

# Example-based Support Vector Machine for Drug Concentration Analysis

Wenqi You  
Integrated Systems Laboratory  
EPFL, Switzerland 1015

Nicolas Widmer  
University Hospital Center and  
University of Lausanne, Switzerland 1011

Giovanni De Micheli  
Integrated Systems Laboratory  
EPFL, Switzerland 1015

**Abstract**—Machine learning has been largely applied to analyze data in various domains, but it is still new to personalized medicine, especially dose individualization. In this paper, we focus on the prediction of drug concentrations using Support Vector Machines (SVM) and the analysis of the influence of each feature to the prediction results. Our study shows that SVM-based approaches achieve similar prediction results compared with pharmacokinetic model. The two proposed example-based SVM methods demonstrate that the individual features help to increase the accuracy in the predictions of drug concentration with a reduced library of training data.

## I. INTRODUCTION

Nowadays, in clinical pharmacology, the dosage of medicines are chosen on the basis of clinical monitoring of patients. However, this experience-based method is not suitable to all kinds of drugs. There is a small group of medicines whose effective therapeutic concentration range is quite narrow and therefore there is a very high risk to under- or over-dose a patient. Under-dosing will lead to an ineffective treatment, while over-dosing will expose the patient to a risk of toxicity. Thus knowing the therapeutic range of a drug is essential to properly carry out the clinical monitoring; in other words, it is necessary to know how the human body affects the disposition of the drug. This area of research is known as *population pharmacokinetics* (PK).

There exist two main methodologies applied to data analysis in the field of pharmacokinetics[1]. One of them uses mathematically simple techniques to obtain an overview of a drug disposition, usually by computing a simple AUC (Area Under the plasma concentration time Curve). This method is mainly used when there are few patients available but a fair number of samples per patient. However, in reality, such AUC could not be accurate if the data cannot support a modeling approach. The other method uses highly-sophisticated techniques to extract pharmacokinetic information from a sample of the populations. In this process some model structures may be ignored due to insufficient data to characterize the process, e.g. the absorption process after an oral dosage.

Many well-designed studies have been analyzed based on a structural modeling approach[2][3][4]. Recently, new statistical approaches have been introduced to pharmacokinetic analysis, such as Bayesian approaches[5][6]. However, most existing methods process real values with a pharmacokinetic model as shown in Figure 1, where  $K_a$  and  $K_{el}$  denote

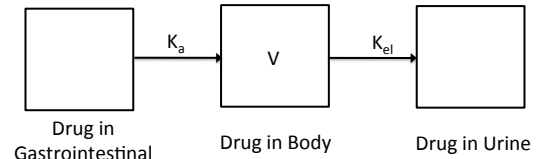


Fig. 1. Example of a One-Compartment PK Model

respectively the coefficients of a drug absorption and elimination in the body,  $V$  stands for volume of distribution. Boolean variables, such as gender, whether taking multiple drugs, etc., create strong discontinuities in the models and are in general not supported. Furthermore, due to the explicit analytical model, adding or removing a parameter is complicated. Hence, in this paper, we introduce a machine learning technology to the domain of personalized medicine, especially dose individualization.

We give a brief review of four representative machine learning algorithms in Section 2. Section 3 discusses the principles of methodologies, including the introduction of general PK model, applying LS-SVM to predict drug concentration and the proposed two example-based SVM strategies. We demonstrate the quality of our algorithm experimentally in Section 4 while Section 5 concludes this study.

## II. RELATED WORK

Machine learning has been applied with some success to solve classification problems in computer vision and pattern recognition in the past few decades[7]. However, there is very little literature on the utilization of machine learning to achieve the dose individualization of personalized medicine.

