BioNanoSci. (2013) 3:378–393
DOI 10.1007/s12668-013-0103-8

Personalized Drug Administrations Using Support Vector Machine

A New Approach in Computer-Aided Dose Analysis

Wenqi You · Alena Simalatsar · Nicolas Widmer · Giovanni De Micheli

Published online: 22 August 2013

Abstract The decision-making process regarding drug dose, regularly used in everyday medical practice, is critical to patients' health and recovery. It is a challenging process, especially for a drug with narrow therapeutic ranges, in which a medical doctor decides the quantity (dose amount) and frequency (dose interval) on the basis of a set of available patient features and doctor's clinical experience (a priori adaptation). Computer support in drug dose administration makes the prescription procedure faster, more accurate, objective, and less expensive, with a tendency to reduce the number of invasive procedures. This paper presents an advanced integrated Drug Administration Decision Support System (DADSS) to help clinicians/patients with the dose computing. Based on a support vector machine (SVM) algorithm, enhanced with the random sample consensus technique, this system is able to predict the drug concentration values and computes the ideal dose amount and dose interval for a new patient. With an extension to combine the SVM method and the explicit analytical model, the advanced integrated DADSS system is able to compute drug concentration-to-time curves for a patient under different conditions. A feedback loop is enabled to update the curve with a new measured concentration value to make it more personalized (a posteriori adaptation).

Keywords Drug dose computation · Support vector machine · Decision support system

1 Introduction

In current clinical pharmacology, the initial drug dose is chosen on the basis of previous medical experiences. It can be consequently modified based on the presence of adverse events or nonresponsiveness of a patient to the treatment. However, this experience-driven method is not suitable to many kinds of drugs. There is a small group of medicines, i.e., drugs for treating HIV, cancers, etc., whose effective therapeutic concentration range is quite narrow, and therefore, there is a very high risk to under- or overdose a patient. Underdosing will lead to an ineffective treatment, while overdosing will expose the patient to a risk of toxicity. Thus, controlling the drug concentration to be within the therapeutic range is essential to properly carry out the clinical monitoring; in other words, it is necessary to know how the human body affects the drug dissipation studied by the *population pharmacokinetics* (PK). The PK studies together with the therapeutic ranges provided by the *pharmacodynamics* (PD) studies on the drug effects form the initial ground for the quantitatively justified decision making regarding the dose adaptation.

W. You (✉) · A. Simalatsar
School of Computer and Communication Sciences,
École Polytechnique Fédérale de Lausanne,
Lausanne, 1015 Switzerland
e-mail: wenqi.you@epfl.ch

A. Simalatsar
e-mail: alena.simalatsar@epfl.ch

N. Widmer
Pharmacien responsable TDM, Division de Pharmacologie et
Toxicologie cliniques CHUV, Hopital de Beaumont,
1011 Lausanne, Switzerland
e-mail: nicolas.widmer@chuv.ch

G. De Micheli
Institute of Electrical Engineering,
Integrated Systems Centre, EPFL,
Lausanne 1015, Switzerland
e-mail: giovanni.demicheli@epfl.ch

Medical decisions are critical to patients' health; the effective use of medical resources [1] and a proper medical decision may improve the quality of health care service. *Decision support systems* (DSSs) are computer-based information systems that support decision-making activities [2]. *Clinical DSSs* (CDSSs) form a special class of DSSs that are designed to aid clinical decision making according to the characteristics of an individual patient. This software is based on a computerized clinical knowledge or expertise and can generate recommendations for a specific patient [3]. In this paper, we focus on a specific class of health-care procedures that we would like to support with a CDSS, i.e., the personalized computation of a suitable drug dose for a new patient based on the prediction of the blood drug concentration taking into account patient's features. There have been several models developed in support of PK studies that are able to predict the drug concentration in the blood. These models can be classified as analytical [4] and statistical [5]. The analytical models are able to account only for the variables with real values, while binary-valued variables, such as gender, create strong discontinuities in the models and are, in general, not taken into account by the methods. Moreover, the analytical models are based on differential equations that are hard to modify if we were to add new parameters. The main drawback of the statistical approach, including the Bayesian approach [6], is that it requires knowledge on data distributions, such as mean and/or deviation values. For newly developed drugs which do not have sufficient studies on the patients, it is difficult to give a proper mean or deviation value to compute the concentrations for the following patients. In our previous works, we have briefly presented a personalized drug concentration prediction method—the *support vector machine* (SVM)-based algorithm [7, 8]. In [9], we enhanced the prediction accuracy of the algorithm. In [10], we presented a drug concentration-to-time (DCT) curve prediction approach, parameterized SVM, which combines the SVM and analytical models.

