Therapeutic Drug Monitoring of Kidney Transplant Recipients using Profiled Support Vector Machines

Gustavo Camps-Valls, *Member, IEEE,* Emilio Soria-Olivas, Juan J. Pérez-Ruixo,
Fernando Pérez-Cruz, *Member, IEEE,* Antonio Artés-Rodríguez, *Member, IEEE,* N. Víctor Jiménez-Torres

***Abstract*— This work proposes a twofold approach for therapeutic drug monitoring (TDM) of kidney recipients using Support Vector Machines (SVM), for both predicting and detecting Cyclosporine A (CyA) blood concentrations. The final goal is to build useful, robust, and ultimately understandable models for individualising the dosage of CyA.**

We compare SVM with several neural network models, such as the multilayer perceptron (MLP), the Elman recurrent network, FIR/IIR networks, and Neural Network ARMAX approaches. In addition, we present a profile-dependent SVM (PD-SVM), which incorporates *a priori* knowledge in both tasks. Models are compared numerically, statistically, and in the presence of additive noise. Data from fifty-seven renal allograft recipients were used to develop the models. Patients followed a standard triple therapy and CyA trough concentration was the dependent variable.

The best results for the CyA blood concentration prediction were obtained using the PD-SVM (mean error of 0.36 ng/mL and root-mean-square-error of 52.01 ng/mL in the validation set) and appeared to be more robust in the presence of additive noise. The propose PD-SVM improved results from the standard SVM and MLP, specially significant (both numerical and statistically) in the one-against-all scheme. Finally, some clinical conclusions were obtained from sensitivity rankings of the models and distribution of support vectors. We conclude that the PD-SVM approach produces more accurate and robust models than neural networks. Finally, a software tool for aiding medical decision-making including the prediction models is presented.

***Index Terms*— Cyclosporine, therapeutic drug monitoring, neural networks, support vector machines, sensitivity analysis, kidney transplantation.**

I. Introduction

Cyclosporine A (CyA) is still the cornerstone of immunosuppression in renal transplant recipients. This immunosuppressive drug shortens average hospital stays after kidney transplantation. At present, despite the appearance of new formulations, 90% of therapeutic guidelines are based on CyA and, consequently, costs continue to rise year after year[1]. Recently, important advances in dose formulation, therapeutic drug monitoring (TDM) and guidelines, and the emerging role of CyA-based combined therapies have resulted in a substantial improvement in clinical outcomes in renal transplant recipients [1]. Nevertheless, CyA is generally considered to be a critical dose drug. Underdosing may result in graft loss and overdosing causes kidney damage, increases opportunistic infections, systolic and diastolic pressure, and cholesterol. Moreover, the pharmacokinetic behaviour of CyA presents a substantial inter- and intra-individual variability, which appears to be particularly evident in the earlier post-transplantation period, when the risk and clinical consequences of acute rejection are higher than in stable renal patients [2]. Several factors such as clinical drug interactions and patient compliance can also significantly alter blood CyA concentrations and, thus, intensive TDM of CyA becomes necessary; however, it influences the patient's quality of life and the cost of the care.

Since the trough blood concentration has traditionally been used to monitor CyA therapy, mathematical models that are capable of predicting the future concentration of CyA and adjusting the optimal dosage become necessary. Population pharmacokinetic models and Bayesian forecasting have been used to predict CyA blood concentrations, but their performance was not optimal [3], [4]. These models predict plasma drug concentrations based on theoretical models of drug distribution and elimination but they often fail when the underlying principles are not sufficiently understood or known to be encoded into a set of relationships [5]. In fact, despite convincing results in many areas, few attempts have been made to use neural networks to predict drug behaviour [6]–[8]. A different approach to TDM of kidney recipients has recently been presented [9], in which the goal is the detection of subtherapeutic and toxic levels. This could aid physicians by providing "alarm signals" for risks that threaten patient evolution. Nevertheless, poor results were achieved regarding subtherapeutic levels detection, which could lead to dramatic kidney rejection processes.

All these limitations, in both prediction and classification methods, have led us to try to solve the problem of TDM using modern neural networks and Support Vector Machines (SVM),

Manuscript received December 2004;

G. Camps-Valls and E. Soria-Olivas are with Grup de Processament Digital de Senyals, GPDS. Dept. Enginyeria Electrònica. Escola Tècnica Superior d'Enginyeria. Universitat de València. C/ Dr. Moliner, 50. 46100 Burjassot (València) Spain. E-mail: gustavo.camps@uv.es.

F. Pérez-Cruz and A. Artés-Rodríguez are with Departmento de Teoría de la Señal y Comunicaciones, Universidad Carlos III de Madrid, 28911 Leganés, Madrid, Spain.

J. J. Pérez-Ruixo was with the Pharmacy Service at the Dr. Peset University Hospital, València (Spain), when this paper was prepared. Currently, he is with the Advanced PK/PD Modeling & Simulation. Global Clinical Pharmacokinetics and Clinical Pharmacodynamics Division. Johnson & Johnson Pharmaceutical Research & Development, a Division of Janssen Pharmaceutica N.V. Beerse (Belgium). E-mail: jperezru@prdbe.jnj.com.

Prof. N. V. Jiménez-Torres is with the Pharmacy and Pharmaceutical Technology Department, Universitat de València (Spain) and with the Pharmacy Service at the Dr. Peset University Hospital, València (Spain). E-mail: victor.jimenez@uv.es.

[1]The total cost in 1997 of CyA in Spain was €52 million and €12 billion in Europe.

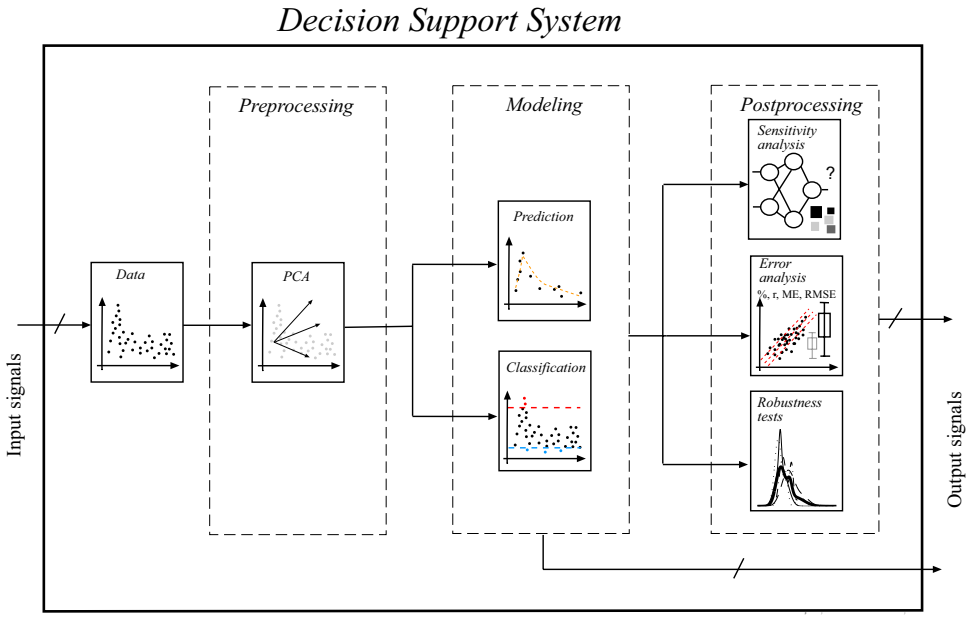

Fig. 1. *Learning scheme considered in this paper. Three main blocks are developed to obtain a validated decision support system (DSS). In the pre-processing stage, data are previously "whitened" through Principal Component Analysis (PCA). Afterwards, two approaches are considered for modeling; time series prediction of the CyA blood concentration and identification of subtherapeutic and toxic levels. After developing the models, we follow a methodology for extracting valid, novel and potentially useful knowledge from the model by using sensitivity analysis and by inspecting support vectors. In addition, we compare models through accuracy and robustness tests.*

which have shown good results in other fields. Therefore, two main objectives can be distinguished in this work:

- *Prediction of CyA trough levels.* This approach tries to predict values of CyA blood concentration from the previous values by following a time-series methodology. This same approach was followed in previous communications [10]–[12] but only short time series and a reduced population were available. In the present paper, we extend these works by developing more neural and kernel predictors using a larger database. We compare SVM with extensively used neural networks in other fields, such as the multilayer perceptron (MLP), Elman recurrent network, FIR/IIR networks, and Neural Network ARMAX (NNARMAX) approaches. Comparison is carried out in terms of accuracy by evaluating classical bias-variance trade-off measurements, and in terms of robustness by analysing performance when noise is introduced at input signals.
- *Prediction of CyA levels class.* Due to the high inter- and intra-subjects variability, non-uniform sampling, and non-stationarity of the time series, the difficulty of prediction task is well known. In fact, intensive TDM tries to keep CyA blood levels in the therapeutical range (usually established in the range 150-400 ng/mL) by making adjustments in the patient's drug regimen. Therefore, an alternative approach to the time-series prediction methodology consists of identifying future toxic ($>$400 ng/mL) and subtherapeutical levels ($\leq$150 ng/mL). Two different schemes are used for that purpose: one-against-all classifiers and multi-classifiers. We compare performance of SVM and MLP in both schemes with recognition rates and receiver-operating curves (ROC).

In addition, we present the so-called *Profile-Dependent* SVM (PD-SVM) to incorporate *a priori* knowledge in both the prediction and the classification models. The PD-SVM gives different confidence weights to different parts of the training data to focus the training on the *a priori* most important regions, the most recent samples for time series prediction, and samples near the decision thresholds for levels classification. Finally, we perform sensitivity analyses on the MLP and inspect the distribution of the support vectors in order to gain knowledge about the problem. This paper constitutes the natural extension of the works [12], [13], in which only a limited number of prediction approaches was carried out and no attention was paid to the structure of the "black-box" models. In this paper, a common methodology consisting of three basic steps is followed to develop a Decision Support System (DSS) for TDM based on modern techniques. A general learning scheme for our proposal is illustrated in Fig. 1. The basic characteristics of a general DSS are ensured in our case study as follows: (a) the system is robust, as it incorporates models that show good performance in noisy environments; (b) the system is accurate and unbiased, as models show good bias-variance trade-offs; and (c) the system incorporates *a priori* knowledge (profiled model) and allows the user to extract rules for dosage adjustment (sensitivity analysis). Rather than providing the clinician with dosage prediction, the system gives an estimation of future CyA blood concentration, alarm signals for subtherapeutic or toxic levels, feature ranking, and follow-up statistical information.

The rest of the paper is organized as follows. Data collection is introduced in Section II. Section III presents a review of the

predictive techniques used with special emphasis on SVM. The results are presented in Section IV. A discussion is provided in Section V along with some concluding remarks and a proposal for further work.