Four of the most representative machine learning techniques are *Decision Trees* (DT), *Neural Networks* (NN), *Support Vector Machine* (SVM), and *Boosting* in which *Adaptive Boosting* or *AdaBoost* (AB) is the standard one. With the extension to solve regression problems, these techniques became popular in various other domains such as image superresolution[8], object tracking[11], etc.. Among the four, DT is the simplest and thus the fastest approach, but it is not as precise as the other three, especially for regression to give a prediction on continuous numbers[12]. NN is the oldest technique of the four inspired from neurobiological knowledge, but it is often regarded as a black box due to the high complexity of the model it builds[13]. SVM employs a hyperplane in a high- or infinite- dimensional space to do

classification or regression tasks[16]. It is convenient due to both its clear mathematic understanding and its control of the overfitting problem. AB is a meta-algorithm used in conjunction with other weak classifiers iteratively, but it is sensitive to outliers or noisy data[17].

Here we have chosen the SVM technique for our modeling system because of its appropriate complexity, efficiency and strength in data regularization[9]. In the paper, we propose two example-based learning approach based on least square SVM (LS-SVM), which learn from selected smaller training sample subsets. Instead of solving a convex *quadratic programming problem* (QP), the LS-SVM classifier manages to give a solution simply by solving a set of linear equations[15]. The experimental results show that drug concentration predicted by SVM-based approaches are similar or closer to the measured values than those predicted by the PK model, depending on the input features that the algorithm used. Further study also analyzes the importance of each biological or demographical characteristic of a patient, such as the gender, age, etc.

### III. METHODOLOGY

#### A. General Pharmacokinetic Model

In the simplest situation, we consider the drug concentration after a single *intravenous* (IV) bolus dose, which has a quick and concentrated drug effect, to a one-compartment pharmacokinetic model[1]. In the one-compartment model, the human body is considered as one unique chemical and biological system and the drug's concentration is computed by a first-order linear differential equation as shown in (1) and (2).

$$\frac{dC}{dt} = -k_{el} \cdot C \quad C_{init} = \frac{\text{dose}}{V} \quad (1)$$

$$C = \frac{\text{dose}}{V} \cdot e^{-k_{el} \cdot t} \quad (2)$$

where  $V$  denotes the volume of distribution and  $k_{el}$  stands for the elimination rate of the drug inside a body, which generally depends of the liver metabolic capacity or of the renal function (expressed by creatinine clearance), depending of the drug.

Furthermore, if we assume that the drug is taken orally, we need to consider one more component, i.e. the mechanism of absorption from the *gastrointestinal* (GI) tract to the arteriovenous system as illustrated in Figure 1. This way after a single dose, the concentration of the drug is calculated using (3):

$$C = \frac{F \cdot \text{dose} \cdot k_a}{V \cdot (k_a - k_{el})} \cdot \{e^{-k_{el} \cdot t} - e^{-k_a \cdot t}\} \quad (3)$$

where  $k_a$  is the absorption rate and  $F$  is an extent factor called bioavailability.

Practically, in real clinical scenarios, multiple dose regimens are more interesting to clinicians. After  $m$  dosages, the

drug concentration could be estimated as:

$$C = \frac{F \cdot \text{dose} \cdot k_a}{V \cdot (k_a - k_{el})} \cdot \left\{ \left[ \frac{1 - e^{-m \cdot k_{el} \cdot \tau}}{1 - e^{-k_{el} \cdot \tau}} \right] \cdot e^{-k_{el} \cdot t} - \left[ \frac{1 - e^{-m \cdot k_a \cdot \tau}}{1 - e^{-k_a \cdot \tau}} \right] \cdot e^{-k_a \cdot t} \right\} \quad (4)$$

where  $\tau$  stands for the dosage interval and  $t$  for the time when the drug concentration is measured.