In this paper, we present a systematic combination of the various parts of the algorithm into an advanced Drug Administration Decision Support System (DADSS) that is aimed to assist medical doctors in decision making regarding the drug dose adaptation during the different phases of the treatment. We show a detailed step-by-step elaboration of the SVM-based prediction algorithm, comparison, and interrelation and evaluation of its various parts under a different point of view. DADSS is able to recommend the dose and the intake time interval for a new patient in a personalized manner. The core prediction algorithm is represented with two optional algorithms: a point-wise prediction of the drug concentration in patient blood, essential for the a priori drug dose adaptation, and the parametrized SVM that allows the modeling of the concentration–time curve instead of the prediction of unique drug concentration point. This extension also gives a possibility to study the effect of the residual drug concentration after previous intakes, which is essential for the a posteriori drug dose adaptation. The parametrized SVM algorithm does not show a major enhancement in accuracy compared to the point-wise prediction SVM algorithm with random sample consensus (RANSAC); however, it allows the visualization of the DCT curve and curve adaptation with real therapeutic drug monitoring (TDM) measurements.

The paper is organized as follows. Section 2 introduces the fundamental knowledge on clinical decision support system and the related work in domains using a support vector machine. Section 3 talks about the background information of our research, including the addressed problems, terminologies used in the paper, and the statistics of the drug *imatinib*. Section 4 presents the DADSS combining our previous work and extends to analyzing within different dose groups. Section 5 draws the conclusion of this study.

2 Related Work

In this section, we first talk about the background of decision support system. Then we introduce the SVM algorithm and its applications in various decision support domains.

2.1 Clinical Decision Support System

In the literature, there exist many definitions of a decision support system according to various purposes, within which clinical DSSs form a special type of DSSs that provide clinicians with medical guidelines of best practices in patient care according to clinical knowledge. Previous studies [11–14] on comparing clinical DSSs to the professional clinicians have concluded that using a reliable clinical DSS helps improve the patients' treatment process in effectiveness and safety. Hickling et al. [15] have also shown that some measurements of blood plasma during treatments helps increase the accuracy of blood concentration analysis for one individual. In [16], the authors have demonstrated that using the Bayesian approach other than empirical choice can reduce the number of hospital stay so as to save the cost.

When one develops a CDSS, the two main problems that need to be addressed are as follows: (1) the *medical knowledge acquisition*, which is devoted to build a medical knowledge database in a structural way, and (2) *medical knowledge representation*, which analyzes the data of the medical databases in order to produce inferences helping in medical decision making. Many tools aimed at combining knowledge acquisition and representation were developed in the past three decades. These tools form a class of knowledge-based decision support systems among which we can name PROforma [17], Prodigy [18], EON [19], GLIF [20], SAGE [21], Guide [22], GLARE [23] and Asbru [24], and [25]. They approach the problem of generalization of medical guideline (GL) representation. However, due to the big variety of the GLs and information sources as well as the continuous growth of the medical knowledge database, there has been no common standard for GL representation presented until now. The proposed integrated DADSS system will be applied as part of the computer-assisted medical guideline control flow to predict the concentration values and therefore give the advice on dose amount and time interval for a patient. It can be considered as a solid brick for a general decision support system aimed at assisting medical doctors when applying the therapeutic drug monitoring approach. For instance, it can close the verification loop of the TAT-based medical protocol representation [26] by bridging the modeling gap between treatment and patient's reaction to the treatment.