II. Patients and data collection

Sixty-seven renal allograft recipients treated in the Nephrology Service of the Dr. Peset University Hospital in the city of València (Spain) were initially included in this study. The exclusion criteria took into account patients with grave affection of other vital organs, active neoplasia or metastasis risk, active infections, presence of HIV virus, presence of urinary or vascular abnormalities, and patients older than 70. In addition, patients who did not fulfil the prescribed posology or who received metabolic inducers or inhibitors were excluded because they modify the pharmacokinetic profile of CyA.

Patients received a standard immunosuppressive regimen (triple therapy basis) with a microemulsion formulation of CyA (Sandimmun Neoral®), mycophenolate mofetil (2 g/d), and prednisone (0.5-1 mg/kg/d). The initial oral dose of CyA (5 mg/kg b.i.d) was reduced according to the measured CyA blood concentration and the desired target range (150-300 ng/mL) [14]. Steady state blood samples were withdrawn 12-14 hours after dose administration. CyA blood levels were measured by a specific monoclonal fluorescence polarisation immunoassay (Abbott, TDx), with inter- and intra-assay variation coefficients of less than 7.5% [15].

The study collected many potentially relevant variables of all monitored patients:

- *Anthropemetrical factors:* weight (Kg), age (years), and gender.
- *Biochemical factors:* urea (mg/dL), creatinine (mg/dL), creatinine clearance (mL /min), total protein and albumin (g/dL), bilirubin (mg/dL), cholesterol and triglicerides (mg /dL).
- *Hepathical enzymes:* aspartate aminotransferase (IU/L), alanine aminotransferase (IU/L), gamma-glutaril transpeptidase (IU/L), and alkaline phosphatase (IU/L).
- *Hemathological factors:* hematocrit (%), haemoglobin (g/dL), and leukocites (U /mm^3).
- *Clinical factors:* systolic and diastolic arterial pressure (mmHg).

At this point, two analyses were performed:

1) We detected 10 patients who contained more than 10% of statistical outliers (more than two standard deviations from the mean) in the original data distribution of many variables. In fact, when considering all available descriptors, 23% of samples were outliers. These patients were withdrawn for developing the models and hence, a final cohort of fifty-seven patients was considered.
2) At the same time, we analyzed the available data using classical statistical analyses (correlation analysis, normality plots, higher-order statistics), Principal Component Analysis (PCA), Self-Organising Maps (SOM) [16], Classification and Regression Trees (CART) [17], and Multivariate AutoRegressive Splines (MARS) [18] in order to get a preliminary subset of relevant features [19]. The obtained results showed a high degree of concordance with NONMEM and linear modeling [20], [21]. From this study, we obtained a reduced subset of eleven relevant patient factors that, along with dosage, CyA blood concentration, and post-transplantation days, were used to build the models. Some basic population statistics of these factors are shown in Table I. It is worth noting that almost all variables have significant non-normal distributions (z-values for skewness and kurtosis greater than ± 3.08, $p < 0.001$). This problem was addressed applying suitable transformations to raw data [22].

Two-thirds of the patients were used to train the models and the rest were used for their validation using the hold out method. Population was randomly assigned to two groups: 39 patients (665 patterns) were used for training the models and 18 patients (427 patterns) for their validation. This process was repeated until three basic conditions were met: (1) variables should take variations in mean and variance between training and validation sets lower than 15%; (2) each subset should contain an approximately similar proportion of male and female patients; and (3) patients who were monitored for longer period of time were assigned to the validation set. When the three conditions could not be fulfilled simultaneously, condition (3) was adopted and (2) relaxed. This three-step randomising methodology ensures balanced datasets to avoid population differences that could bias the models [23].

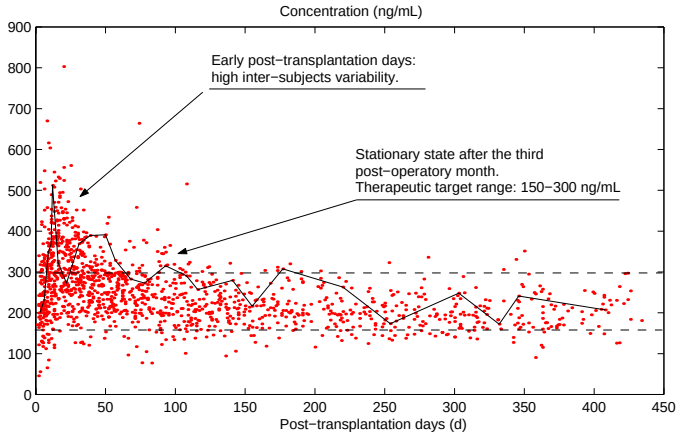

Fig. 2. Distribution of blood concentration (ng/mL) of CyA. The solid line represents a specific patient profile. Dashed lines indicate the desired (therapeutic) target range (150–300 ng/mL).

The health care team in the hospital decides daily the next dose to administer by assessing the patients' factors and their evolution. This protocol, nevertheless, produces three undesired features in the time series:

1) *High variability.* High inter-subject variability is found (coefficient of variation, CV=31%), especially remarkable in the early post-transplantation days (PTD) (CV = 43%). In this period, it becomes necessary to raise or lower dosage while closely monitoring the patient's concentration, as shown in Fig. 2.
2) *Non-stationarity.* Therapy tries to keep CyA levels in the target range in order to avoid nephrotoxicity or

TABLE I

CHARACTERISTICS OF PATIENTS IN THE STUDY FOR THE TRAINING AND THE VALIDATION (IN BRACKETS) SETS. RESULTS ARE PRESENTED AS MEAN ± STANDARD DEVIATION (SD) AND THE RANGE, EXCEPT FOR THE GENDER WHICH IS GIVEN AS THE NUMBER OF SUBJECTS. THE SHAPE DESCRIPTORS (KURTOSIS AND SKEWNESS) ARE CALCULATED OVER THE WHOLE POPULATION.

Variable	Range	Mean	Standard Deviation	Skewness (z-value)†	Kurtosis (z-value)†
♯ **Samples/patient**	5 - 30 (19 - 30)	19 (26)	8.42 (3.14)	-	-
Gender		21 (14) male and 18 (4) female			
Age (yr)	22.79 - 68.21 (28.82 - 69.43)	48.95 (51.28)	11.68 (10.46)	-0.26 (-3.49*)	2.29 (15.42*)
Post-transplantation days (d)†	1 - 434 (2 - 426)	100.31 (121.96)	106.34 (114.74)	1.08 (14.58*)	3.07 (20.74*)
CyA concentration (ng/mL)	55.69 - 615.98 (45.79 - 664.17)	249.93 (248.60)	78.95 (78.74)	0.95 (12.80*)	4.33 (29.21*)
Daily dosage/Weight (mg/Kg/d)	0.71 - 12.04 (1.88 - 9.45)	5.27 (4.47)	2.28 (1.87)	0.75 (10.09*)	2.90 (19.54*)
Urea (mg/dL)	17 - 332 (22 - 279)	75.94 (69.45)	49.55 (38.41)	2.25 (30.42*)	8.44 (56.95*)
Creatinine (mg/dL)	0.70 - 10.60 (1 - 10.20)	1.93 (2.03)	1.37 (1.22)	3.27 (44.14*)	15.38 (103.74*)
Creatinine clearance (ml/min)	7.77 - 117.42 (7.77 - 107.03)	50.44 (52.25)	20.96 (19.59)	0.04 (0.56)	2.56 (17.28*)
Alkaline phosphatase (IU/L)	62 - 683 (65 - 724)	195.93 (219.87)	85.54 (118.23)	1.96 (26.49*)	8.25 (55.65*)
Bilirubin (mg/dL)	0.22 - 6.10 (0.30 - 2)	0.88 (0.82)	0.63 (0.32)	3.66 (49.37*)	24.33 (164.12*)
Hematocrit, HTO (%)	22 - 55 (22 - 52)	35.50 (36.48)	6.18 (6.40)	0.32 (4.26*)	2.61 (17.58*)

† The z values are derived by dividing the statistics by the corresponding standard errors of $\sqrt{6/N}$ (skewness) and $\sqrt{24/N}$ (kurtosis). * Significant at the 0.001 level.

transplant rejection. This provokes the presence of non-stationary processes in the time series.

3) *Non-uniform sampling.* The individualisation procedure directly affects the sampling of the time series.

To deal with these problems, some issues should be taken into account. For example, non-stationarity was treated efficiently with dynamic neural networks. The problem of non-uniform sampling was initially addressed using the classical strategy of interpolation and resampling, but this produced overoptimistic results. Therefore, we decided to alleviate the problem by incorporating post-transplantation days into the model *globally*, i.e. no time-series differences were thus undertaken. Finally, some modifications on the training algorithms of the models were carried out, as will be shown in the next section.

III. METHODS AND EXPERIMENTAL SETUP

A. Static and dynamic neural networks

A common approach to time series prediction is the AutoRegressive Moving Average (ARMA) model. However, ARMA models are not suitable to our problem due to the non-linearity, non-uniform sampling, and non-stationarity of the time series. Hence, many researchers have turned to the use of non-linear models, such as neural networks, in which few assumptions must be made. The multilayer perceptron (MLP) is the most commonly used neural network, which is composed of a layered arrangement of artificial neurons in which each neuron of a given layer feeds all the neurons of the next layer. This model forms a complex mapping from the n-dimensional input to the binary output, $\psi : \mathbb{R}^n \longrightarrow \{0, 1\}$. For regression purposes, the MLP mapping has the form $\psi : \mathbb{R}^n \longrightarrow \mathbb{R}$. However, it is a static mapping; there is no internal dynamics [24], [25]. This problem can be easily addressed by including an array of unit-delay elements, called a tapped-delay line model, to make use of spatially-converted temporal patterns. This approximation was previously followed in [13], [26].

Neural Network ARMAX modeling is intimately related to the latter approach. In a NNARMAX model [27], given a pair of input-output discrete time series, a multilayer perceptron (MLP) is used to perform a mapping between them, in which past inputs, past outputs and past residuals can be fed into the input layer. Selecting a model structure is much more difficult in the nonlinear case than in the linear case (classical ARMA modeling). Not only is it necessary to choose a set of regressors but also a network architecture is required. Therefore, several regression structures are available. In this work, we use the ones that best fit to our interests; NNARMAX2 (the regression vector is formed by past inputs, past outputs and past residuals), NNSSIF (the regressor is in the form of state space innovations), and NNOE (the regression vector is formed by past inputs and estimations). All these models are extensively described in [27]. The use of NNARMAX models in control applications and nonlinear system identification has expanded in the past years. Its main advantage is the use of a non-linear regressor (usually an MLP) working on a fully tailored "state" vector [28]. This makes the model specially well suited to our problem because we can design the endowed input state vector to accommodate our non-stationary dynamics. This can be done, for example, by adding more "memory" in the form of error terms if we observe that prediction error is not completely exploited.

