# Parameterized SVM for Personalized Drug Concentration Prediction*

Wenqi You[1], Alena Simalatsar[2] and Giovanni De Micheli[3]

*Abstract*— This paper proposes a parameterized Support Vector Machine (ParaSVM) approach for modeling the Drug Concentration to Time (DCT) curves. It combines the merits of Support Vector Machine (SVM) algorithm that considers various patient features and an analytical model that approximates the predicted DCT points and enables curve calibrations using occasional real Therapeutic Drug Monitoring (TDM) measurements. The RANSAC algorithm is applied to construct the parameter library for the relevant basis functions. We show an example of using ParaSVM to build DCT curves and then calibrate them by TDM measurements on *imatinib* case study.

## I. INTRODUCTION

Drug administration is one of the most common clinical routines in hospitals. Once a new patient is admitted, an initial drug dose and intake interval must be prescribed, which we call an *a priori* drug dose adaptation. In current clinical practice, the dose and its intake interval are chosen on the basis of averaging the population values and doctors' medical experience. This approach does not guarantee a proper prescription to a patient with condition different from those patients based on whose samples the library datasets were built. Therefore, such an experience-driven method is not suitable for many kinds of drugs, i.e. drugs for treating HIV or cancer. These drugs have a relatively narrow effective therapeutic range for the value of drug concentration in blood. Since this value after the same administrated dose to diverse patients differs due to inter- and intra- patients' variability, it may result in under- or over-dosing patients. If under-dosed, patients may have a suboptimal response to the drug treatment, while overdosing may expose them to a risk of toxicity.

Therapeutic Drug Monitoring (TDM) [1] is an approach aimed at controlling the drug concentration in a patient's blood over time by performing several realtime measurements to further adjust the dose, which we call an *a posteriori* adaptation. TDM, however, has several disadvantages: it is expensive, time-consuming (it may take a day to obtain one measurement value) and requires an invasive measurement of an individual drug concentration. Therefore, a novel approach that can overcome these drawbacks, gives relatively accurate predictions of drug concentrations and allows calibration of the prediction every time a new measurement of drug concentration is available is required.

[1]W. You is with Lab of Integrated Systems, Ecole Polytechnique Federale de Lausanne (EPFL) `wenqi.you at epfl.ch`

[2]A. Simalatsar is with Lab of Integrated Systems, Ecole Polytechnique Federale de Lausanne (EPFL) `alena.simalatsar at epfl.ch`

[3]G. De Micheli is with Faculty of Electrical Engineering and Head of Lab of Integrated Systems, Ecole Polytechnique Federale de Lausanne (EPFL) `giovanni.demicheli at epfl.ch`

Recent works [2] have shown a great potential of applying machine learning methods, i.e. SVMs [3], to predict the drug concentrations in blood. Compared with the traditional analytical Pharmacokinetic (PK) method [4], SVM algorithm has several advantages: it is able to process the binary/Boolean value inputs and allows us to easily add or remove features from the set. In [2] it is shown that the SVM algorithm gives a similar prediction accuracy to the one obtained by using the PK method. In addition, the influence of different patients' features on the accuracy of predicting the drug concentrations is also illustrated. In [5] an extension of the prediction algorithm by combining the RANdom SAmple Consensus (RANSAC) algorithm with SVM was proposed. This extension has improved the prediction accuracy by about 40%. A Drug Administration Decision Support System (DADSS) [6] based on this RANSAC-SVM algorithm has also been proposed. However, all these methods are able to preform only a point-wise drug concentration prediction and have no knowledge about the dependency among all predicted values for one patient with fixed parameters. Therefore, every time a new measured concentration value is available, it must be added to the initial database and the whole relearning and predicting process must be applied. It is computationally heavy, especially if we need to have complete information about the drug disposition over time. Solving similar engineering problems [7] uses regression analysis and [8] proposes a combination rule of evidence theory. However, these methods do not consider a large number of features nor future data adaptation.

In this paper, to utilize the advantages of the two, we present a novel method that combines the SVM algorithm and an analytical model which describes the DCT curve derived using a curve fitting process. The parameter library of the basis functions used for curve interpolation are computed using the RANSAC algorithm previously applied in [5] to separate inliers and outliers from a set of (noisy) data. Using this method, we are able to process as many features as possible, consider binary inputs, visualize the data represented in a form of personalized DCT curve for each patient, and adjust it structurally to make it more personalized with a given measured concentration value. Moreover, based on the SVM and RANSAC algorithm, it is also less computationally expensive and more precise (36% more accurate) than the PK method [4]. We can further extend the DCT curve from the single dose regimen to multiple-dose by calibrating each subsequent computed curve with the value of the residual drug concentration after previous intake.