2.2 Support Vector Machines

Machine learning has been applied with some success to solve classification problems in computer vision and pattern recognition in the past few decades [27].

Four of the most representative machine learning techniques are decision trees, *neural networks* (NN), support vector machine, and *AdaBoosting* (AB). With the extension to solve regression problems, these techniques became popular in various domains such as image superresolution [28], object tracking [29], etc. Among the four, decision trees (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 [30]. 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 [31]. SVM uses a nonlinear mapping to transform the original training data into a higher dimensional space, within which it searches for the linear optimal separating hyperplane, or "decision boundary" to separate the two classes [32]. It is convenient both due to its clear mathematical 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 [33].

Here, we have chosen the SVM technique for our modeling system because of its appropriate complexity, efficiency, and strength in data regularization [34]. It was invented by Vapnik in 1979 and applied to classification and regression problems in 1995 [35]. Besides common areas such as object recognition, handwritten digit detection, etc., SVMs have also been applied in decision support systems where prediction-based decision making is required. In [36], the authors use an SVM and an artificial neural network as bases for their heart disease classification DSS. The SVM was used to separate the disease data into two classes, showing the presence or absence of heart diseases with 80.41 % accuracy. In [37], the authors propose a medical diagnosis DSS with an extension to the SVM algorithm to classify four types of acid–base disturbance. Besides clinical cases, SVMs have also been used in DSSs for hard landing of civil aircrafts [38], electric power information systems [39], etc. While all these works rely on the classification ability of SVMs, in our paper, we will present a DSS for drug administration using SVMs for regression [40] to predict the drug concentration in the blood and then use it to compute an appropriate dose and a dose administration interval for a chosen patient in our decision support system.

3 Background Information

In this section, we introduce the background information of current therapeutic drug monitoring (Section 3.1) and the statistics of the drug *imatinib* in our research (Section 3.2). The link between our previous work and this manuscript is also presented.

3.1 Therapeutic Drug Monitoring

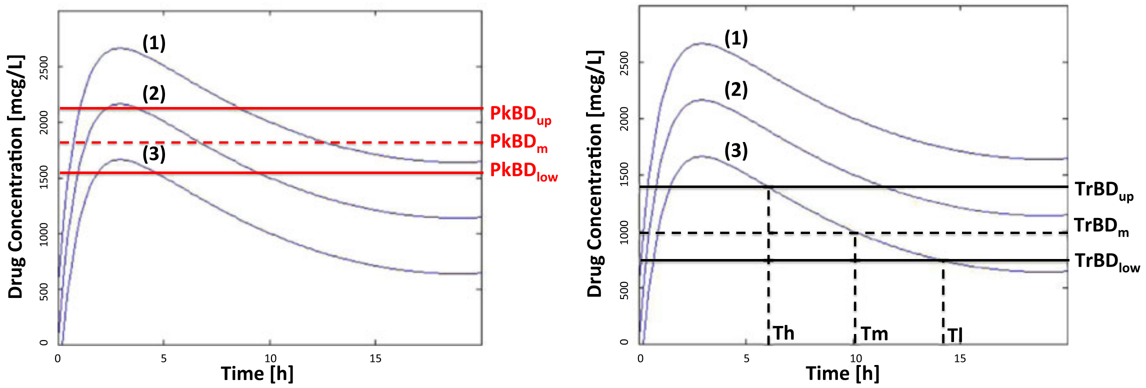
The variation of DCT in blood depends on the drug dose administrated to a patient as well as the patient's features, and this dependency is not trivial. PK is a branch of pharmacology focused on studying the drug dissipation in the human body, and pharmacodynamics (PD) is the study on the biochemical and physiological effects of drugs on the body. In clinical practice, to determine whether a patient is prescribed a right dose, doctors measure the drug concentration in blood and compare it with the drug's therapeutic range. For a drug with a narrow therapeutic range, it is easy to under- or overdose a patient. The approach that unifies the PK-PD studies in the medical practice focused on keeping the drug concentration within the limited ranges is named therapeutic drug monitoring [41]. In general medical practice, TDM specializes in the measurement of drug concentrations in blood by performing regular sampling of the blood in order to maintain its value within the therapeutic concentration ranges [42]. It is an invasive, time-consuming, and quite expensive measurement procedure. Several PK computational models [42] have been presented earlier to replace this procedure. However, due to the complexity of the human body system and many external factors that may affect the patients' health condition, it is difficult to build a general analytical model for the drug concentration prediction that would be efficient for any patient.