In clinical setting, it is often not possible to collect many data points for each patient. Thus, even though a pharmacokinetic model may require two or more parameters, only one or two data points may be available. In such cases, nonlinear regression trough the whole samples of the patient population available is performed, using approach such as NONMEM[10]. This approach could somehow be compared to a Bayesian analysis performed to obtain estimations of the patient pharmacokinetic parameter values by minimizing:

$$\sum_{i=1}^{N_1} \frac{(C_{obs_i} - C_{calc_i})^2}{\text{variance}_i} + \sum_{j=1}^{N_2} \frac{(P_{pop_j} - P_{calc_j})^2}{\text{variance}_j} \quad (5)$$

where  $C$  stands for drug concentration,  $P$  for some other parameters,  $N_1$  and  $N_2$  are the number of data points and parameters respectively[19]. In Section 3, we compare our proposed method with the analytical model mainly using equation (4) and (5).

#### B. Least Square Support Vector Machine (LS-SVM)

Let's assume that there are  $N$  patients, each of whom has been examined  $n_i$  times. Each sample takes into account  $d$  features. Hence the training data are given as  $\{(x_1^1, y_1^1), (x_2^1, y_2^1), \dots, (x_{n_1}^1, y_{n_1}^1), \dots, (x_{n_N}^N, y_{n_N}^N)\}$ , where  $y_{n_i}^j$  denotes the output drug concentration and  $x_{n_i}^j$  is the space of input patterns  $\mathbb{R}^d$ , e.g. the age, gender, body weight, dosage of the drug, interval between dosing and measuring, etc.. Our goal is to find a linear function

$$f(x) = w \cdot \phi(x) + b \quad (6)$$

which approximates the relationship between the data set points and can be used to estimate the output  $y$  with respect to a new input patient data. In (6),  $\phi(x)$  maps the input samples to a higher-dimensional feature space by finding a non-linear function in the original space, and  $w$  stands for the the weights of the feature space and  $b \in \mathbb{R}$  an offset constant.

Any regression problem uses a loss function  $\mathcal{L}(y, f(x))$  to describe how the predicted values deviate from the measured ones. We can find many forms of loss functions in the literature. For our purposes, we apply the quadratic loss function which is defined as:

$$\mathcal{L}(y, f(x)) = (y - f(x))^2 \quad (7)$$

In order to find the optimal solution, function (7) has to be minimized. Besides the task of minimizing the loss function, we also need to make sure that the function  $f$  is as *flat* as possible in order to prevent overfitting the problem. The

flatness of the function  $f$  can be ensured by keeping the norm  $\|w\|^2$  small. Thus the regression problem can be rewritten as optimizing the following objective function:

$$\min_{w,b} \frac{1}{2} \|w\|^2 + C_0 \sum_{j=1}^N \sum_{i=1}^{n_j} [y_i^j - w \cdot \phi(x_i^j) - b]^2 \quad (8)$$

where the constant  $C_0$  determines the tradeoff between the flatness of  $f$  and the amount up to which deviations between the predicted and measured values are tolerated. According to [15], the optimal  $w$  can always be expressed by  $w = \sum_{j=1}^N \sum_{i=1}^{n_j} \alpha_i^j \phi(x_i^j)$ , where  $\alpha$  and  $b$  are found by solving the linear system:

$$\underbrace{\begin{bmatrix} \mathbf{K} + \frac{1}{C_0} \mathbf{I} & \mathbf{1} \\ \mathbf{1}^T & 0 \end{bmatrix}}_H \begin{bmatrix} \alpha \\ b \end{bmatrix} = \begin{bmatrix} y \\ 0 \end{bmatrix} \quad (9)$$

where  $\mathbf{K}$  is the kernel matrix and is defined as  $K_{ab} = \phi(x_{i_a}^{j_a})^T \phi(x_{i_b}^{j_b})$ . The use of the kernel matrix greatly helps reducing the computational complexity without explicitly computing  $\phi(x)$ , making use of the fact that the SVM algorithm depends only on dot products between sample patterns [18]. Hence, after defining the kernel function, the least-square optimization problem could be solved simply by inverting the first term  $H$  in the left-hand side of (9).