The paper is organized as follows: Section II discusses the background information of a specific drug *imatinib*
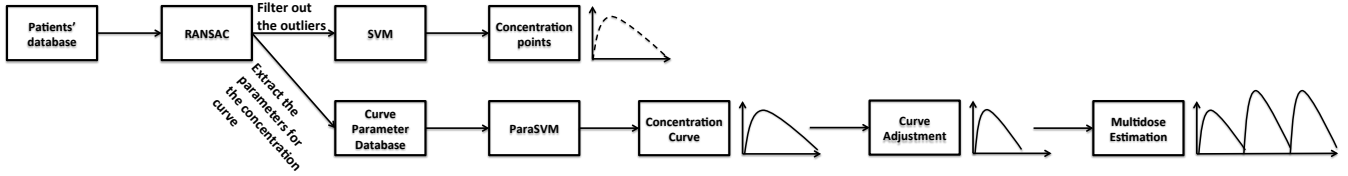
Fig. 1. Flow of Applying Support Vector Machine Algorithms in Building the Drug Concentration Curves. First row: point-wise concentration value prediction approach; second row: parameterized Support Vector Machine approach.

considered in this work. Section III describes the proposed parameterized SVM approach, the curve-adjusting method and the extension from the single dose regimen to the multiple dose one. Section IV gives a brief conclusion.

## II. BACKGROUND INFORMATION

*Imatinib* [4], a drug used to treat chronic myeloid leukemia and gastrointestinal stromal tumors, is the drug considered in this paper. Until now, only a trough's therapeutic range of this drug has been proposed and is presently being validated in a randomized clinical study in leukemia patients (I-COME; ISRCTN31181395). Relying on the data presented in [4], [1] we assume that the trough range has a lower bound at $750\mu g/L$ and upper bound at $1500\mu g/L$. The training data used in this paper includes 54 patients and 252 samples in total, while the testing data contains 65 patients presenting gastrointestinal stromal tumors with 209 samples. The set of input features of patient profile data includes: Gender, Age, Body Weight, and Type of pathology.

In the current clinical practice, a standard prescription of $400mg$ dose every 24 hours is administrated to patients in the chronic phase. Therefore, the trough concentration value is considered to be measured at 24h (supposing the initial dose is taken at the time point 0h). In the TDM procedure [9], a blood sample is taken and the drug concentration in blood is analyzed several hours after the patient takes a dose. If the analysis result shows a tendency that the trough concentration value will exceed the upper bound of the therapeutic range ($1500\mu g/L$), the clinician then reduces the dose for the patient, and vice versa.

As mentioned in Section I, in [2] we proposed to use the SVM algorithm to predict the drug concentrations at any time point taking into account the patients' personal features. Compared to TDM, it is a fast, non-invasive and less expensive method. It is able to give a personalized amount and interval for the initial dose according to the therapeutic range of the trough concentration [6]. However, multiple computations can only resemble the form of the DCT curve, as illustrated in the first row of Figure 1. This point-wise prediction approach does not capture the structure information of the DCT curve itself, hence causing difficulties in adjusting the curve, i.e. when measured concentration values are available. Further in this paper we describe a parameterized SVM (ParaSVM) approach which, based on SVM, also gives the structural information of the DCT curve, including a curve adjusting method and the construction of the DCT curve after multiple doses.

## III. ALGORITHMS

This section mainly discusses the methods used to realize the second row of Fig. 1. Unlike applying the SVM algorithm to predict the concentration values directly after removing the outliers using the RANSAC algorithm as shown in the first row, here we first utilize the RANSAC algorithm to compute the parameter database, or parameter library (III-A), then apply a parameterized SVM to predict the parameters used to construct the DCT curve for each patient based on their personal features (III-B), further adjust the curve parameters to better illustrate the concentration variation with one measured concentration value, or *a posteriori* adaption, (III-C), and finally construct the DCT curves for multiple-dose cases. The algorithms proposed in this section are all developed for the drug *imatinib*; however, the methods can be applied to any other drug concentration analysis.

### A. Parameter Library

To apply the SVM algorithm, we need to have a set of training data being the library, which contains training inputs (patients' features) and training outputs (DCT curve parameters). This library is built using the RANSAC algorithm [10], which was originally used to separate inliers and outliers from a set of (noisy) data with respect to given basis functions. As presented in [5], the algorithm is randomly selecting a very small subset of the given input data, computes the parameters (or weights) of each basis function considering the small subset, and then determines the inliers and outliers for the rest of the data with a given distance value (threshold).