In addition, there are two more approaches for introducing dynamic capabilities into a static neural network:

1) *Synapses as digital filters.* To substitute the static synaptic weights for dynamic connections, which are usually linear filters. The FIR neural network models each synapsis as a Finite Impulse Response (FIR) filter [29]. There are striking similarities between this model and

the MLP. Notationally, scalars are replaced by vectors and multiplications by vector products. These simple analogies carry through when comparing standard back-propagation for static networks with *temporal backpropagation* for FIR networks [24]. FIR neural networks are appropriate to work in non-stationary environments or when non-linear dynamics are observed in the time series because *time* is treated naturally in the synapsis itself, that is, the networks have internal dynamics. In fact, they have demonstrated good results in problems with those characteristics, such as speech enhancement [30], and time series prediction [31].

In addition to the FIR network, we have used the gamma network [32], a class of Infinite Impulse Response (IIR) filter-based neural network which includes a local feedback parameter. In this structure, the FIR synapsis that uses the standard z-Transform delay operator z^{-1} is replaced by the so-called gamma operator

$$G(z) = \frac{\mu}{z - (1-\mu)}, \tag{1}$$

where μ is a real parameter which controls the memory depth of the filter. As pointed out in [33], gamma filters are theoretically superior to standard FIR filters in terms of number of parameters required to model a given dynamics. The filter is stable if $0 < \mu < 2$, and $G(z)$ reduces to the usual delay operator for $\mu = 1$. This filter also provides an additional advantage: the number of degrees of freedom (order K) and the memory depth remain decoupled, as shown in [33]. A proposed measurement of the memory depth of a model, which allows us to quantify the past information retained, is given by K/μ and has units of time. Hence, values of μ lower than the unit increase the memory depth of the filter. The use of the gamma structure in a neural network can be two-fold: we could substitute each scalar weight in an MLP with a gamma filter, or we could use a gamma unit delay line as the first layer of a classical MLP, which yields the so-called *focused gamma network*. The latter is the adopted approach in our work since it is more simple and allows us to scrutinize the needed memory depth for the problem by analysing the weights of this layer. The use of this network has focused attention for de-noising, communications, and time series prediction [34]. In general, the gamma network can deal efficiently with complex dynamics and lower number of network parameters, which could make it *a priori* well-suited to our problem since patient dynamics are local in the early post-transplantation days.

2) *Recurrent networks.* To construct loops in the connections between neurons or layers of the network. The Elman recurrent network is a simple recurrent model with feedback connections around the hidden layer. In this architecture, in addition to the input, hidden and output units, there are also context units, which are only used to memorize the previous activations of the hidden units [35]. The application of recurrent neural networks has traditionally been linked to applications such as speech and language processing. Additionally, Elman networks can result in efficient models to both detect or generate time-varying patterns [35], hence, its suitability to our problem.

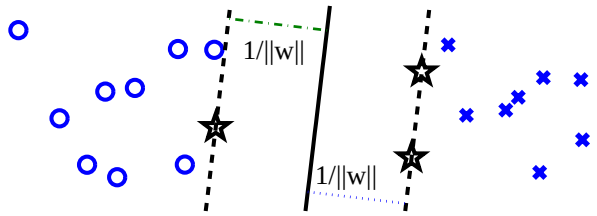

Fig. 3. The Optimal Decision Hyperplane (ODH) in a linearly separable problem. Maximizing the margin is equivalent to minimizing $\|\mathbf{w}\|$. Only support vectors (stars) are necessary to define the ODH.

Despite the fact that dynamic neural networks have been extensively used in areas such as signal processing and control, its use in biomedical engineering and medicine in general, and in TDM in particular, has received little attention. These networks are introduced here to deal more efficiently with time-varying patterns. However, some serious difficulties have been found in its application to our problem. The main limitations to the use of these networks are the need for long time series and for an unconstrained number of filter parameters through the networks. Also, the use of these methods to predict CyA blood concentration becomes more complicated since we develop individual models rather than population ones. This forces us to change the adaptation rules of the network when data coming from a new patient is presented to the networks in each iteration (epoch). In these situations, we updated the corresponding internal states (contextual neurons or filter coefficients in FIR/IIR synapses) of the network to the same parameters the patient had in the previous epoch and then we applied the usual updating rules. This procedure could be interpreted as a patient-based *batch* learning. This process produces oscillation of the training error, which was alleviated with a correct choice of initialisation parameters, and, in some cases, by using high values of the momentum term.

B. Support Vector Machines

Support vector machines (SVMs) are state-of-the-art tools for linear and nonlinear input-output knowledge discovery [36], [37]. The SVM was first proposed to solve nonlinear binary classification in [38]. Since then it has been extended to Regression [39] or to multiclass problems [36], among others. In the following, we revise the implementations used in this paper.

1) Support Vector Classifiers: Given a labeled training data set $\{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}$, where $\mathbf{x}_i \in \mathbb{R}^N$ and $y_i \in \{-1, +1\}$, and a nonlinear mapping $\phi(\cdot)$, usually to a higher (possibly infinite) dimensional (Hilbert) space, $\phi : \mathbb{R}^N \longrightarrow \mathcal{H}$, the SVM method solves:

$$\min_{\mathbf{w},\xi_i,b} \left\{ \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_i \xi_i \right\} \tag{2}$$

constrained to:

$$y_i(\boldsymbol{\phi}^T(\mathbf{x}_i)\mathbf{w}+b) \geq 1-\xi_i \qquad \forall i=1,\ldots,n \quad (3)$$

$$\xi_i \geq 0 \qquad \forall i=1,\ldots,n \quad (4)$$

where $\mathbf{w}$ and b define a linear classifier in the feature space. The non-linear mapping function $\boldsymbol{\phi}$ is performed in accordance with Cover's theorem, which guarantees that the transformed samples are more likely to be linearly separable in the resulting feature space (see Fig. 3). The regularization parameter C controls the generalization capabilities of the classifier and it must be selected by the user, and ξ_i are positive slack variables enabling to deal with permitted errors.

Due to the high dimensionality of vector variable $\mathbf{w}$, primal function (2) is usually solved through its Lagrangian dual problem, which consists of solving

$$\max_{\alpha_i}\left\{\sum_i \alpha_i - \frac{1}{2}\sum_{i,j}\alpha_i\alpha_j y_i y_j \boldsymbol{\phi}(\mathbf{x}_i)\cdot\boldsymbol{\phi}(\mathbf{x}_j)\right\} \quad (5)$$

constrained to $0 \leq \alpha_i \leq C$ and $\sum_i \alpha_i y_i = 0$, $i=1,\ldots,n$, where auxiliary variables α_i are Lagrange multipliers corresponding to constraints in (3). It is worth noting that all $\boldsymbol{\phi}$ mappings used in the SVM learning occur in the form of inner products. This allows us to define a kernel function K:

$$K(\mathbf{x}_i,\mathbf{x}_j) = \boldsymbol{\phi}(\mathbf{x}_i)\cdot\boldsymbol{\phi}(\mathbf{x}_j), \quad (6)$$

and then a non-linear SVM can be constructed using only the kernel function, without having to consider the mapping $\boldsymbol{\phi}$ explicitly. Then, by introducing (6) into (5), the dual problem is obtained. After solving this dual problem, $\mathbf{w} = \sum_{i=1}^{n} y_i\alpha_i\boldsymbol{\phi}(\mathbf{x}_i)$, and the decision function implemented by the classifier for any test vector $\mathbf{x}$ is given by

$$f(\mathbf{x}) = sgn\left(\sum_{i=1}^{n} y_i\alpha_i K(\mathbf{x}_i,\mathbf{x}) + b\right), \quad (7)$$

where b can be easily computed from the α_i that are neither 0 nor C, as explained in [40].

The SVM extension for multi-class problems is far from being unique and none of them seems to be superior to the others [41]. In this paper, we have used the one-against-all implementation and one-against-one approach. In the one-against-all approach, each class is compared with all the others together [42]. In the one-against-one scheme, the problem can be casted directly as a generalisation of the binary classification scheme [36], [43], [44].

2) Support Vector Regressor (SVR): The Support Vector Regressor (SVR) is the support vector implementation for regression and function approximation. Following the previous notation, SVR methods, find the minimum of:

$$\min_{\mathbf{w},\xi_i,\xi_i^*,b}\left\{\frac{1}{2}\|\mathbf{w}\|^2 + C\sum_i(\xi_i+\xi_i^*)\right\} \quad (8)$$

with respect to $\mathbf{w}$, ξ_i, ξ_i^* and b, subject to:

$$y_i - \boldsymbol{\phi}^T(\mathbf{x}_i)\mathbf{w} - b \leq \varepsilon+\xi_i, \quad \forall i=1,...,n \quad (9)$$

$$\boldsymbol{\phi}^T(\mathbf{x}_i)\mathbf{w} + b - y_i \leq \varepsilon+\xi_i^*, \quad \forall i=1,...,n \quad (10)$$

$$\xi_i,\xi_i^* \geq 0, \quad \forall i=1,...,n \quad (11)$$

where now $y_i \in \mathbb{R}$, and ξ_i, ξ_i^* and C are, respectively, positive slack variables to deal with training samples with a prediction error that is larger than ε ($\varepsilon > 0$) and the penalisation applied to them. The usual procedure for solving SVRs introduces the linear restrictions using Lagrange multipliers, computes the Karush-Kuhn-Tucker conditions and solves Wolfe's dual problem using quadratic programming procedures [45], [46].