Once the  $\alpha$  and  $b$  have been obtained, the output could be estimated via the following prediction function:

$$f(x) = \sum_{j=1}^N \sum_{i=1}^{n_j} \alpha_i^j \mathbf{K}(x_i^j, x) + b \quad (10)$$

Till now, LS-SVM trains the inputs as a whole and applies the model to new patient indifferently. Experiments show that this improves the prediction results but degrades greatly with the decrease of input data samples.

### C. Optimization Using Example-based SVM (E-SVM)

Similarly to the scenario described above, we introduce the idea of example-based SVM. Although LS-SVM approach could estimate the drug concentration with a large number of input features and also manage to simplify the addition or deletion of a new feature, it still treats the input training samples without selection for each of the new testing sample before running the algorithm. However, when comparing to other fields of using SVM, such as computer vision or pattern recognition, clinical samples are more often much fewer than the data in these domains. Therefore, *bad* samples are going to greatly affect the model built by SVM, which could cause an inaccuracy in the predictions.

Hence we propose our Example-based SVM approach to optimize the prediction, which carefully selects useful training patients in the library to be our examples and builds the personalized model specifically for a new patient. As illustrated in Figure 2, our system first takes all the features  $f_d$  from a new patient and extracts a subset of them  $f_s$  to be the selection criteria for training samples. The weights for

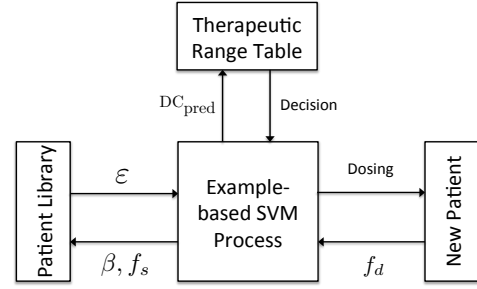


Fig. 2. Flow of Example-based SVM Approach

each feature in  $f_s$  are denoted by  $\beta$ . For each training patient in the library, we assign a coefficient  $\varepsilon$  according to:

$$\varepsilon_i^j = \begin{cases} 0 & \|\beta \cdot (f_s - f_j)\|_a > t_s \\ 1 & \|\beta \cdot (f_s - f_j)\|_a \leq t_s \end{cases} \quad (11)$$

where  $i, j$  stands for the  $i^{\text{th}}$  sample of the  $j^{\text{th}}$  patient in the library,  $t_s$  is the threshold, and  $f_j$  is the corresponding criteria features. Both  $f_j$  and  $f_s$  are normalized in advance by subtracting the mean and dividing the standard deviation of each feature. We indicate  $a$  as an  $l_a$ -norm having the form  $\|x\|_a = \sqrt[a]{\sum_{i=1}^n |x_i|^a}$ .

Two strategies are used in our approach to select the training examples:

- 1) **Uniform** strategy. In this strategy, we consider that the features of a patient remain constant throughout the whole monitoring phase, e.g. the weight and age of a patient are supposed to be unchanged. Therefore, the value of  $\varepsilon$  follows  $\varepsilon_1^j = \varepsilon_2^j = \dots = \varepsilon_{n_j}^j$ . In practice, we take the average values of each feature from all the samples of a patient, compare in a brute force way to find the *closest* example of the patient in the patients library. This guarantees the experiments to be closer to the real scenarios.
- 2) **Discriminatory** strategy. When we do not have sufficient training patients data, the uniform strategy will give less accurate results because the chances of choosing an example that is *far* from the new patient are high. Therefore, we propose this discriminatory strategy, in which we treat each sample of a new patient as a separate set of data and search in the patients library the closest training samples; thus, for each sample the  $\varepsilon$  could be different. Moreover, the use of this strategy could give more precise predictions, especially in the case of a smaller number of training data in the patients library.

Both strategies try to find the *closest* training samples in the patients library to build the model for a new patient under different situations. The selection of the training data via feature comparisons is the simplest and fastest way to remove the outliers. When the data increases, the first strategy is faster than the second one but requires the model for a patient to remain unchanged. The second strategy tries to find the *closest* data for each sample of a new patient, and is therefore more time-consuming, but it is more accurate when doing predictions of drug concentration.