Here, instead of separating inliers and outliers, the RANSAC algorithm is applied to compute the parameters of the basis functions for each patient. To remove the outliers and keep enough data samples to build the subset at the same time, for each patient we take into account all his/her samples in addition to the randomly-selected samples from the rest of the patients to build the subset. We use the common basis functions $\beta^j = \{t^{-2}, \log(t), 1 - e^{-t}\}$, respecting the shape of DCT curve obtained from the PK method [4], where $t$ stands for time [5]. Therefore, the target is to obtain the parameters $y$ for the weights of $\beta$:

$$f_{concentration} = y \cdot \beta = \begin{bmatrix} y^1 & y^2 & y^3 \end{bmatrix} \begin{bmatrix} \beta^1 \\ \beta^2 \\ \beta^3 \end{bmatrix} \quad (1)$$

These parameters $\{y^1, y^2, y^3\}$ together with patients' features form the Parameter Library being used as the training data.

## B. Parameterized SVM (ParaSVM)

Instead of predicting the drug concentration values [2], we apply the SVM algorithm to learn the mathematic relationship between the parameters of the basis functions extracted in Section III-A and then predict the parameter values of the DCT curve for a new patient in the testing dataset. To build one DCT curve in our case, three parameters are required. Therefore, the main difference is the inclusion of minimizing the objective function which considers a combined difference between the predicted parameter values and the ones in the parameter library (training dataset).

In the case of modeling $N$ patient samples, the form of patient samples becomes $(x_i, y_i^1, \cdots, y_i^j, \cdots, y_i^{N_P})$, where $i$ is the ID of a sample $i \in \{1, 2 \cdots, N\}$, $x_i$ represents the feature values of $i$-th patient, $y_i^j$ denotes the $j$-th parameter value of this patient, and $N_P$ is the number of parameters. The goal is to find $N_P$ linear functions $f^j(x) = w^j \cdot \phi^j(x) + b_j$ to describe the relationship in between the dataset points and to estimate the parameter value $y_i^j$ according to a new input dataset. Based on [2], we minimize the following modified objective function:

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

where $H$ takes into account the combined difference of all three predicted values plus the ones in the parameter library. Note that this objective function has Root of Sum of Square (RSS) fitting error and a regularization term, which is also a standard procedure for the training of Multi Layer Perceptron (MLP) and is related to ridge regression [11], [12]. Applying Lagrangian analysis to solve the optimization problem of the objective function, we obtain $w$ as: $w^j = \sum_{i=1}^{N} \alpha_i^j \phi^j(x_i)$, which gives a linear system together with Equation(2):

$$\begin{bmatrix} \mathbf{K}^j + \frac{1}{C_0}I & 1 \\ 1^T & 0 \end{bmatrix} \begin{bmatrix} \alpha^j \\ b^j \end{bmatrix} = \begin{bmatrix} y^j \\ 0 \end{bmatrix} \quad (3)$$

where each entry of the kernel matrix $\mathbf{K}^j$ is defined to be $K_{ab}^j = \phi^j(x_a)^T \phi^j(x_b)$. A Gaussian Kernel is applied in a similar way as in [2]. Therefore, the prediction function for the DCT curve parameters becomes: $Para^j(x) = \sum_{i=1}^{N} \alpha_i \mathbf{K}^j(x_i, x) + b^j$. Table I compares the prediction accuracy in absolute differences of the proposed ParaSVM, RANSAC-SVM[5] (RSVM) and the PK method towards the measured concentration values. The percentages indicate the improvement of the methods when comparing with the PK one. It shows that ParaSVM improves the average prediction accuracy by about 36.13% to PK method, which is 5% less accurate than the performance of RSVM. However, ParaSVM further provides a way of adapting the DCT curve structurally as discussed in the following subsections.