3) The profile-dependent SVM (PD-SVM): It is a common practice in classification problems with unbalanced classes to set different penalisation factors for each class, which are usually proportional to the class priors [47], according to fuzzy-rules [48], or posteriors [49]. This approach combats the presence of false positives, which obviously produces "unbalanced" models. This problem is specially relevant in bioengineering applications and could prevent models from their use in a real clinical environment. This way, the SVM learns to classify patterns independently from the class they belong to (we have control over its learning); something that is not possible when using an overall penalisation C. In the field of time series prediction, this approach can be extended by considering that the most recent samples contain, in principle, more information. Therefore, problems with non-stationary processes can be alleviated using a different penalisation factor (or insensitivity zone) for each training sample t according to a certain *confidence function* C_t on the samples. This allows the regression machine to follow, in principle, the probability density function variations over time [50], [51]. In this paper, we tailor specific profiles for each problem:

- *Prediction of CyA trough levels.* This approach was previously presented in [51] for CyA level prediction. In [13], profiles were defined in terms of clusters rather than fixed *a priori* for the same problem and, in [52], we designed profiles for another complex pharmacokinetic problem. These experiences suggested tailoring profiles based on exponential memory decay functions, which has also been observed by other authors [53], [54]. Therefore, a good practice is to consider an exponential memory decay based on the confidence of past samples:

$$C_t = \lambda^{t_n - t_i}, \ \lambda \in [0,1] \quad (12)$$

  where t_n is the actual time sample and t_i is the time instant for sample i. This profile reduces the penalisation parameter and enlarges the ε-insensitive region of previous samples as new samples are obtained.
- *Prediction of CyA levels class.* The same approach can be used in a classification task by increasing penalisation near the decision borders (150 ng/mL and 400 ng/mL) to avoid false detections. With regard to the formulation, one has only to substitute the standard penalisation parameter C with a time-dependent penalisation C_t, as follows:

$$C_{150,t+1} = \frac{1}{150}[k_2C_t + k_1C_{t-1} + k_o] \quad (13)$$

$$C_{400,t+1} = \frac{1}{400}[k_2C_t + k_1C_{t-1} + k_o] \quad (14)$$

  With this approach, we intuitively increase the penalisation of errors as we approach the decision border. The additional penalisation factors k_i can be fixed *a priori* or

computed in an adaptive way by taking advantage of the Iterated Re-Weighted Least Squares (IRWLS) procedure [50]. In our application, we only considered a heuristic approach in which several combinations were tested. Further work will consider refined updating rules.

The inclusion of a temporal confidence function in the SVM formulation offers some advantages. Essentially, the overall number of SV remains constant through time and better results are obtained when abrupt changes appear in the time series, as demonstrated in [50], [54]. In addition, note that the computational complexity of the proposed method is the same than for the standard SVM formulations, since the functional and the number of constraints is the same. The only shortcoming in the PD-SVM is the design of the confidence function (Eq. (12) for regression or Eqs. (13)-(14) for classification). In our previous experience, the inclusion of *a priori* meaningful profile functions alleviated this restriction, resulting in more elegant solutions to the problem [52], [55], [56].

IV. Results

This section is organized as follows. First, we include a detailed description about pattern building, which serves equally well both for prediction and classification. Then, we show and discuss results for the prediction of CyA levels. Afterwards, we compare MLP and SVM for the identification of subtherapeutic and toxic levels. Finally, we perform sensitivity analysis of the MLP and analyze the SV distribution.

A. Building the input-output data

In order to develop a model, the input patterns from the available data for each patient must be previously built. Both for the problem of time series prediction and classification, we built the input pattern following a time series methodology. For this purpose, we tested several sizes of the time window (from one to five post-operatory days) [19]; however, given that prediction in the early post-transplantation period is strictly necessary, time-series were lagged by only one sample. Therefore, an input pattern for a given patient contains two samples from each variable in Table I (except for the gender) at time t and $t-1$, resulting in a total input dimension of 21 (10×2 + 1). This scheme allows a prediction in the first four post-operatory days as a mean and, as a result, yields a fixed unique model.

The prediction task consisted on estimating the value of CyA blood concentration at time $t+1$ from the input pattern. The classification task used the same input pattern to identify the range of concentration at $t+1$. More details on this scheme are given in Section IV-C.

B. Prediction of CyA levels

1) Model development: With regard the MLP and NN-ARMAX models, we varied the number of hidden neurons (< 20 to avoid overfitting), the initialisation of weights and the learning rate (between 0.01 and 1) in order to determine the best topology. We also penalized committed errors larger than 50 ng/mL by a penalization factor P, which was varied from 1 to 3. Additionally, we left weights unaltered if the committed error for a pattern was below an error threshold ($\varepsilon < 25$ ng/mL). If not, we proceeded to the application of the typical equations of the on-line back-propagation (BP) learning algorithm. In terms of statistical learning, the latter approach can be referred to as the quadratic ε-insensitive cost function. These modifications produced higher recognition rates than the MLP trained with the standard BP and, in turn, reduced the computational burden.

The training process for the FIR network was difficult because of its complexity since the number of free parameters increases geometrically with the number of inputs, as shown in [29]. In order to obtain accurate models, a great many sweeps were performed, varying the number of hidden neurons (from 2 to 25), the number of taps per synaptic connection in a layer (from 1 to 2) and the learning rate (typically between 0.0001 and 0.01). Initialisation of the weights depending on the structure of the net was also used, as proposed by many authors [25], [29]. Models with few taps ($<$4) and long training (number of epochs $>$10000) were necessary to attain satisfactory results. The training of the Elman network depends to a large extent on the learning rate and the number of context neurons. In fact, the network must be trained using high values of the momentum term ($\alpha > 0.8$) because this prevents weight oscillations in the training process.

In the case of SVMs, we tested linear, polynomial and Radial Basis Function (RBF) kernels to obtain the SVR solutions. There are many reasons to select the RBF kernel *a priori*: it has less numerical difficulties than linear, polynomial or sigmoid kernels, and only the Gaussian width has to be tuned. In addition, sigmoid kernels are non-positive definite kernels in all situations, which precludes their practical application [40], [57], [58]. Note that one or more free parameters must be previously settled in the nonlinear kernels. We used exponentially increasing sequences of C ($C = 10^{-2}, 10^{-1}, \cdots, 10^{6}$), and σ ($\sigma = 1, \cdots, 50$). The tube size ε was tuned linearly ($\varepsilon = 0, \cdots, 0.75$). The polynomial order d was varied in the range 1 to 8, as suggested in the literature [59]. An additional parameter for the PD-SVR was λ for the exponential memory decay, which was varied from 0.70 to 1 in steps of 0.05. During the development of the models, the data were pre-processed to give zero mean and unit variance. All models were developed in MATLAB® environment (Mathworks, Inc). Since the computational burden was high, *m-files* were translated to MEX-files and the programs run on a Pentium III (1.8GHz) with 256MB RAM.

The criterion used to select a candidate model for the final system was based on the model predictive performance in the validation data set. Bias was measured using the mean error (ME). The root-mean-square error (RMSE) was used as a measure of precision. In addition, we measured blood levels accurately predicted (%BLAP) if an error margin of 20% is fixed, as proposed in [3], [4]. We used the mean of the absolute prediction error to compare the precision of the models using the one-way ANOVA method. The results were also assessed by inspecting the correlation coefficient (r) as a measure of goodness-of-fit. The model accuracy was tested by using the one-way analysis of variance (ANOVA) method.

TABLE II

BEST RESULTS FOR THE CYA BLOOD CONCENTRATION PREDICTION OF ALL THE MODELS. THE ROOT-MEAN-SQUARE ERROR (RMSE), THE MEAN ERROR (ME) AND THE CORRELATION COEFFICIENT (r) ARE GIVEN. BLOOD LEVELS ACCURATELY PREDICTED (20%BLAP) ARE ALSO SHOWN. 95% CI ARE GIVEN IN BRACKETS, WHICH WERE CALCULATED USING BOOTSTRAP METHODS FOR THE CASE OF RMSE.

Model	**r**	**ME (ng/mL) ±95% CI**	**RMSE (ng/mL) ±95% CI**	**BLAP20% (%)**
MLP	0.74	3.23 (-1.76,8.22)	52.65 (45.28,59.11)	69.79
NNARMAX2	0.74	2.33 (-2.72,7.38)	53.25 (46.52,59.22)	69.79
NNOE	0.76	2.23 (-2.61,7.08)	51.11 (44.65,56.84)	70.49
NNSSIF	0.76	2.24 (-2.63,7.10)	51.27 (44.79,57.03)	70.49
ELMAN	0.74	0.30 (-4.71,5.31)	52.72 (46.03,58.66)	70.02
FIR	0.79	4.25 (-0.39,8.89)	49.06 (42.82,54.61)	73.30
GAMMA	0.74	4.62 (-0.42,9.67)	53.33 (46.54,59.36)	70.49
SVR	0.74	0.38 (-4.68,5.44)	53.28 (46.43,59.34)	69.79
PD-SVR RBF kernels†	0.75	0.36 (-4.57,5.33)	52.01 (45.22,58.11)	71.20

† The distribution was divided into three monitoring regions (period ♯1: 0–60 days, period ♯2: 60–160 days, period ♯3: 160–400 days) and a different PD-SVR was applied in each of them. Both ε and C were tuned using an exponential memory decay: $\lambda_1 = 0.98$, $\lambda_2 = 0.995$, $\lambda_3 = 0.998$.

2) Numerical analysis: Table II shows results in the prediction task for all models in the validation set. Results do not yield significant numerical differences between models in terms of accuracy (RMSE) or goodness-of-fit (r) (see Table II). A better performance is achieved with PD-SVR, NNOE, NNSSIF, and FIR models regarding accuracy and success rates. The PD-SVR method performs better than MLP, FIR, and the standard formulation of the SVR. The less biased models are the Elman recurrent network (0.30 ng/mL), SVR (0.38 ng/mL), and PD-SVR (0.36 ng/mL) methods. In addition, the size of the difference between CI95% is similar for all models, but SVR and PD-SVR keep them symmetrically distributed. An ANOVA test shows that no statistical differences in bias (F=0.6292, p=0.7540) or accuracy (F=0.3912, p=0.9259) are obtained.

Figure 4 shows predicted-versus-observed and predicted-versus-residuals plots in the validation set for the PD-SVR, which yields the best compromise between accuracy and bias of the estimations. Good determination coefficients ($r^2 = 0.78$) and negatively biased estimations (linear regression; slope±IC95%: 1.112±0.097, intercept ± IC95%: -27.524±24.755) are observed. Residuals do not indicate any trend. As the figure shows, all the models capture abrupt changes in the time series of CyA blood concentration in the

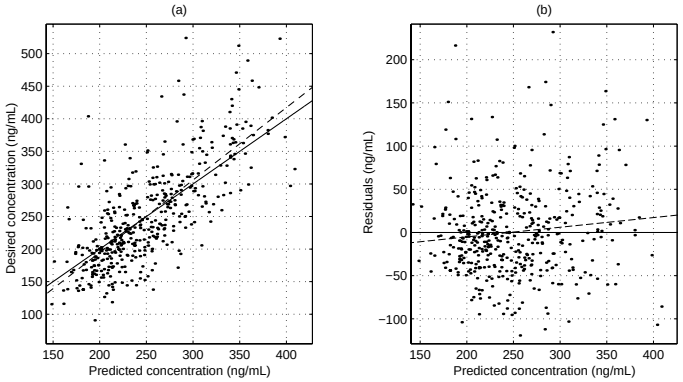

Fig. 4. PD-SVR performance in the validation set for the CyA blood concentration prediction. The solid line represents the line of identity, and the dotted one is the regression line. (a) Predicted versus observed dosages and (b) predicted versus residuals, in which the slope of the linear regression is a measure of the expected systematic concentration-related deviation in the data.

patients and perform similarly in the early post-transplantation period. Nevertheless, three patients have very poor predictions (RMSE $>$ 60 ng/mL) which could be due to errors in drug dosage administration, to the inter- and intrasubject variability in the drug absorption process, to the recording of blood sampling times, or abrupt changes in each patient's clinical condition. A linear dependency on interindividual variability could explain those results. In fact, patients with highly biased estimations are precisely those with high values of coefficient of variation (CV$>$30% on the whole series and CV$>$45% in the early post-transplantation month). When these patients are not considered, results improve drastically (r=0.78, ME=0.27 ng/mL, RMSE=49 ng/mL, BLAP20%=75%).

