# Cyclosporine Concentration Prediction using Clustering and Support Vector Regression Methods

G. Camps-Valls, E. Soria-Olivas, J. J. Pérez-Ruixo,
F. Pérez-Cruz, A. R. Figueiras-Vidal and A. Artés-Rodríguez

*Abstract*—**This paper proposes a combined strategy of clustering and Support Vector Regression (SVR) methods to predict Cyclosporine A (CyA) concentration in renal transplant recipients. Clustering combats the high variability and non-stationarity of the time series and reports knowledge gain in the problem. The SVR outperforms other classical neural networks.**

*Introduction* Despite progress with newer agents, Cyclosporine A (CyA) is still the cornerstone of immunosuppression in patients who have undergone kidney transplantation. However, 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 behavior of CyA presents a substantial inter- and intra-individual variability which appears to be particularly evident in the earlier post-transplantation period (<3 months), when the risk and clinical consequences of acute rejection are higher than in stable renal patients (>6 months) [1]. Several factors such as clinical drug interactions and patient compliance can also significantly alter blood CyA concentrations and thus intensive therapeutic drug monitoring of CyA becomes necessary but it influences the patient's quality of life and the cost of the care.

Models capable of predicting the future concentration and determining the optimal dosage of CyA usually aid to individualize therapy. Few studies have been done and none, to our knowledge, using machine learning or neural networks. We propose the use of Support Vector Machines (SVM) for solving this task since they do not rely on any *a priori* assumption about the problem and have proven to be effective techniques in a wide range of applications [2]. To deal with the non-uniform sampling (NUS), the presence of non-stationary processes, and the high variability in the time series, we have previously clustered the data.

*Support Vector Regressor* SVMs are state-of-the-art tools for non-linear input-output knowledge discovery [2]. The Support Vector Regressor (SVR) is for regression and function approximation. Given a labelled training data set ($\{(\mathbf{x_i}, \mathbf{y_i}), \mathbf{i} = \mathbf{1}, \ldots, \mathbf{n}\}, \mathbf{x_i} \in \mathbb{R}$ and $y_i \in \mathbb{R}$) and a nonlinear mapping to a higher dimensional space $\phi(\cdot)$ ($\mathbf{x} \in \mathbb{R} \longrightarrow \phi(\mathbf{x}) \in \mathbb{R}^{\mathbb{H}}$, $d \leq H$), the SVR solves:

$$\min_{\mathbf{w},\mathbf{b},\xi,\xi_\mathbf{i}^*,\varepsilon} \frac{1}{2}\|\mathbf{w}\|^\mathbf{2} + \mathbf{C}\sum_{\mathbf{i=1}}^{\mathbf{n}}(\xi_\mathbf{i} + \xi_\mathbf{i}^*) + \mathbf{Cn}\nu\varepsilon \tag{1}$$

subject to the following constraints:

$$y_i - \boldsymbol{\phi}^T(\mathbf{x_i})\mathbf{w} - \mathbf{b} \leq \varepsilon + \xi_i \qquad \forall i = 1, \ldots, n \tag{2}$$

$$\boldsymbol{\phi}^T(\mathbf{x_i})\mathbf{w} + \mathbf{b} - \mathbf{y_i} \leq \varepsilon + \xi_i^* \qquad \forall i = 1, \ldots, n \tag{3}$$

$$\xi_i, \xi_i^* \geq 0 \qquad \forall i = 1, \ldots, n \tag{4}$$

where $\mathbf{w}$ and $b$ define a linear regressor in the feature space, nonlinear in the input space unless $\phi(\mathbf{x_i}) = \mathbf{x_i}$. In addition, $\xi_i$, $\xi_i^*$ and $C$ are, respectively, positive slack variables to deal with training samples with a prediction error larger than $\varepsilon$ and the penalization applied to these ones. The tube size $\varepsilon$ is traded off against model complexity and slack variables via a constant $\nu \in [0, 1)$ which can be regarded as an upper bound on the fraction of errors and a lower bound on the fraction of Support Vectors (SV). This formulation is known as the $\nu$-SVR [3]. The usual procedure for solving SVRs introduces the linear restrictions using Lagrange multipliers into Eq. (1), computes the Karush-Kuhn-Tucker conditions and solves the Wolfe's dual problem using quadratic programing procedures [2], [3]. We will instead use an alternative procedure that consists in solving iteratively a series of weighted least square problems [4], known as Iterative Re-Weighted Least Square (IRWLS) procedure, that is summarised in the next steps:

1. Solve the linear system:

$$\begin{bmatrix} \mathbf{H} + \mathbf{D}^{-\mathbf{1}}_{(\mathbf{a}+\mathbf{a}^*)} & \mathbf{1} & \mathbf{E} \\ \mathbf{1^T} & 0 & 0 \\ \mathbf{E^T} & 0 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ b \\ \varepsilon \end{bmatrix} = \begin{bmatrix} \mathbf{y} \\ 0 \\ C\mathbf{n}\nu \end{bmatrix} \tag{5}$$

where we have defined:

$$(\mathbf{H})_{\mathbf{ij}} = \boldsymbol{\phi}^T(\mathbf{x_i})\boldsymbol{\phi}(\mathbf{x_j}) = \kappa(\mathbf{x_i}, \mathbf{x_j}) \tag{6}$$

$$(\mathbf{D}^{-\mathbf{1}}_{\mathbf{a}+\mathbf{a}^*})_{\mathbf{ij}} = \frac{\delta(i-j)}{a_i + a_i^*} \tag{7}$$

$$\mathbf{E} = \left[\frac{a_1 - a_1^*}{a_1 + a_1^*}, \ldots, \frac{a_n - a_n^*}{a_n + a_n^*}\right]^T \tag{8}$$

and, in order to work with reproducing kernels in Hilbert Space, we require $\mathbf{w}$ to be a linear a combination of subset of training samples $\mathbf{w} = \sum_{\mathbf{i=1}}^{\mathbf{n}} \beta_\mathbf{i}\boldsymbol{\phi}(\mathbf{x_i})$.

2. Recompute

$$a_i^{(*)} = \begin{cases} 0, & e_i^{(*)} < 0 \\ \dfrac{C}{e_i^{(*)}}, & e_i^{(*)} \geq 0 \end{cases} \tag{9}$$

where $e_i = y_i - \boldsymbol{\phi}^T(\mathbf{x_i})\mathbf{w} - \mathbf{b} - \varepsilon$ and $e_i^* = \boldsymbol{\phi}^T(\mathbf{x_i})\mathbf{w} + \mathbf{b} - \mathbf{y_i} - \varepsilon$

3. Repeat until convergence.

The column vectors $\mathbf{y}$, $\mathbf{a}$, $\mathbf{a}^*$, $\boldsymbol{\beta}$ and $\mathbf{1}$, present the obvious expressions and $\mathbf{H}$ is known as the *kernel* matrix, since it is only formed by inner products of the training samples in the feature space. Consequently, neither the minimizing procedure nor the use of the regressor needs to know explicitly the form of the nonlinear mapping, $\boldsymbol{\phi}(\cdot)$, but only its kernel representation $\kappa(\cdot, \cdot)$. The needed transformations to obtain the IRWLS procedure from the minimization of Eq. (1) are detailed in [4].

*Data collection* Fifty-seven renal allograft recipients treated in the Nephrology Service of the Hospital Universitari Dr. Peset in the city of València (Spain) were included in this study. Patients received a standard immunosuppressive regimen of CyA (Sandimmun Neoral®). Steady state blood samples were withdrawn 12-14 hours after dose administration and measured by a specific monoclonal fluorescence polarization immunoassay. We collected 11 patient factors such as age, gender, creatinine plasma levels, creatinine clearance, alkaline phosphatase, hematocrit, urea and bilirubin, along with dosage, CyA blood concentration and post-transplantation days to build the models. Each *pattern* was formed by the present and past values of the these variables in order to perform one-step-ahead prediction. We split the data into two groups: two-thirds of the patients were used to train the models and the rest for their validation using the cross-validation method.

*Results* The high intersubjects variability (coefficient of variation, CV = 31%) led us to set up clusters and then building individual predictive models for each one of them. We have used the well-known $K$-means clustering algorithm and selected the optimal partition by evaluating the root-mean-square error (RMSE) of dedicated models through 3-fold cross-validation experiments. Four clusters were identified with this methodology. Since the second cluster was the largest (42% of the patterns), models yielded poor results in it (RMSE>60 ng/mL) and its variability still held high values (CV=27%), we decided to perform re-clustering on it. Once again a four-clusters partition was employed for the posterior prediction. The clustering reduced the RMSE and revealed postoperatory days, creatinine clearance, CyA blood concentration and serum alkaline phosphatase as decisive factors.

In Table I results are benchmarked with a multilayer perceptron (MLP) trained with the familiar back-propagation algorithm and the Elman recurrent neural network [5] both with and without a previous clustering.

Elman network fails with a clustering approach since NUS becomes more evident since samples from the same patient can be in different clusters and, thus, contextual neurons do not deal efficiently with past time samples. The $\nu$-SVR model outperforms the MLP since the CV[%] is subsequently reduced in each partition. Blood levels accurately predicted (%BLAP) if an error margin of 20% is fixed have reached a value of 70%, which is an excellent result considering the time series characteristics of our population. Fig. IV shows predictions in two validation patients.