## C. The a posteriori Adaptation

In an *a posteriori* dose adaptation, we refine the predicted DCT curve computed by ParaSVM. It is done by calibrating the current DCT curve using one or several measured data

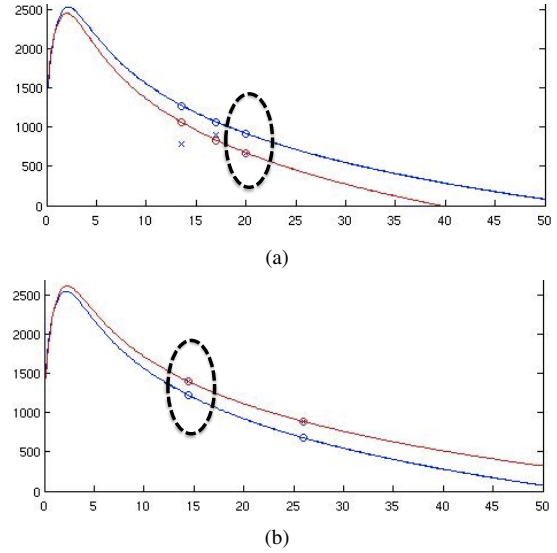| | PK | RSVM [5] | | ParaSVM | |
|---|---|---|---|---|---|
| Mean [$\mu g/L$] | 421.42 | 247.61 | 41.24% | 269.16 | 36.13% |
| STD [$\mu g/L$] | 384.55 | 211.31 | 45.05% | 259.82 | 32.44% |



(a)



(b)

Fig. 2. Two Examples of Parametrically Refined Drug Concentration to Time Curves for the Drug *imatinib*. Blue curve: DCT curve computed by the ParaSVM approach; Red curve: DCT curve adjusted with one measurement; Crosses: measured concentrations; Circles: predicted concentration based on the DCT curve. Dotted black circle: measured point chosen for curve refinement. X-axis: time [$h$], Y-axis: concentration value [$\mu g/L$]

point under certain constraints. Taking into account that these measurements are done for the same patient, the calibration makes the DCT curve more personalized with each new measurement. The same DCT curve adjustment approach is also applied to build the concentration curve for multiple dose regimens using the computed trough concentration value from the previous cycle (computation) as a measurement.

Once there is a new measurement we first predict the basis function's parameters using ParaSVM and then search within a certain radius $\delta D$ around each parameter value with a step $\delta d$ to find the best set of parameters that satisfies the following conditions:

- The modified DCT curve has to pass through the given measured concentration value;
- After giving the dose, the concentration value should increase with time: $\frac{\partial f_{concentration}}{\partial t}|_{t=0} > 0$;
- After several hours, the concentration value reaches the peak value and starts to decrease: $\frac{\partial f_{concentration}}{\partial t}|_{t=T_p} < 0$, where $T_p$ is a time point after the peak value, i.e. we set it as $T_p = 24h$.
- Considering the trough value or residual value, from the previous dose, the difference between the starting value of DCT curve ($t = 0$) and the ending one ($t = 24h$, since *imatinib* is usually administrated once a day), should be within a certain range (R), i.e. $< 50\mu g/L$: $|f_{concentration}^{t=0} - f_{concentration}^{t=24}| < R$.
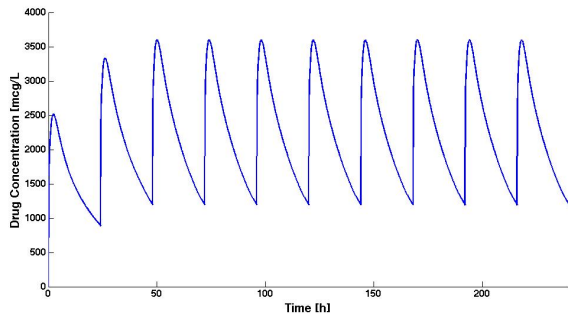- The DCT curve whose shape is the closest to the curve previously predicted by ParaSVM will be chosen:

Fig. 3. Example of Multiple dose Estimation for the Drug Concentration to Time Curve Over 10 Days of Drug *imatinib*. X-axis: time [*h*], Y-axis: concentration value [$\mu g/L$]

$\min_{g_r} \sum_{j=0,\cdots,N_s} (g_r^{t=j} - g^{t=j})^2$, where $g^{t=j}$ stands for the concentration value at time $j$ and $g_r^{t=j}$ is the one in the refined curve. The set of parameters $y$ corresponding to the best $g_r$ are selected.

Fig. 2 shows two examples of parametrically refined DCT curves for the drug *imatinib*. The upper curve (blue) in Fig. 2(a) and the lower curve in Fig. 2(b) are computed by ParaSVM, on which the points (blue) are the drug concentrations that correspond to the measuring time in the testing database. Similarly, the lower curve (red) in Fig. 2(a) and the upper curve (red) in Fig. 2(b) are curves adjusted with one measurement meeting the above constraints, on which the points (red) are the concentrations after the adjustment. The crosses are the measured concentration values in the testing database. The black dotted circle in both figures represent the measured point chosen for the adjustment process. From these figures one can see, the proposed ParaSVM approach is able to minimize the difference between the predicted concentration values and the measured ones.