After obtaining  $\varepsilon$  for each patient,  $w$  could then be

calculated by:

$$w = \sum_{j=1}^N \sum_{i=1}^{n_j} \varepsilon_i^j \alpha_i^j \phi(x_i^j) \quad (12)$$

where the equation still satisfies the  $\varepsilon_i^j \alpha_i^j \geq 0$ . For those patients samples with  $\varepsilon_i^j = 0$ , they are regarded as *faraway* data in advance for a specified new patient in order not to affect the personalized model. Hence equation (9) could be written as:

$$\begin{bmatrix} (\mathbf{K} + \frac{1}{C_0}I)\varepsilon & 1 \\ \varepsilon & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ b \end{bmatrix} = \begin{bmatrix} y \\ 0 \end{bmatrix} \quad (13)$$

The values of  $\alpha$  and  $b$  could be computed similarly by adding  $\varepsilon$  to Equation (10). The accuracy of the method has been shown experimentally in Section 4 to be more precise and stable than the traditional PK modeling or LS-SVM method.

Once the system gets the prediction of the drug concentration, it checks the value to be effective or not according to the drug's therapeutic range table. If it is above the range, the dosage has to be reduced. After considering the amount of reduction, the system will redo the modeling step to give a new concentration prediction. The dose has to be increased respectively for the case of under-dosing.

#### IV. EXPERIMENTS AND COMPARISONS

Our approaches have been evaluated on a set of data related to the anticancer drug *imatinib*, which was designed to treat chronic myeloid leukemia and gastrointestinal stromal tumors[20]. The training data set consists of 54 patients and 252 samples (obtained at time of a previous clinical trial[20]), while the validation set contains the data of 65 patients and 209 samples (patients followed latter on a routine basis, in the context of an amendment of the initial trial). We compare the prediction results of our system to the ones using a general population pharmacokinetic model.

As shown in equations (8), (10) and (11), we need to determine several parameters both for LS-SVM and the Example-based SVM method as follows:

- 1)  $\mathbf{K}$ : the effectiveness of the SVM method depends on the choice of the kernel function. A common choice is to use a Gaussian Kernel which has a single parameter  $\sigma$  that has to be estimated.
- 2)  $C_0$  and  $\sigma$ : after selecting the kernel, the performance of SVM highly depends on the kernel parameters and the margin factor  $C_0$ . The best combination of  $C_0$  and  $\sigma$  is found by a grid-search with exponentially growing sequences via 10-fold cross validation. In practice, our  $C_0, \sigma \in \{10^{-2}, 10^{-1}, \dots, 10^3, 10^4\}$ . The 10-fold cross validation is done by separating randomly the original training data into 10 subsample groups. In our experiment, one of the subsamples is used as the set of validation data while the remaining 9 groups are used as training sets.
- 3)  $f_s$ : determines the features used to find the *closest* training examples. In our experiments, the selected feature set is composed of {gender, age, weight}.
- 4)  $\beta$  and  $a$ : both parameters  $\beta$  and  $a$  are critical to the selection of training data by determining the value of  $\varepsilon$ .

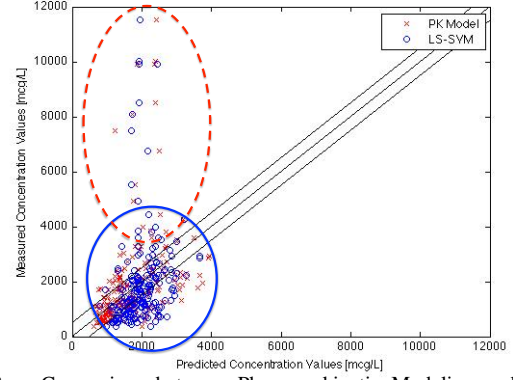


Fig. 3. Comparison between Pharmacokinetic Modeling and LS-SVM Methods based on 100% Utilization of the Patients Library,  $x$ -axis: Predicted Concentration Values,  $y$ -axis: Measured Concentration Values.