Ideally, two different therapeutic ranges are defined for a drug: one for the peak and another one for the trough values of the drug concentration. Figure 1a, b depicts three examples of the drug concentration curves each, which are defined by PK studies. The therapeutic peak range ($\mathrm{PkBD}_{\mathrm{low}}$ up to $\mathrm{PkBD}_{\mathrm{up}}$) and the ideal value ($\mathrm{PkBD}_m$) are defined in Fig. 1a, while the trough range ($\mathrm{TrBD}_{\mathrm{low}}$ up to $\mathrm{TrBD}_{\mathrm{up}}$) and the ideal value ($\mathrm{TrBD}_m$) are presented in Fig. 1b. The ideal drug concentration curve is the one whose peak and trough values are as close as possible to the corresponding ideal peak and trough values.

3.2 Statistics of the Drug Imatinib

Imatinib [6] is a drug used to treat chronic myeloid leukemia and gastrointestinal stromal tumors, which is considered in our study. Until now, only a trough therapeutic range of this drug has been proposed and is presently being validated in a randomized clinical study in leukemia patients (I-COME; ISRCTN31181395). The trough range has a lower bound at 750 mg/L, upper bound at 1,500 mg/L, and target value at 1,000 mg/L [43]. The available training data in our research are 251 collected from 54 patients and 209 testing data from 65 patients, which distribute with respect to different doses as shown in Table 1. The set of input features of patient profile data includes gender, age, and body weight, which, together with the “dose amount”, the “measuring time”, and the “measured drug concentrations”, consist of the data library for [7–10].

In [7], we have presented the SVM algorithm for drug concentration predictions able to account for different feature parameters of a patient, where we found that the feature measuring time is the most important feature to calculate the drug concentration (DC) values in blood. However, the mean absolute difference (MAD) between the predicted and the measured concentrations was still large. We concluded that this large difference was possibly caused by two factors: (a) measured data samples are noisy and (b) insufficient types of features are considered. In [9], we applied the random sample consensus (RANSAC) algorithm to remove the “outliers,” or noisy data, from all our data samples and analyzed the influence of using different features to find the outliers, which has improved the prediction accuracy. In [8], we presented a DADSS which uses all the predicted DC values from 1.0 to 24.0 h after a patient has taken a dose to compute the ideal dose amount and ideal time interval for this patient, namely the a priori adaptation, according to the therapeutic range. The paper also illustrated the potential feature library to be considered in the future clinical practice. However, to collect sufficient clinical data on newly proposed patient features is demanding in the sense of time, i.e., it will probably take years. Therefore, one solution is to adjust the predicted concentrations with one or more measured values, namely the a posteriori adaptation. Nevertheless, the previously proposed methods were all considered to be “point-wise” DC value predictions; hence, it is difficult to adjust all the DC values given with one measurement. The analytical model is an effective approach to overcome this problem, thanks to its explicit description of curve’s structural information. Current PK model [6] is one of the commonly used analytical models. However, the basis functions of this PK model rely, exponentially, on several other parameters such as drug absorption rate and elimination rate, which might also vary due to an intra-patient variation of the parameters [7]. On the other hand, explicit PK model makes it difficult to consider more patient features when they are available. In [10], we presented a DCT curve prediction approach, parameterized SVM, which combines the SVM and analytical models. It keeps the merits of SVM, such as being able to be extended to a larger feature library in the future, and also adds to the advantages of analytical model’s being structurally adjustable.