In the example of Fig. 2(a), the refined DCT curve with respect to the measurement in the dotted circle improves the prediction accuracy for the other two measurements. In Fig. 2(b), the refined DCT curve also goes through the second measured value, while the original predicted curve predicts the concentrations to be smaller than the measured ones.

### D. Multiple Dose Estimation

Knowing how the concentration varies with time after multiple doses is important to clinicians and patients in order to monitor a long term therapeutic procedure. Fig. 3 shows an example of estimating the drug concentration over 10 days based on ParaSVM taking into account the residual drug concentration from the previous day. In an *a posteriori* case, after we adjust the concentration curve with a given measurement, the multi-dose DCT curve can be obtained by computing over days the updated one-dose DCT curve. As the drug sometimes takes more than 24 hours to be clear from the human body, we also consider the residual concentration values from the previous dose(s).

The DCT curve for multiple doses makes it visually easy to obtain the peak and trough concentration values and to check whether they are within the therapeutic ranges or not. As shown in Fig. 3, the residual concentration affects both the peak and the trough concentration values, especially at the beginning of the treatment.

## IV. CONCLUSIONS

In this paper, we have presented a novel approach, ParaSVM, that complements the SVM method with an analytical model , which allows us to consider various patient features, build the DCT curve, adjust it with occasional real TDM measurements, and construct the multi-dose curve afterwards. The main contribution of this work is the combination of an implicit machine learning algorithm with an explicit analytical model. The ParaSVM model shows slight reduction in the prediction accuracy (5%) compared with SVM approach, however, it is still more than 35% accurate than the PK method.

## ACKNOWLEDGMENT

## REFERENCES

[1] V. Gotta, N. Widmer, L. Decosterd, C. Csajka, M. Duchosal, Y. Chalandon, D. Heim, M. Gregor, and T. Buclin, "Therapeutic drug monitoring (TDM) of imatinib: Effectiveness of bayesian dose adjustment for CML patients enrolled in the imatinib concentration monitoring (I-COME) study," ser. Swiss Medical Forum, 2010, p. 285.

[2] W. You, N. Widmer, and G. De Micheli, "Example-based support vector machine for drug concentration analysis," in *Engineering in Medicine and Biology Society,EMBC, 2011 Annual International Conference of the IEEE*, Aug. 30-Sept. 3 2011, pp. 153 –157.

[3] C. Cortes and V. N. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[4] N. Widmer, L. Decosterd, C. Csajka, S. Leyvraz, M. A. Duchosal, A. Rosselet, B. Rochat, C. B. Eap, H. Henry, J. Biollaz, and T. Buclin, "Population pharmacokinetics of imatinib and the role of $\alpha_1$-acid glycoprotein." *British Journal of Clinical Pharmacology*, vol. 62, no. 1, pp. 97–112, 2006.

[5] W. You, A. Simalatsar, and G. De Micheli, "Ransac-based enhancement in drug concentration prediction using support vector machine," in *Proceedings of the International Workshop on Innovative Simulation for Healthcare (IWISH)*, September 2012, pp. 21 – 5.

[6] W. You, A. Simalatsar, N. Widmer, and G. De Micheli, "A drug administration decision support system," in *Bioinformatics and Biomedicine Workshops, IEEE International Conference*, Oct 2012, pp. 122 – 9.

[7] K. V. Kumar, K. Porkodi, and F. Rocha, "Isotherms and thermodynamics by linear and non-linear regression analysis for the sorption of methylene blue onto activated carbon: comparison of various error functions," *J Hazard Mater*, pp. 794–804, Mar. 2008.

[8] R. Sun, H.-Z. Huang, J. Yang, D. Ling, and Q. Miao, "Curve fitting with weight assignment under evidence theory combination rule," in *Quality, Reliability, Risk, Maintenance, and Safety Engineering, 2011 International Conference on*, june 2011, pp. 929 –934.

[9] T. Buclin, V. Gotta, A. Fuchs, N. Widmer, and J. Aronson, "An agenda for UK clinical pharmacology monitoring drug therapy," *British Journal of Clinical Pharmacology*, vol. 73, no. 6, pp. 917–23, 2012.

[10] M. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[11] G. Golub and C. F. Van Loan, *Matrix Computations*. John Hopkins University Press, 1989.

[12] J. D. Brabanter, "Ls-svm regression modelling and its applications," *PhD thesis*, pp. 25–28, 2004.