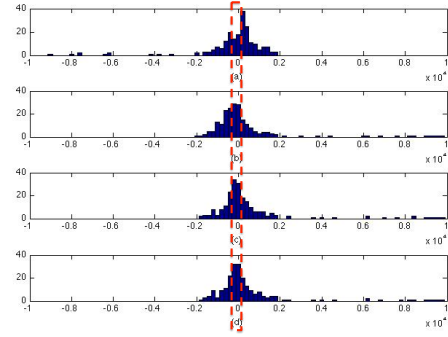


Fig. 4. Histogram of the Mean Absolute Difference in the Drug Concentration Predictions Among the four Approaches: (a) PK Model, (b) LS-SVM, (c) E-SVM Uniform, (d) E-SVM Discriminatory [bar unit = 200mcg/L]

In our experiments, both values are also estimated by 10-fold cross validation on the training data set before running the SVM approaches.

- 5)  $t_s$ : is the threshold used to decide whether one training data is an outlier or not. This parameter could be set as a constant regardless of the different physical conditions of the testing patients. However, to make our system adaptive to each specific patient, we choose a specific value that for each data set. Practically, it performs as a filter to pass only the first  $M\%$  *closest* training data.
- 6)  $f_d$ : decides the features to be used in the training data samples and to perform the prediction of drug concentration as well. As the main advantage of using SVM-based approaches, adding or removing one feature in building up the model is simpler than the classical pharmacokinetic method. The features to be considered as training data are: {Amount of Dosage, Time of Measuring, Gender, Age, Body Weight}, which correspond to {A, B, C, D, E} in Table I. Additionally, we also compare the influence of using subsets of these features.

Figure 3 shows the comparison of drug concentration predictions ( $x$ -axis) versus the measured concentration ( $y$ -axis) between LS-SVM method and the traditional pharmacokinetic modeling approach using 100% of the data in the patient library. From this figure we see that the prediction based on LS-SVM provides results similar to the ones of the

TABLE I  
COMPARISONS OF THE MEAN ABSOLUTE DIFFERENCES AMONG LS-SVM (LS), E-SVM UNIFORM (E1) AND E-SVM DISCRIMINATORY (E2)  
(FEATURES: A-DOSAGE, B-MEASURING TIME, C-GENDER, D-AGE, E-BODY WEIGHT) [UNIT :  $mcg/L$ ]

M% Features	100%			70%			50%			30%			10%		
	LS	E1	E2	LS	E1	E2	LS	E1	E2	LS	E1	E2	LS	E1	E2
A+B	<b>822.3</b>	850.4	<b>830.6</b>	836.3	858.2	<b>832.4</b>	832.7	857.0	<b>843.5</b>	849.9	971.9	957.0	<b>938.9</b>		
A+B+C	835.3	848.0	837.6	838.3	893.5	833.6	833.5	856.7	848.1	848.9	982.1	957.2	942.3		
A+B+C+D	846.5	867.1	851.8	855.8	834.6	876.3	853.5	951.4	876.7	859.9	980.6	986.8	960.3		
A+B+D	853.0	878.2	853.4	852.8	890.1	853.3	851.5	904.0	853.1	868.1	973.4	953.5	945.6		
A+B+E	853.8	837.5	860.2	851.6	882.2	855.9	852.0	880.5	872.1	868.1	995.8	964.8	940.7		
A+B+D+E	868.1	880.6	889.3	885.3	882.1	909.5	893.1	860.7	895.8	890.4	979.0	991.9	963.6		
A+B+C+D+E	903.6	882.7	879.1	875.2	867.8	915.1	884.9	855.5	891.3	885.7	984.2	991.9	963.6		
A+B+C+E	849.1	865.6	851.9	855.8	854.0	876.3	853.5	926.4	876.7	859.9	975.9	986.8	960.3		