(a) Example of Peak Concentration Range

(b) Example of Trough Concentration Range

Fig. 1 **a** Example of peak concentration range. **b** Example of trough concentration range. An example of the drug concentration curves intersecting with the peak and trough therapeutic ranges. The dose of curve (*1*) is 800 mg; curve (*2*), 600 mg; and curve (*3*), 400 mg

Table 1 Distribution of patient samples with respect to different doses in the training and testing library

Dose unit: mg	100	200	300	400	500	600	700	800	Total
Training library	3	11	18	193	–	10	–	16	251
Testing library	1	1	7	176	1	13	–	10	209

The RANSAC algorithm has been applied to filter the outliers in both training and testing datasets, while in real clinical practice, it is hard to distinguish whether a newly measured concentration value is an outlier or is due to a sudden intra-variation happening to the patient. Under the condition that we do not have a sufficient number of features, we need to search for a curve adaptation method to help the prediction to be more personalized for one patient. In addition, all the analysis in the above work has considered the dose amount as one of the input features for SVM to predict the drug concentrations. The results in Table 1 show a large inequality distribution of the samples in different dose groups. Hence, in this paper, we also investigate the influence of different dose groups on the prediction accuracy.

4 Decision Support System

The integrated Drug Administration Decision Support System takes, the patient features, as input data, and produces, as output data, the recommendation for drug's dose amount and time interval. As shown in Fig. 2, the original DADSS system [8] can be viewed as the upper part of the flowchart which is composed of three main modules: "preprocess", "prediction", and "selection". The lower part of the flowchart shows the parameterized SVM and a feedback loop that uses the measured drug concentration values to adapt the concentration curve in order to be more personalized to a patient. The main differences between the two parts are: (a) input data: patients' original features for the upper part and the generated curve parameters for the lower part; (b) outputs from the prediction module: predicted concentration points for the upper part and predicted parameters for building the concentration curve for the lower part.

In this section, we discuss the methods used in each module with detailed experimental comparisons.

4.1 Preprocess Module

Input data, or "training" data to a model, are one of the fundamental factors in deciding whether the model will be correctly built and able to predict a future case with a reasonable accuracy. Traditional PK model considers a limited number of patient features, and, as a result, it does not guarantee a good prediction accuracy for a future patient. In our work, we apply the SVM approach, a method able to analyze a larger number of features, to tract this problem. In Section 4.1.1, a set of potentially relevant features are presented for future clinical study.

As depicted in Fig. 2, there are two types of input data. The first type is based on the initial dataset that was provided by *Cantonal Hospital of Lausanne, Switzerland* (CHUV). It contains several patient's feature data such as age, gender, and body weight for each measured concentration value corresponding to a time stamp of the measurement, as presented in Section 3.2. The drug concentration values were collected by CHUV from 2002 to 2004. An extension of patient feature data is suggested in Section 4.1.1.

The other library, *ParaLibrary*, is a derivative of the initial dataset that contains the concentration curve parameters extracted from the concentration values in the first type of library. These parameters are obtained using the RANSAC