As an example of these situations, we show evolution (CyA dosage and blood concentrations) of three patients with good (Fig. 5a), acceptable (Fig. 5b) and poor (Fig. 5c) predictive performances. In the same figure, we show CyA concentration predictions of MLP, FIR network, and the SVR method. We usually obtain good predictions (RMSE$<$40 ng/mL) in patients under 50 years old, with total body weight higher than 50 Kg and with low number of subtherapeutic or toxic CyA blood levels ($<$10%). We have observed average over-estimations (ME$>$0) in patients with accurate predictions even in the early post-transplantation days or when receiving moderate doses. Abrupt changes in the time series produce appreciable loss in the prediction, as shown in Fig. 5b, where in two weeks, a toxic level is reached with a moderate dosage. Even with these difficulties, models efficiently capture the dynamics in the first post-operatory month. The poor results obtained in Fig. 5c can be mainly due to the presence of abrupt changes in the first post-transplantation month. These poor predictions, even with moderate initial doses ($<$8 mg/Kg), could be due to the higher variability in the absorption and disposition process of CyA during the first four weeks of post-transplantation. In fact, under-prediction in this period is a common characteristic of the models for almost all the patients, and the prediction of CyA blood concentration presents serious difficulties. This problem is lessened as long as blood levels become more stable, which is when slight over-predictions are obtained (see

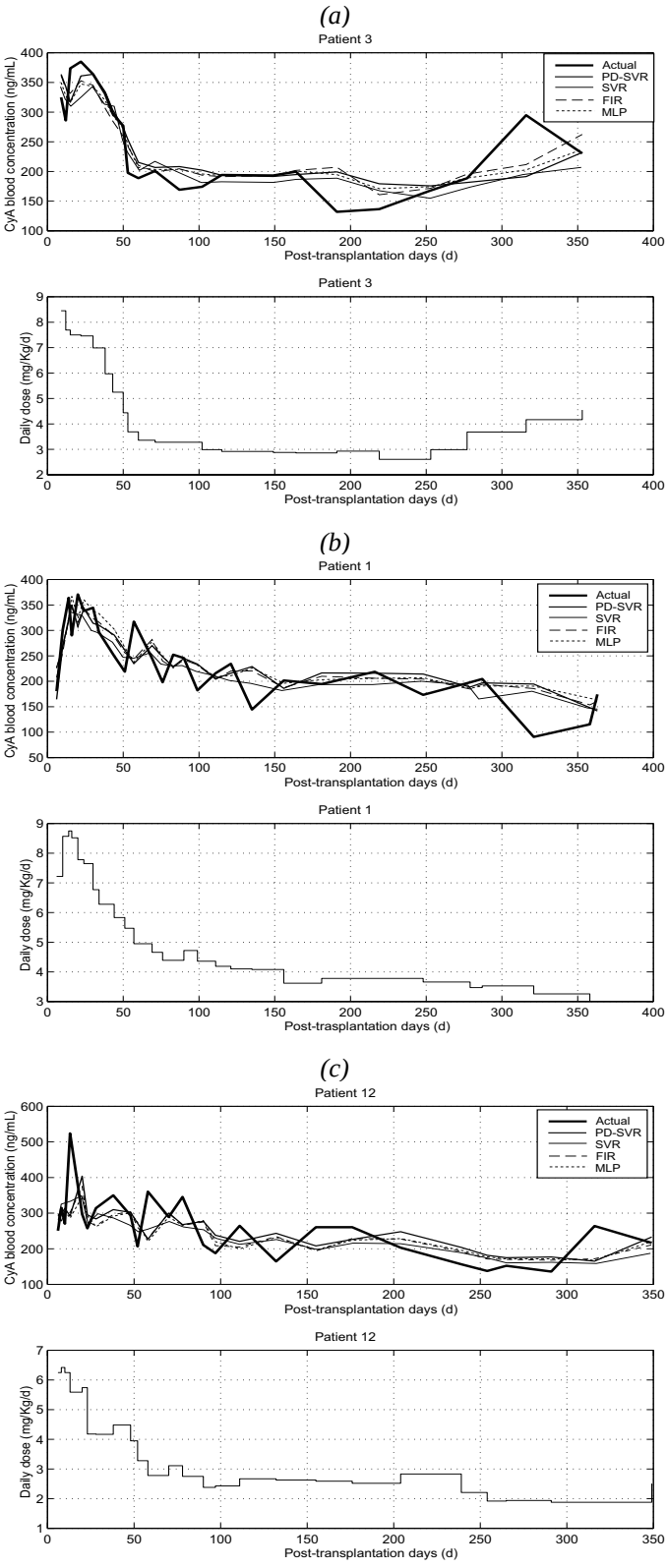

Fig. 5. Plots of evolution of three individual patients showing (a) good (RMSE=34.8 ng/mL), (b) acceptable (RMSE=42.9 ng/mL), and (c) unsatisfactory (RMSE=73.1 ng/mL) predictive performance. For each patient, the upper panel shows observed (thick solid line) and predicted CyA blood concentration (ng/mL) using PD-SVR (thin solid line), SVR (thinner solid line), FIR (thin dashed line), and MLP (thin dotted line). In each bottom panel, the oral dose (mg/Kg/d) versus post-operatory day is represented for proper analysis.

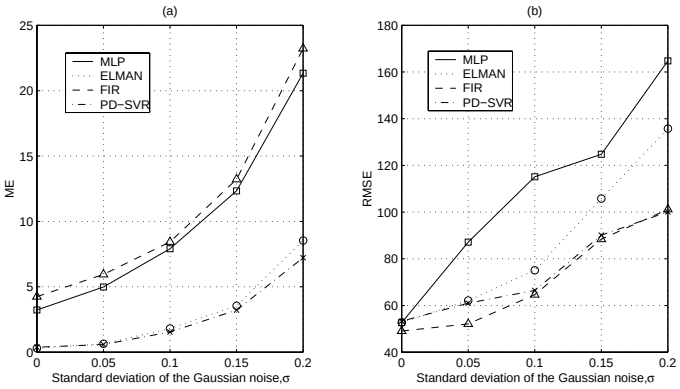

Fig. 6. Evaluation of the (a) (absolute) mean error and the (b) RMSE measurement when additive Gaussian noise with zero mean and standard deviation σ is introduced in the predictive models. Results refer to the validation set and were repeated 100 times, which represents a reasonable confidence margin for the measurements.

patients in Fig. 5a and 5b).

Since no numerical or statistical differences were observed between the neural and kernel models, we decided to test their robustness by introducing additive noise at models inputs. This can simulate situations such as blood sampling errors, patient compliance and the sensitivity of the model to exact input values. This process was tested with the most precise and unbiased models (ELMAN, FIR, PD-SVR) and the classical MLP network. In Fig. 6, we show the performance (bias and accuracy) in models when different levels of noise variance (σ) are introduced. Both measurements increase as the noise level is increased. However, as σ is increased, PD-SVR shows an excellent behaviour regarding bias and accuracy. We conclude that PD-SVR offers excellent robustness capabilities when low noise levels are introduced (σ <0.05), which indicates less sensitivity to exact input values in normal situations. Certainly, regularisation not only provides smoother solutions but also improves stability of predictions. This issue has been extensively demonstrated in the literature in general, and in our problem in particular.

C. Levels identification

Even though the previous forecasting models are accurate, they do not capture abnormal CyA levels (only 5% of toxic levels and 4% of subtherapeutic levels are correctly predicted), and thus they would not aid in preventing nephrotoxicity or transplant rejection. An alternative approach consists in predicting whether the next CyA blood level increases (decreases) to a toxic (subtherapeutic) level. With such an approach, the prediction task becomes a classification problem with three classes (CyA levels <150, [150, 400] or >400 ng/mL) and thus $N_c = 3$. For this purpose, we developed two schemes: (1) the one-against-all classification scheme, in which each of the three binary classifiers is trained to distinguish the samples in a given class from the samples in the two remaining classes; and (2) the one-against-one scheme, in which $N_c(N_c-1)/2$ binary classifiers are developed to distinguish a pair of classes.

1) Model development and comparison: In this task, two approaches were considered; one-against-all and one-against-

TABLE III
CONFUSION MATRICES IN THE VALIDATION SET OF THE MLP (IN BRACKETS), SVM (ITALICS) AND PD-SVM (BOLD FACE) MODELS FOR PREDICTING TOXIC AND SUBTHERAPEUTIC LEVELS USING ONE-AGAINST-ALL (TOP) AND ONE-AGAINST-ONE SCHEMES (BOTTOM)†.

ONE-AGAINST-ALL CLASSIFICATION SCHEME			
	Actual CyA levels [ng/mL]		
Prediction	**<150**	**[150,400)**	**≥400**
<150 ng/mL	**24** *24* (21)	**78** *123* (92)	**2** *2* (1)
[150,400) ng/mL	**1** *1* (4)	**264** *194* (244)	**2** *2* (2)
≥400 ng/mL	**1** *1* (1)	**42** *67* (48)	**13** *13* (14)
ONE-AGAINST-ONE SCHEME			
	Actual CyA levels [ng/mL]		
Prediction	**<150**	**[150,400)**	**≥400**
<150 ng/mL	**25** *24* (24)	**143** *150* (144)	**0** *0* (1)
[150,400) ng/mL	**1** *2* (2)	**184** *170* (173)	**1** *1* (1)
≥400 ng/mL	**0** *0* (0)	**57** *64* (67)	**16** *16* (15)

† The best results for the PD-SVM were obtained using k_2=1, k_1=0.3, k_o=0 for identifying toxic levels, k_2=-1, k_1=-0.4, k_o=600 for subtherapeutic levels, and k_2=1, k_1=0.12, k_o=0.05 in the one-against-one scheme.

one classifiers. With regard to the MLP models, we varied the number of hidden neurons (< 20 to avoid overfitting), the weight initialisation range, and the learning rate (between 0.01 and 3) in order to determine the best topology. We tested linear, polynomial, and Radial Basis Function (RBF) kernels to obtain the SVM solutions. The same ranges as the ones presented in Section IV-B.1 were used. An additional parameter for the PD-SVM was k_i, which were heuristically tuned.

