 5 classes. Longer bars (both positive and negative) correspond to more important attributes. Each group of 5 bars relates to one attribute. Attributes omitted by the GA are represented with a zero weight, i.e. w_2 for $F0$ and $F4$.

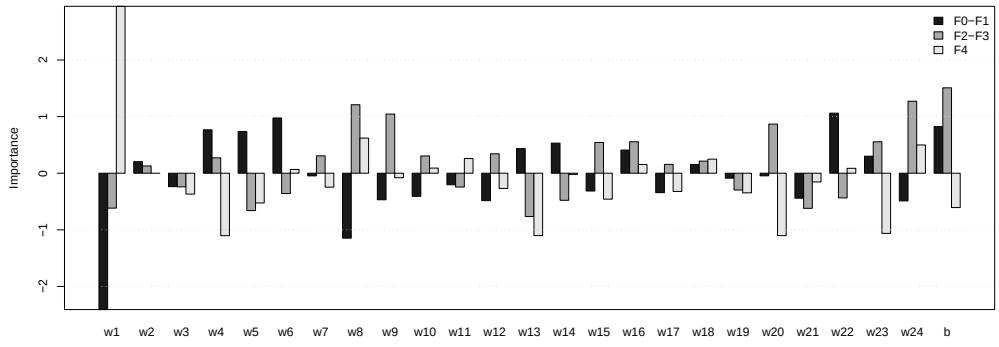

Figure 4: The coefficients discovered by the GA-ESVM for each attribute and for each of the 3 classes. Longer bars (both positive and negative) correspond to more important attributes. Each group of 3 bars relates to one attribute. Attributes skipped by the GA have the weight value set to zero.