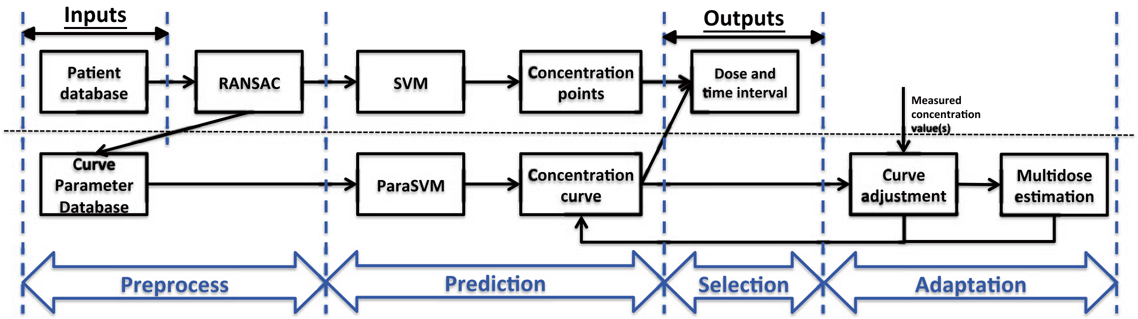


Fig. 2 Flowchart of the integrated Drug Administration Decision Support System

algorithm which is originally applied to file out the outliers of a dataset (Section 4.1.2) and modified to obtain the parameters for ParaLibrary (Section 4.1.3).

4.1.1 Extension of Patient Feature Data

As presented in the previous sections, two factors influence the accuracy of the drug concentration prediction model:

- Accuracy of a measurement, which may introduce noise in the training/evaluation datasets;
- Limitation of the patients' feature database, where an insufficient number of feature could lead to an inaccuracy of drug concentration computation.

While it is difficult to adjust the PK method such that it accounts for a bigger number of features, the SVM approach is flexible in this term and is able to account for various patient features, allowing one to study a large number (several dozens) of new parameters and determine which are the important ones. Choosing new valuable features is critical to enhance the accuracy of the drug concentration prediction, as well as it can be potentially used to enhance the existing methods. Hence, we present an extension of the feature database such that it accounts for all kinds of *standard* (such as age and weight) and *nonstandard* (not used in the current clinical practice) patient features. In Table 2, we classify the dataset features into three types: *patient profile*, *physical measurements*, and *external parameters*.

Table 2 List of the three groups of potentially relevant patient features

Patient profile
Individual features
Age, gender, weight, height, etc.
Clinical features
Diabetes, hypertension, high cholesterol, heart disease, cancer, etc.
Genomic features
Family disease history, gene polymorphism
Physical measurements
Vascular features
Blood sugar, pH value, cholesterol level, etc.
Physical features
Blood pressure, heart rate, renal function, respiration frequency, respiration rhythm, respiration deepness, etc. External parameters
Symptoms features
Vomiting, fever, dizziness, headache, convulsion, somnolence, shock, dreams, etc.
Habitual features
Amount of water, milk, smoke, alcohol, tea, coffee, sports, etc.
Environmental features
Humidity, temperatures, pressure, etc.

The patient profile class is given by the user as semi-static data which rarely change during the treatment. The physical measurements and the external parameters can be collected by using clinical tools, questionnaires, etc. [8]. Though the features are not considered explicitly in the current drug concentration prediction models, i.e., the PK method, they are critical to help in determining whether a drug has been prescribed properly. If we take *physical features* as an example, the abnormal responses of the patient might be a signal of overdosing, i.e., during the period when the patients take the drug, there is a sudden and big increase in the blood pressure.

4.1.2 RANSAC for Filtering Outliers

The *preprocess* module prepares the input data for the *prediction*. Apart from applying RANSAC to remove the outliers from all the dataset, first of all, the system checks the completion of patient features. When the features of a new patient are available only partially, it replaces the missing data by an average value of the corresponding feature in the library. Moreover, since each feature considered in clinical scenarios has different absolute values in different metrics, we normalize all the feature values using "zero mean, unit variance" technique as follows:

$$\text{norm}(\text{feature})_i = \frac{\text{feature}_i - \text{mean}(\text{feature}_i)}{\text{STD}(\text{feature}_i)} \quad (1)$$

RANSAC is applied between the above two steps.