In a multiclass problem, one usually optimizes a *global* measurement of model performance, such as the overall success rate (SR[%]). However, in unbalanced datasets, it is more useful to pay attention to the sensitivity/specificity in order to avoid skewed results. In our case, models were selected by evaluating the average of the sensitivity (SE) and specificity (SP) factors obtained in the *classes of interest* (toxic and subtherapeutic levels), since the final goal is to obtain a highly sensitive, robust classifier capable of providing 'alarm signals' for patient's monitoring. The sensitivity determines the percentage of true results that are correctly classified, e.g. CyA levels greater than 400 ng/mL correctly classified, and specificity determines the percentage of false results that are correctly classified, e.g. greater than 150 ng/mL correctly classified. Since the distribution of classes is highly unbalanced (4.95% of the cases are subtherapeutic levels and 5.77% of the cases are toxic levels), committed errors were systematically penalized according to the number of cases in each class [47]. In the case of the PD-SVM, the former priors-based penalisation multiplies the penalisation provided by the specific profile (Eqs. (13) and (14)).

2) Numerical comparison: Table III[top] shows one-against-all models performance. High rates of the (SE+SP) criterion ($> 81\%$) were obtained for all classifiers. This result enables its use as sensitive models for the real clinical practice. Specifically, the SVM classifier yields a higher sensitivity ratio than MLP (size of the difference is 11.54%) but it yields a lower recognition rate (size of the difference is 5.88%) when predicting subtherapeutic levels. In any case, sensitivity is much better than in a previous work by Hirankarn et al. [9], in which accuracy in subtherapeutic ranges was about 62%. This issue has been addressed by using the PD-SVM, which improves results of the standard SVM and MLP models, especially significant for subtherapeutic level detection.

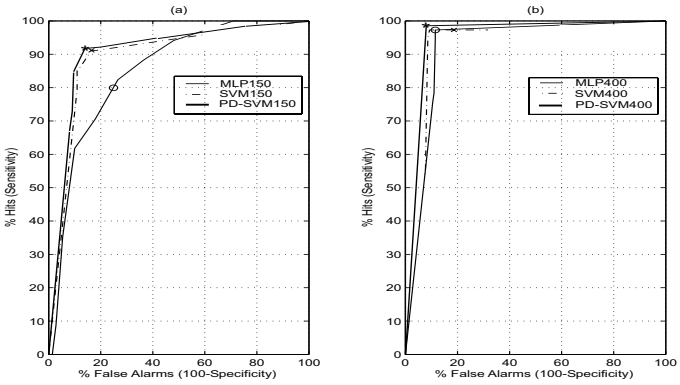

Fig. 7. Receiver operating characteristic (ROC) curve of the MLP and the SVM methods in the validation set when used as dedicated prediction models of (a) subtherapeutic levels and (b) toxic levels. The decision limit γ was varied throughout the output range [–1,+1] to obtain this curve. Circles (MLP), crosses (SVM) and stars (PD-SVM) represent the origin of the decision limit (γ=0).

In order to assess models' performance, we also calculated the area under a receiver operating characteristic (ROC) curves for the dedicated classifiers of toxic and subtheratpeutic blood levels (Fig. 7). The plot of false alarms versus hits provides a useful way to compare models for a wide range of 'levels of discrimination', and thus it has become a traditional method for model assessment. In our models, the area under the ROC curve (AUC) is higher with the PD-SVM method in detecting both subtherapeutic (MLP150: 76.47%, SVM150: 82.77%, PD-SVM150: 87.43%) and toxic levels (MLP400: 89.83%, SVM400: 95.11%, PD-SVM400: 97.45%). In addition, the PD-SVM reduces the number of false positives for both models (see Table III[top]) and the number of SVs (28% against 35% of the whole training set for the standard formulation). An additional comment refers to the distance between discrimination levels. When large jumps from one level of discrimination to another are found in these curves, a lack of knowledge of the classifier's behaviour in that area is present. In this sense, it is worth noting that very similar curves are observed for toxic level models (Fig. 7b), but (slightly) lower confidence can be obtained by using an MLP for predicting subtherapeutic levels (Fig. 7a). Similar results are observed between the SVM and its profiled version. The latter conclusions can also be observed by inspecting the false alarm rate in Table III[top]. In general, the ROC curves relative to the PD-SVM are above and on the left compared with the ROC curves relative to the other models.

Table III[bottom] shows the one-against-one confusion matrix. Several conclusions can be drawn: (1) The PD-SVM method yields better results than the rest, specificially the MLP, with a raise of 2.23% in terms of SR[%] and 6.25%

in terms of SE+SP; (2) a dramatic error in classifying toxic patterns is committed by the MLP; (3) the PD-SVM improves sensitivity of detection of subtherapeutic levels and drastically reduces the misclassification rate of therapeutic levels; and (4) once again model complexity is reduced with the profile-dependent technique, by which we obtain a mean reduction of 4% of SVs *per* class.

3) Statistical comparison: As we did in the prediction approach, we have analyzed the numerical but also statistical differences among classifiers and schemes. For this purpose, we have computed the statistical pairwise comparison of two classifiers through Z-scores [60]. In general, SVM-based models give better performance in terms of sensitivity, but MLP is (slightly) better in specificity. In particular, PD-SVM yields better (SE+SP) scores so it is more sensitive, specific and balanced classifier in all schemes. Statistical tests yielded Z scores higher than 1.96 for all classifiers, and thus results are significant and classifications are better than random choice. An interesting result is that the PD-SVM and MLP are preferred statistically when working in one-against-all schemes ($Z_{\text{PD-SVM}} = 5.38$, $Z_{\text{SVM}} = 3.00$, $Z_{\text{MLP}} = 4.12$) than in one-against-one schemes ($Z_{\text{PD-SVM}} = 3.00$, $Z_{\text{SVM}} = 2.60$, $Z_{\text{MLP}} = 2.62$), in which no appreciable differences appear. Performing pairwise statistical comparisons, one can conclude that only the PD-SVM is significantly different than the other classifiers in one-against-all scheme, and no statistical differences appear in the one-against-one schemes. These results match the ones shown in [61], in which the authors pointed out that, in some ocassions, a one-against-all scheme can be as accurate as any other.

D. Models analysis

Knowledge discovery is defined as "the process of identifying valid, novel, potentially useful, and ultimately understandable structure in data" [62]. The scientific community is not only searching for methods that provide accurate estimations of the underlying system function, but for methods that also explain those complex, and often non-linear, relationships from the input-output mapping performed by the models. In this paper, sensitivity analyses for the MLP and insight on SV distribution for the PD-SVR have been used in order to gain knowledge about the problem.

1) Sensitivity analysis: Sensitivity analysis is used to study the influence of input variables on the dependent variable and consists of evaluating the changes in training error that would result if an input were removed from the model. This measure, commonly known as *delta error* in the literature, produces a valuable ranking of the relevance of the variables. Two additional sensitivity measures, which are based on perturbing an input and monitoring network outputs, can be computed: the *Average Gradient (AG)* and the *Average Absolute Gradient (AAG)*. All these measurements are extensively described in [63].

In Table IV, different rankings in accordance with these measurements are shown for the MLP. Only the top seven relevant inputs are shown. Several conclusions can be drawn. The most informative variables considered by the model are past value of dosage, CyA blood concentration and creatinine level. In a second level of relevance, we find the post-transplantation days along with past hematocrit and creatinine clearance levels. By analysing the sign of DE and AAG we can conclude that, on average, an increase in past dosage produces an increase in future CyA blood concentration, which indicates that model captures correctly this issue. On average, lower creatinine levels are associated to an increase in CyA blood concentration. These results agree with those obtained when other machine learning approaches [19] and NONMEM modeling (ANOVA and univariate analysis methods) [20], [21] were used.

TABLE IV
RANKING OF INPUT VARIABLES ACCORDING TO THE DELTA ERROR (DE), AVERAGE GRADIENT (AG) AND AVERAGE ABSOLUTE GRADIENT (AAG) MEASUREMENTS FOR THE BEST MLP. THE MOST RELEVANT INPUT VARIABLES ARE DAILY DOSAGE (DD), CYA BLOOD CONCENTRATION (C), CREATININE (CR), POST-TRANSPLANTATION DAYS (PTD), AND HEMATOCRIT (HTO) FOR POST-TRANSPLANTATION DAYS t AND $t-1$.

♯	Variable	DE	Variable	AG	Variable	AAG
1	DD(t)	1.00	DD(t)	1.00	DD(t)	1.00
2	C(t)	0.90	C(t)	0.93	C(t)	0.94
3	DD(t-1)	-0.44	DD(t-1)	-0.41	DD(t-1)	0.41
4	PTD(t-1)	0.18	CR(t)	-0.34	CR(t)	0.34
5	C(t-1)	0.14	C(t-1)	0.25	C(t-1)	0.26
6	PTD(t)	0.10	PTD(t-1)	-0.25	PTD(t-1)	0.25
7	HTO(t-1)	0.10	HTO(t-1)	-0.24	HTO(t-1)	0.24

2) Distribution of the Support Vectors: SVMs have demonstrated to be well-suited techniques in classification and regression tasks. An additional advantage also arises from their use: the solution is expressed as a linear combination of some instances and, thus, their analysis offers some knowledge gain about the problem. Indeed, the final model is a good compromise between accuracy (almost 80% of predictions with errors under 20%) and simplicity (24% of samples become support vectors). Support vectors are mainly scattered around CyA blood levels of 320 ng/mL (± 100, $p > 0.05$) and in patients who weigh more than 45 Kg and who are over 50 years old.

We also compared distributions of the entire training set and the obtained support vectors using Principal Component Analysis (PCA). After their respective diagonalisation and standardising, we evaluated the scatter degree in every subset as a measure of the distance, $d(i,j)$, among the eigenvectors, $\mathbf{v}_i$ and $\mathbf{v}_j$, weighted by their eigenvalues, λ_i and λ_j:

$$\mathbf{DT} \equiv d(i,j) = \|\lambda_i \mathbf{v}_i - \lambda_j \mathbf{v}_j\|. \tag{15}$$

This distance was averaged over all possible pairs (i,j) of different eigenvectors. No geometrical differences were found between the original ($\bar{\mathbf{DT}} \pm \sigma_{\mathbf{DT}}$: 6.22 ± 3.14) and the SV data set (6.44 ± 3.64), which suggests that SVs scatter in a way similar to that for the whole data set and, consequently, reveals that a robust solution has been achieved.