There is, nevertheless, 14% of patients with poor predictions which can be due to errors in drug dosage administration, in recording blood sampling times or abrupt changes in each patient's clinical condition. An additional hypothesis for this could be some kind of liver dysfunction since alkaline phosphatase has resulted in a critical clustering factor. If we discard these patients, %BLAP increases to a

**TABLE I.** Mean error (ME [ng/mL]) and root-mean-square error (RMSE [ng/mL]) of models both for training and validation.

| **Model** | **ME(T) (± CI95%)** | **ME(V) (± CI95%)** | **RMSE(T) (± CI95%)** | **RMSE(V) (± CI95%)** |
|---|---|---|---|---|
| **MLP** | 0.13 | 3.17 | 52.71 | 52.64 |
| | (-3.88,4.14) | (-1.81,8.16) | (48.27,56.82) | (45.27,59.11) |
| **MLP with** | 0.18 | 3.23 | 51.79 | 51.60 |
| **clustering** | (-3.82,4.18) | (-1.37,7.83) | (47.98,55.60) | (46.90,56.30) |
| **Elman** | -5.85 | 0.30 | 53.69 | 52.72 |
| **21×2×1** | (-9.91,-1.79) | (-4.71,5.31) | (49.32,57.73) | (46.03,58.66) |
| **Elman with** | -6.12 | 0.40 | 56.31 | 55.21 |
| **clustering** | (-10.21,-1.99) | (-4.75,5.55) | (52.01,60.61) | (48.41,62.01) |
| **$\nu$-SVR** | -11.49 | 0.36 | 50.08 | 51.39 |
| **RBF kernels** | (-15.18,-7.19) | (-4.68,5.44) | (45.83,54.30) | (45.60,58.81) |
| **$\nu$-SVR with** | -9.91 | 0.32 | 49.25 | 48.25 |
| **clustering** | (-12.92,-6.90) | (-3.69,4.33) | (45.95,52.55) | (44.10,52.40) |

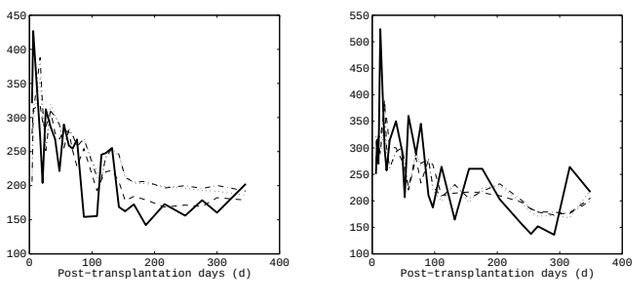


**Fig. 1.** CyA trough concentration predictions within two validation patients. Actual, − − −− $\nu$-SVR, · · · · · · · MLP, · — · — · ELMAN

88% with the $\nu$-SVR and to a 75% with the MLP. Support Vectors are mainly placed in the early post-operatory period (68% of them in the early three months) and change of cluster according to high levels in CV[%].

*Conclusions* In this paper we have proposed the combination of clustering and a state-of-the-art technique for knowledge gain and accuracy improvement in a complex pharmacokinetic prediction problem. By means of clustering we can identify patients' state and their future evolution, specify *confidence* intervals for each cluster prediction and discover important and meaningless patient' factors all the time. The power and versatility of the SVR machines allowed fast and reliable prediction schemes.

Further work is tided up to benchmark Incremental Learning with SVR approaches and Non-Linear Mixed-Effects Modelling (NONMEM) which is a commonly used method in population pharmacokinetics.

*References*

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

[2] Vladimir N. Vapnik, *Statistical Learning Theory*, John Wiley & Sons, New York, 1998.

[3] 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, 1999, pp. 330 – 336, MIT Press.

[4] F. Pérez-Cruz and A. Artés-Rodríguez, "An IRWLS procedure for $\nu$-SVR," in *Proceedings of the ICASSP'01*, Salt Lake City, Utah. U.S.A., May 2001.

[5] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice Hall, 1999.

**Authors' affiliations:**

G. Camps-Valls and E. Soria-Olivas. Dept. Enginyeria Electrònica. Universitat de València. C/ Dr. Moliner, 50, 46100 - Burjassot (València). Spain. *gcamps@uv.es*.

J. J. Pérez-Ruixo. Global Clinical Pharmacokinetics and Clinical Pharmacodynamics Department. Janssen Research Foundation. Turnhoutsweg, 30, B-2340 Beerse. Belgium.

F. Pérez-Cruz, A. R. Figueiras-Vidal and A. Artés-Rodríguez. Dpto. Teoría de la Señal y Comunicaciones. Universidad Carlos III de Madrid. C/ Butarque 15, 28911 - Leganés (Madrid). Spain.