The RANSAC [44] algorithm works as described in Algorithm 1. The number of trials M is set to be big enough to guarantee that at least one of the sets of possible inliers does not include any outlier with a high probability p. Usually, p is set to 0.99. Let us assume, that u is the probability that any selected data point is an inlier, then $v = 1 - u$ is the probability of selecting an outlier. M trials of sampling each K data points are required, where $1 - p = (1 - u^K)^M$. This implies

$$M = \frac{\ln(1-p)}{\ln(1-(1-u)^K)}. \quad (2)$$

From the above algorithm description, we could find that the RANSAC algorithm, though powerful, does rely on several empirical choices: (a) the threshold to decide whether a point is an outlier or inlier, and (b) the basis functions used to construct the structure of dataset. Figure 3 shows an analysis for different thresholds influencing the SVM prediction accuracy for the *training* data in different dose groups. RANSAC is used to select inliers of each dose group to build the SVM model, and the model is then tested on the whole training database of the same dose group. The thresholds vary from 10 to 5,000 with a step of 10 μg/L. The result shows that small threshold values exclude too many

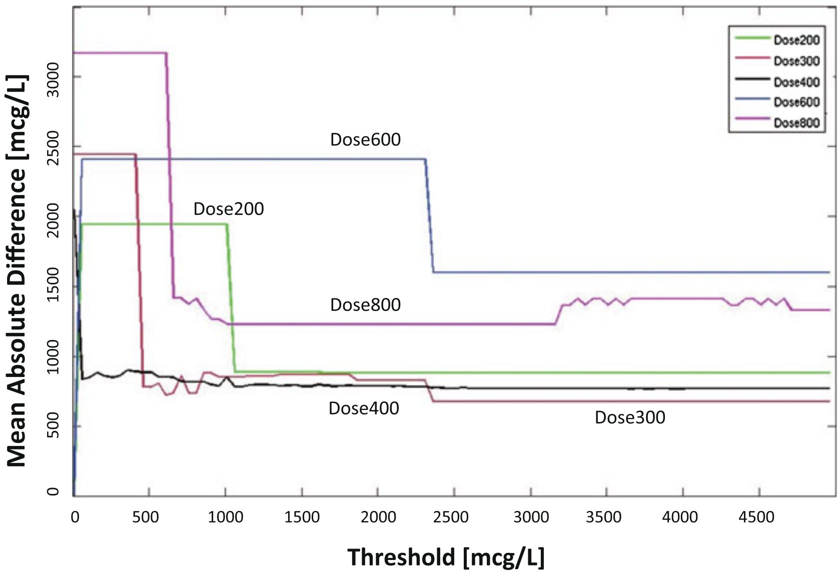


Fig. 3 The influence of prediction accuracy using different threshold values with respect to different dose groups

data points, and the resulting inliers are not enough to build a decent model to predict the concentrations.

The model of the RANSAC algorithm is a linear combination of several basis functions. The number of basis functions corresponds directly to the minimum number of points K required to fit the model. The parameters of the model are the weights of each basis function. Due to the fact that some dose groups have limited number of training data, i.e., 100- or 600-mg groups, we would like to keep the K as small as possible. In the preprocess phase, the basis of the RANSAC algorithm is chosen to be a combination of some typical functions: $\{x^{-2}, x^{-1}, x, x^2, x^3, \log(x), \cos(x), (1 - \exp(-x)), \exp(x)\}$. This requires a $K = 9$ data points to be randomly selected each time to compute the weights of these basis functions. However, in our experiments, we find that not all the listed basis functions are utilized to estimate the inliers and outliers. Table 3 shows the utilization results ("1" for "used", when the corresponding weight of the basis function is nonzero; "0" for "unused", when the corresponding weight is 0 zero) of each basis function with respect to different thresholds Th, which indicates a tolerable difference between the measured concentration and the predicted one. In clinical scenario, we expect that the threshold is as small as possible to minimize the prediction inaccuracy. Therefore, we combine the first two rows of the basis functions and get $f(x) = \{x^{-2}, x, x^3, \log(x), \cos(x), (1 - \exp(-x)), \exp(x)\}$. However, the basis functions should also reflect the distribution of the data points with a meaningful structure. To further minimize the value of K and as indicated in [6], the values of drug concentration in blood varying with time, or DCT curve, follows the specifications of a "concave" form, such that it monotonically grows in the beginning till reaching the peak value, and then starts (monotonically) decreasing. Therefore, after evaluating the shape of the above basis functions, we select $f(x) = \{x^{-2}, \log(x), (1 - \exp(-x))\}$ to be our final ones. We test it on the dose group with the largest number of data points (400 mg) for each basis function with a fixed threshold of 500 μg/L. The results are presented in Fig. 4, where the small cross points are the training data points and the green circles indicates the