V. DISCUSSION AND CONCLUSIONS

In this paper, we have presented time series prediction and classification approaches for a complex TDM problem.

We have compared state-of-the-art support vector machines and neural networks. Model comparison has been carried out in terms of accuracy and robustness. A novel kernel-based approach has been presented, which allows the incorporation of *a priori* knowledge and improves results in both approaches. Finally, we have analyzed model structure by performing sensitivity analyses on the best MLP model and inspecting the distribution of the support vectors on the best SVM model. These methods not only provide a ranking of relevant variables, but also constitute a methodology for model assessment.

The prediction of immunosuppressive blood concentrations is a challenging issue and leads to difficulties in selecting the optimum dose drug to avoid graft rejection and minimize adverse effects. Intensive drug monitoring is necessary in order to keep blood concentrations within the proper range. In this context, we have presented the formulation of state-of-the-art models that could help to individualize the CyA posology. Blood concentration models have been built to achieve accuracy and robustness. By predicting concentration instead of dosage we achieve objectivity and usefulness since the latter is based on a certain protocol of dosage administration that could disturb the final goal of TDM. In fact, this approach attempts to assist the health care team in dosage individualisation since physicians could take the blood concentration estimation as a helpful guide for dose administration. In [10], [12], we presented a scheme of two chained models where the CyA blood concentration predicted by the concentration prediction model constituted an input to the dosage prediction model. This system could serve as a dosage guide to the clinician, but it presented two basic problems. First, dosing follows a therapeutic guideline, which makes predicting dosage an indirect way of predicting doctors protocol. Second, from our particular point of view and experience, a direct translation of concentration to dosage could influence, or even replace, the doctor's own decision. A decision-support system can be defined as "a computer-based algorithm that assists a clinician with one or more component steps of the diagnostic process" [64], and thus, it should only aid doctors, rather than influence them or substitute them. In contrast, the scheme presented in this paper provides the clinician with different signals, rankings, and follow-up information.

Despite the fact that the results obtained for the blood concentration prediction are acceptable, they are inferior to those for classification purposes. Nevertheless, its joint consideration could be a valuable help in TDM. The best results for the CyA blood concentration prediction were obtained using the PD-SVR, where an ME of 0.36 ng/mL and a RMSE of 52.01 ng/mL were observed in the validation set. Our results clearly improve a previous work that followed the time series methodology [26], in which an MLP with lagged inputs was used to predict CyA levels in renal allograft recipients and the results were not optimal (bias: 25 ng/mL, precision: 74 ng/mL in the test set). From a statistical point of view, there are no significant differences between the neural and kernel models developed in our work. However, from our analysis of model robustness, we can conclude that although dynamic neural models give good results, PD-SVM yields a more robust solution, which is a direct consequence of using a regularized and tailored model.

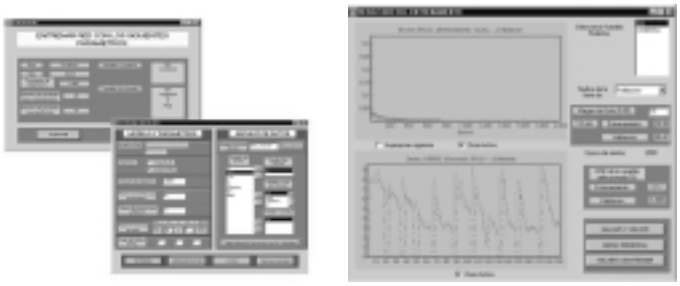

Fig. 8. Windows of the application for predicting the CyA blood concentration.

With regard to the classification problem, both one-against-all and one-against-one schemes have been attempted. Models could be useful for clinicians in designing dosage regimens to avoid toxic and subtherapeutic cyclosporine ranges and, in turn, to reduce costs to the Health Care System. The SVM method performs slightly better than the MLP in both schemes regarding sensitivity rates and AUC. Once again, designing specific penalisation profiles have produced better results.

Based on these outcomes, the application of SVM in the context of TDM can become a clinically useful tool. In this sense, we implemented the best model in an easy-to-use computer program in order to aid in the individualisation of dosage and pharmacotherapeutical attention (Fig. 8), which in turn brings state-of-the-art models closer to clinicians [65]. The main limitation encountered in this work is due to the group's location. Since patients were all from the same nephrology units, they had a series of characteristics in common. For example, the treatment guidelines and protocol administration for the patients were similar, which means that extrapolations to other centres should be treated with caution. Furthermore, a strict test should be performed before using the application in new situations. This, however, should not prevent the use of SVM methodology in other nephrology units, where they should be implemented taking into account the local population characteristics and dosing protocols.

Further studies are necessary in order to explore statistical differences between methods, the influence of clinical covariates, and the expansion of the predictive performance up to long-term follow-up. However, there is no doubt that the appearance of new protocols based on two-hour post-dosing monitoring (C_{2h}) constitutes the new cornerstone in CyA TDM. At present, there are only limited C_{2h} data in our hospital, but in a few years, we expect a substantial amount to be collected. The poor preliminary results obtained with neural networks and ARMA modeling [66] encourage the use of SVM in this new application.

ACKNOWLEDGMENTS

The authors want to express their gratitude to Prof. Antonio J. Serrano-López (Universitat de València, Spain) for his useful discussions and references on Decision-Support Systems, to Sergio Sánz (TISSAT S.A., Spain) for his valuable help in software development, to Prof. Ángel Navia-Vázquez

(Universidad Carlos III, Spain) for his useful comments on recurrent networks for uneven sampling problems, and to Dr. Begoña Porta-Oltra (Pharmacy Service of the Dr. Peset University Hospital of València, Spain) for the stimulating clinical discussions and careful data collection.

REFERENCES

[1] P. Belitsky, "Neoral used in the renal transplant recipient," *Transplantation Proceedings*, vol. 32, no. 3A Suppl. Review., pp. S10–S19, May 2000.

[2] L. A., "Factors influencing the pharmacokinetics of cyclosporine in man," *Therapeutic Drug Monitoring*, vol. 13, no. 6, pp. 465–477, Nov 1991.

[3] J. Parke and B. G. Charles, "NONMEM population pharmacokinetic modeling of orally administered cyclosporine from routine drug monitoring data after heart transplantation," *Therapeutic Drug Monitoring*, vol. 20, no. 3, pp. 284–293, Jun 1998.

[4] B. Charpiat, I. Falconi, V. Bréant, R. W. Jellife, J. M. Sab, C. Ducerf, N. Fourcade, A. Thomasson, and J. Baulieux, "A population pharmacokinetic model of cyclosporine in the early postoperative phase in patients with liver transplants, and its predictive performance with Bayesian fitting," *Therapeutic Drug Monitoring*, vol. 20, pp. 158–164, 1998.

[5] M. E. Brier, J. M. Zurada, and G. R. Aronoff, "Neural network predicted peak and trough gentamicin concentrations," *Pharmaceutical Research*, vol. 12, no. 3, pp. 406–412, 1995.

[6] A. S. Hussain, R. D. Johnson, N. N. Vachharajani, and R. W. A., "Feasibility of developing a neural network for prediction of human pharmacokinetic parameters from animal data," *Pharmaceutical Research*, vol. 10, no. 3, pp. 466–469, Mar 1993.

[7] P. Veng-Pedersen and N. Modi, "Application of neural networks to pharmacodynamics," *Journal of Pharmaceutical Sciences*, vol. 82, pp. 918–926, 1993.

[8] A. E. Gaweda, A. A. Jacobs, M. E. Brier, and J. M. Zurada, "Pharmacodynamic population analysis in chronic renal failure using artificial neural networks—a comparative study," *Neural Networks*, vol. 16, no. 5-6, pp. 841–845, 2003.

[9] S. Hirankarn, C. Downs, W. Street, and R. A. Herman, "Prediction of two ranges of cyclosporine level (subtherapeutic and toxic) using feature subset selection and artificial neural networks," in *AAPS Annual Meeting*, vol. 2 (2), Orlando, U.S.A., 2000, abstract 1274.

[10] G. Camps-Valls, E. Soria-Olivas, B. Porta-Oltra, J. J. Pérez-Ruizo, J. D. Martín-Guerrero, A. J. Serrano-López, and N. V. Jiménez-Torres, "A neural approach to cyclosporine dose prediction," *World Congress on Medical Physics and Biomedical Engineering*, July 2000.

[11] G. Camps-Valls, E. Soria-Olivas, J. D. Martín-Guerrero, J. J. Pérez-Ruixo, and N. V. Jiménez-Torres, "Neural networks ensemble for cyclosporine concentration monitoring," in *International Conference on Artificial Neural Networks*, vol. 2130. Vienna, Austria: Lecture Notes in Computer Science. Springer–Verlag., Aug 2001, pp. 706–711.

[12] G. Camps-Valls, B. Porta-Oltra, E. Soria-Olivas, J. D. Martín-Guerrero, A. J. Serrano-López, J. J. Pérez-Ruixo, and N. V. Jiménez-Torres, "Prediction of cyclosporine dosage in patients after kidney transplantation using neural networks," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 4, pp. 442–448, April 2003.

[13] G. Camps-Valls, E. Soria-Olivas, J. Pérez-Ruixo, A. Artés-Rodríguez, F. Pérez-Cruz, and A. Figueiras-Vidal, "Cyclosporine concentration prediction using clustering and Support Vector Regression methods," *IEE Electronics Letters*, vol. 38, no. 6, pp. 568–570, June 2002.

[14] M. Oellerich, V. W. Armstrong, B. Kahan, L. Shaw, D. W. Holt, R. Yatscoff, A. Lindholm, P. Halloran, K. Gallicano, and K. Wonigeit, "Lake Louise consensus conference on cyclosporin monitoring in organ transplantation: report of the consensus panel," *Therapeutic Drug Monitoring*, vol. 17, pp. 642–654, Dec 1995.

[15] T. A. S. Assays, "Manual analitique," Rundix Cedex, France: Laboratories ABBOTT, Division Diagnostic, XII-CYCLO-MONO-13.

[16] T. Kohonen, *Self-Organizing Maps*, 3rd ed. Springer Series in Information Sciences, Vol. 30, 2001.

[17] L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees*. Monterey, CA: Wadsworth and Brooks, 1984.

[18] A. Abraham and D. Steinberg, "Is neural network a reliable forecaster on earth? a MARS query!" in *Connectionist Models of Neurons, Learning Processes, and Artificial Intelligence*, J. M. . A. P. (Eds.), Ed. Lecture Notes on Computer Science. LNCS2084. Springer-Verlag, 2001.