traditional PK model in predicting a high drug concentration values as shown in the red dash ellipse. However, as to the data in the blue solid circle, the Figure 4 shows that the three SVM-based approaches are better than the PK modeling method at predicting the drug concentration to be close to the measured value. The red dash rectangle highlights the number of the prediction difference within  $[-200, 200]mcg/L$ , in which the number of predictions for the PK modeling method is 78.1% less than the LS-SVM method, and the number for LS-SVM method is 14.0% and 12.3% less than the two example-based methods respectively.

Another disadvantage of using traditional PK model is that it is often unable to consider Boolean inputs. Neither could it be modified easily to analyze the importance of each feature. Nevertheless, SVM-based approaches could deal with such problems. As shown in Table I, we compare the mean absolute difference between the measured drug concentration and the predicted ones using LS-SVM (LS), E-SVM Uniform (E1), and E-SVM Discriminatory (E2). Feature A and B stand for dosage and measuring time after one dose respectively, which are thought to be the key features and are used through all the experiments. Using the PK model, the mean value of the differences between the predicted concentrations and the real values is 842.1. Hence the 3 SVM-based methods provide predictions of similar accuracy.

Moreover, Table I reveals that the best (lowest) differences for each subset of the data are the ones using only Feature A and Feature B. Therefore, when applying the SVM method to build a model for a new patient with 100% of the library, knowing only the dosage and the measuring time is sufficient to obtain comparable results, while the PK model depends on analyzing all the parameters needed in Equation (2), (3) or (4). In addition, with a reduced library of patients, the two E-SVM approaches surpass the LS-SVM thanks to a search of the *closest* patients (or samples) to be the training data set instead of a random selection. Table I also shows that as an overall, E2 method obtains better results than E1 for 75% of all the experiments, while E1 is better than E2 if we consider the features {Gender, Age, Body Weight} only during the training patients' selection phase.

## V. CONCLUSIONS

Assisting the prediction of drug concentrations in clinical situation with support vector machine was proved to yield some similar results as the traditional pharmacokinetic model. Despite its advantages, the limitation of this approach resides in the high dependency of the results to the quality of the

initial data obtained from patients. Indeed, no assumption on the theoretical disposition of the drug in the body, according to the classical pharmacokinetic behavior of drugs, is made. Under certain conditions, it, however, improves the predictions using fewer input features. Two example-based SVM strategies have been analyzed in the paper and both are superior to the LS-SVM method when dealing with a reduced library of patients. This approach however deserves formal clinical validation with more data sets from various drugs undergoing TDM. Future work has also shown that SVM-based approached could build personalized model for each individual patient.

## ACKNOWLEDGMENT

The authors would like to thank Alena Simalatsar and Carlotta Guiducci from EPFL for the help in manuscript revision, Thierry Buclin and Verena Gotta from CHUV for the precious suggestions on clinical data modeling.

The research work presented in this paper is funded by the ISyPeM Project "Intelligent Integrated Systems for Personalized Medicine", with a grant from the Swiss Nano-Tera.ch initiative, evaluated by the Swiss National Science Foundation.