[19] G. Camps-Valls, "Redes neuronales y máquinas de vectores soporte para la predicción y modelización de la concentración valle de ciclosporina A (CyA) en pacientes con trasplante renal," Ph.D. dissertation, Departament d'Enginyeria Electrònica. Universitat de València, July 2002, available from http://www.uv.es/∼gcamps.

[20] B. Porta, J. J. Pérez-Ruixo, N. V. Jiménez, A. Sancho, and L. M. Pallardó, "Individualización posológica de ciclosporina en pacientes con trasplante renal: propuesta de un modelo farmacocinético en predicción." *Farmacia Hospitalaria*, vol. 22, no. 4, pp. 181–187, 1998.

[21] B. Porta, "Modelado farmacocinético de ciclosporina en pacientes con trasplante renal," Ph.D. dissertation, Departament de Farmàcia Hospitalaria i Galènica. Universitat de València., 2002.

[22] J. F. Hair, R. E. Anderson, R. L. Tatham, and W. C. Black, *Multivariate Data Analysis*, 5th ed. New Jersey, U.S.A.: Prentice-Hall International, Inc., 1984.

[23] B. D. Kahan, W. G. Kramer, C. A. Wideman, S. M. Flechner, M. Lorber, and C. T. van Buren, "Demographics factors affecting the pharmacokinetics of cyclosporine estimated by radioinmunoassay," *Transplantation*, vol. 41, pp. 459–464, 1986.

[24] A. S. Weigend and N. A. Gershenfeld, *Time Series Prediction. Forecasting the Future and Understanding the Past. Proceedings of the NATO Advanced Research Workshop on Comparative Time Series Analysis held in Santa Fe, New Mexico, May 14–17, 1992. Proceedings Volume XV*. Addison–Wesley, 1994, vol. XV.

[25] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 3rd ed. New Jersey, U.S.A.: Prentice Hall, 1999.

[26] M. E. Brier, "Empirical pharmacokinetic predictions for cyclosporine using a time series neural network," *Pharmaceutical Research*, vol. 12, no. Suppl. S363, 1995.

[27] M. Nørgaard, O. Ravn, and N. Poulsen, "NNSYSID & NNCTRL – tools for system identification and control with neural networks," *IEE Computing & Control Engineering Journal*, vol. 12, no. 1, pp. 29–36, Feb. 2001.

[28] L. Ljung, *System Identification. Theory for the user*, 2nd ed. New Jersey, U.S.A.: Prentice-Hall International, Inc., 1999.

[29] E. A. Wan, "Finite Impulse Response neural networks with applications in time series prediction," Ph.D. dissertation, Department of Electrical Engineering. Stanford University, November 1993, available at http://www.ece.ogi.edu/∼ericwan/.

[30] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, and K. Lang, "Phoneme recognition using time-delay neural networks," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, pp. 328–339, 1989.

[31] E. A. Wan, "Modeling nonlinear dynamics with neural networks: Examples in time series prediction," in *Proceedings of the Fifth Workshop on Neural Networks: Academic/Industrial/NASA/Defense, WNN93/FNN93.*, San Francisco, U.S.A., November 1993, pp. 327–332, available at http://www.ece.ogi.edu/∼ericwan/.

[32] B. de Vries and J. C. Principe, "The Gamma model – a new neural model for temporal processing," *Neural Networks*, vol. 5, no. 4, pp. 565–576, 1992.

[33] J. C. Principe, B. deVries, and P. G. deOliveira, "The gamma filter – A new class of adaptive IIR filters with restricted feedback," *IEEE Transactions on Signal Processing*, vol. 41, no. 2, pp. 649–656, Feb 1993.

[34] J. C. Principe, B. de Vries, J. Kuo, and P. Guedes-de Olivera, "Modeling applications with the focused Gamma net," in *Neural Information Processing Systems, NIPS*, 1991, pp. 143–150, available from http://books.nips.cc/papers/files/nips04/.

[35] J. L. Elman, "Finding structure in time," *Cognitive Science*, vol. 14, pp. 179–211, 1988.

[36] V. N. Vapnik, *Statistical Learning Theory*. Wiley, 1998.

[37] B. Schölkopf and A. Smola, *Learning with kernels*. MIT Press, 2002.

[38] B. Boser, I. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proc. 5th Ann. Workshop on Computational Learning Theory*, D. Haussler, Ed. ACM Press, 1992, pp. 144–152.

[39] V. N. Vapnik, S. Golowich, and A. Smola, "Support vector method for function approximation, regression estimation, and signal processing," in *Neural Information Processing Systems*, M. Mozer, M. Jordan, and T. Petsche, Eds. Cambridge, MA: M.I.T. Press, 1997, pp. 169–184.

[40] B. Schölkopf and A. Smola, *Learning with Kernels – Support Vector Machines, Regularization, Optimization and Beyond*. MIT Press Series, 2001.

[41] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Transaction on Neural Networks*, vol. 13, no. 2, 3 2002.

[42] B. Schölkopf, K.-K. Sung, C. J. Burges, F. Girosi, P. Niyogi, T. Poggio, and V. N. Vapnik, "Comparing support vector machines with Gaussian kernels to radial basis function classifiers," *IEEE Transaction on Signal Processing*, vol. 45, no. 11, pp. 2758–2765, Nov. 1997.
[43] J. Weston and C. Watkins, "Multi-class support vector machines," in *ESANN*, 1999.
[44] U. H. G. KreBel, *Pairwise classification and support vector machines, In: Advances in Kernel Methods: Support Vector Learning*. Cambridge, MA, U.S.A.: The MIT Press, Cambridge, MA, 1999.
[45] V. N. Vapnik, *Statistical Learning Theory*. New York: John Wiley & Sons, 1998.
[46] B. Schölkopf, P. L. Bartlett, A. Smola, and R. Williamson, "Shrinking the tube: a new support vector regression algorithm," in *Advances in Neural Information Processing Systems 11*, M. S. Kearns, S. A. Solla, and D. A. Cohn, Eds. Cambridge, MA: MIT Press, 1999, pp. 330 – 336.
[47] Y. Lin, Y. Lee, and G. Wahba, "Support Vector Machines for classification in nonstandard situations," University of Wisconsin-Madison, Department of Statistics TR 1016, 2000.
[48] C.-F. Lin and S.-D. Wang, "Fuzzy support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 464–471, 2002.
[49] Q. Tao, G.-W. Wu, F.-Y. Wang, and J. Wang, "Posterior probability support vector machines for unbalanced data," *IEEE Transactions on Neural Networks*, vol. 16, no. 6, pp. 1561–1573, 2005.
[50] F. Pérez-Cruz, "Máquina de vectores soporte adaptativa y compacta," Ph.D. dissertation, Dpto. Teoría de la Señal y Comunicaciones. Universidad Carlos III de Madrid., Dec. 2000.
[51] G. Camps-Valls, E. Soria-Olivas, J. Pérez-Ruixo, A. Artés-Rodríguez, F. Pérez-Cruz, and A. Figueiras-Vidal, "A profile-dependent kernel-based regression for cyclosporine concentration prediction," in *Neural Information Processing Systems, NIPS 2001. Workshop on New Directions in Kernel-Based Learning Methods*, December 2001.
[52] J. D. Martín-Guerrero, G. Camps-Valls, E. Soria-Olivas, A. J. Serrano-López, J. J. Pérez-Ruixo, and N. V. Jiménez-Torres, "Dosage individualization of erythropoietin using a profile-dependent support vector regression," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 10, pp. 1136–1142, June 2003.
[53] A. N. Refenes, Y. Bentz, D. W. Bunn, A. N. Burgess, and A. D. Zapranis, "Financial time series modeling with discounted least squares backpropagation," *Neurocomputing*, vol. 14, p. 123 –138, 1997.
[54] F. E. H. Tay and L. J. Cao, "Modified support vector machines in financial time series forecasting," *Neurocomputing*, vol. 48, pp. 847–861, 2002.
[55] G. Camps-Valls, A. Chalk, A. Serrano-Lopez, J. D. Martin-Guerrero, and E. Sonnhammer, "Profiled support vector machines for antisense oligonucleotide efficacy prediction," *BMC Bioinformatics*, no. 5, p. 135, available in OpenAccess: http://www.biomedcentral.com/1471-2105/5/135.
[56] G. Gómez-Pérez, G. Camps-Valls, J. Gutiérrez, and J. Malo, "Perceptual adaptive insensitivity for support vector machine image coding," *IEEE Transactions on Neural Networks*, vol. 16, no. 6, pp. 1574–1581, 2005.
[57] S. S. Keerthi and C.-J. Lin, "Asymptotic behaviors of support vector machines with gaussian kernel," *Neural Computation*, vol. 15, no. 7, pp. 1667–1689, 2003.
[58] H.-T. Lin and C.-J. Lin, "A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods," National Taiwan University, Department of Computer Science and Information Engineering, Tech. Rep., 2003, available at http://www.csie.ntu.edu.tw/∼cjlin/papers/tanh.pdf.
[59] C. Cortes and V. Vapnik, "Support vector networks," *Machine Learning*, vol. 20, pp. 273 – 297, 1995.
[60] R. Congalton and K. Green, *Assessing the Accuracy of Remotely Sensed Data. Principles and Practices*, 1st ed. U.S.A.: CRC Press, 1999.
[61] R. Rifkin and A. Klautau, "In defense of one-vs-all classification," *Journal of Machine Learning Research*, vol. 5, no. 1, pp. 101–141, 2004.
[62] P. S. Bradley, U. M. Fayyad, and O. L. Mangasarian, "Mathematical programming for data mining: formulations and challenges," Mathematical Programming Technical Report 98-01, Computer Sciences Department, University of Winsconsin, WI, Tech. Rep. MSR-TR-98-04, Jan 1998.
[63] G. B. Orr and K.-R. Müller, *Neural Networks: Tricks of the Trade*. Springer-Verlag, Berlin, Heidenberg, 1998.
[64] E. S. Berner, *Clinical Decision Support Systems. Theory and Practice*, 1st ed. New-York: Springer–Verlag, 1999.
[65] S. Sáez, E. Soria-Olivas, G. Camps-Valls, J. D. Martín-Guerrero, A. J. Serrano-López, and N. V. Jiménez-Torres, "Aplicación informática basada en redes neuronales temporales para problemas de farmacocinética clínica." in *XVIII Congreso Anual de la Sociedad Española de Ingeniería Biomédica, CASEIB2000*, Cartagena, Spain, Sep 2000, pp. 235–238.
[66] G. Camps-Valls, A. J. Serrano-López, B. Porta-Oltra, J. D. Martín-Guerrero, E. Soria-Olivas, and N. V. Jiménez-Torres, "Neural networks for C_{2h} cyclosporine concentration modelling," in *32nd European Symposium on Clinical Pharmacy, ESCP 2003.*, València, Spain, Sep 2003.