## REFERENCES

- [1] David W. A. Bourne, 'Mathematical Modeling of Pharmacokinetic Data', Technomic Publishing Company, Inc., 1995.
- [2] Reginald F. Brown, 'Compartmental System Analysis: State of the Art', IEEE Transactions on Biomedical Engineering, vol. BME-27, No. 1, January 1980, pp. 1-11.
- [3] Antonello L. G. Caruso, Thomas W. Bouillon, Peter M. Schumacher, Manfred Morari, 'On the Modeling of Drug Induced Respiratory Depression in the Non-Steady-State', 30th IEEE EMBS Conference, Canada, August 2008, pp.5564-68.
- [4] John F Kalafut, Corey A. Kemper, Pal Suryani, U. Joseph Schoepf, 'A Personalized and Optimal Approach for Dosing Contrast Material at Coronary Computed Tomography Angiography', 31st IEEE EMBS, USA, September 2009, pp. 3521-24.
- [5] Shinichi Tsuchiawata, Kiyoshi Mihara, Akifumi Yafune, Hiroyasu Ogata, 'Evaluation of Bayesian Estimation of Pharmacokinetic Parameters', Ther Drug Monit, vol 27, no. 1, February 2005, pp.18-24.
- [6] Gary Blau, Seza Orcun, 'A Bayesian Pharmacometric Approach for Personalized Medicine - A Proof of Concept Study With Simulated Data', Proceedings of the 2009 Winter Simulation Conference, pp.1969-76.
- [7] Elthem Alpaydin, 'Introduction to Machine Learning (Adaptive Computation and Machine Learning)', MIT Press 2004, ISBN 0-262-01211-1.
- [8] Karl S. Ni, Truong Q. Nguyen, 'Image Superresolution Using Support Vector Regress', IEEE Transactions on Image Processing, vol. 16, June 2007, pp. 1596-1610.
- [9] Laura Auria, Rouslan A. Moro, 'Support Vector Machines (SVM) as a Technique for Solvency Analysis', Discussion Papers, DIW Berlin, August 2008.
- [10] Beal SL, Sheiner LB, Part VII - Conditional Estimation Methods. In: Beal SL, Sheiner LB, Boeckmann A, Bauer RJ, editors. NONMEM User's Guides. (1989-2009). Icon Development Solutions, Ellicott City, Maryland, USA, 2009; 1998.
- [11] Guangyu Zhu, Dawei Liang, Yang Liu, Qingming Huang, Wen Gao, 'Improving Particle Filter with Support Vector Regression for Efficient Visual Tracking', IEEE ICIP2005, pp.422-5.
- [12] Leo Breiman, J. H. Friedman, R. A. Olshen, C. J. Stone, 'Classification and Regression Trees', Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software, 1984, ISBN 978-0412048418.
- [13] P. Jeatrakul, K. W. Wong, 'Comparing the Performance of Different Neural Networks for Binary Classification Problems', IEEE SNLP 2009, pp.111-5.
- [14] Zhanpeng Jin, Joseph Oresko, Shimeng Huang, Allen C. Cheng, 'HeartToGo: A Personalized Medicine Technology for Cardiovascular Disease Prevention and Detection', IEEE LiSSA 2009, pp. 80-83.
- [15] J. Suykens, T. V. Gestel, J. D. Brabanter, B. D. Moor, J. Vanderwalle, 'Least Squares Support Vector Machines', World Scientific, 2002.
- [16] Steve R. Gunn, 'Support Vector Machines for Classification and Regression', Technical Report, University of Southampton.
- [17] Yoav Freund, Robert E. Schapire, 'A Short Introduction to Boosting', Journal of Japanese Society for AI, September 1999, 14(5):pp.771-780.
- [18] Aly Farag, Refaat M. Mohamed, 'Regression Using Support Vector Machines: Basic Foundations', Technical Report, CVIP Laboratory, University of Louisville, 2004.
- [19] C. C. Peck, L. Z. Benet, N. Massoud, J. G. Gambertoglio, 'Computer-Assisted Clinical Pharmacokinetics', Pharmacokinetic Basis for Drug Treatment, New York: Raven Press, pp.353-54.
- [20] N. Widmer, L.A. Decosterd, C. Csajka, S. Leyvraz, M. A. Duchosal, A. Rosselet, B. Rochat, C. B. Eap, H. Henry, J. Biollaz, T. Buclin, 'Population Pharmacokinetics of Imatinib and the Role of  $\alpha_1$ -Acid Glycoprotein', Br J Clin Pharmacol 2006; 62:1 pp.97-112.