Algorithm 1 RANSAC algorithm, where *data* is a set of observations (concentration values w.r.t. different time in our case), *model* is a model that can be fitted to the data, K is the minimum number of data points required to fit the model parameters, M is the number of trials performed by the algorithm, Th is a threshold determining if a data point fits a model, and *bestmodel* is the model fitting the highest number of data points.

```
Input: data, model, K, M, Th
Output: bestmodel
  bestinliers ← Ø
  for $i = 1 \rightarrow M$ do
    possibleinliers ← SampleUniformly(data, K)
    possiblemodel ← Fit(model, possibleinliers)
    inliers ← Ø
    for all point ∈ data
      if Distance(point, model) < $Th$ then
        inliers ← inliers ∪ {point}
      end if
    end for
    if |inliers| > |bestinliers| then
      bestinliers ← inliers
    end if
  end for
  return bestmodel ← Fit(model, bestinliers)
```

Table 3 RANSAC basis function analysis with respect to different thresholds

Th	x^{-2}	x^{-1}	x	x^2	x^3	$\log(x)$	$\cos(x)$	$1-\exp(-x)$	$\exp(x)$
250	1	0	1	0	1	1	1	1	0
500	0	0	1	0	0	1	1	0	1
1,000	0	1	0	0	0	0	1	1	0
1,500	0	1	0	0	0	0	0	0	0

selected inliers to compute the parameters for the basis functions. After combining all the three basis functions, Fig. 4d depicts the concave DCT curve for the dose group of 400 mg.

4.1.3 RANSAC for Building ParaLibrary

As discussed in Section 4.1.2, we select $f(x) = \{x^{-2}, \log(x), (1-\exp(-x))\}$ as our basis functions. The final curve in Fig. 4d is built by the curve fitting interpolation using the weighted basis functions. The weights of these functions will be further named as "parameters." To build the parameters' library, or ParaLibrary, we need to extract these parameters from each of the training data samples. The RANSAC algorithm outputs a curve that can be described as follows:

$$g(x) = \alpha \cdot f(x) = [\alpha_1 \;\; \alpha_2 \;\; \alpha_3] \begin{bmatrix} x^{-2} \\ \log(x) \\ 1-e^{-x} \end{bmatrix} \tag{3}$$

where the set of α's compose our ParaLibrary. For each training sample S_i, we first collect all the other samples from the training library with the same dose amount (in the same dose group). Then the RANSAC algorithm is applied M times (Algorithm 1). However, apart from selecting the model by choosing the one with the biggest number of inliers, we also take into consideration that the target sample S_i has to be an inlier and also be as close to the estimated RANSAC curve as possible.

4.2 Prediction Module

The prediction module runs an SVM-based drug concentration prediction using the preprocessed data and predicts the concentration values for a new patient. Instead of solving a convex *quadratic programming problem* (QP) as a theoretical SVM solver, we simply use a *least square SVM* (LS-SVM) classifier to give a solution by solving a set of linear equations [7, 45]. Section 4.2.1 describes the mathematics of the LS-SVM algorithm. There, the simulation results of the influence of the SVM's hyper-parameters are also discussed. Section 4.2.2 modifies the LS-SVM algorithm to predict three parameters $\{\alpha_1, \alpha_2, \alpha_3\}$ for each sample of the testing library.

4.2.1 Least Square Support Vector Machine

The goal of SVM is to find a linear function $f(x) = w \cdot \boldsymbol{\phi}(x) + b$ which approximates the relationship between the training data points and can estimate output y according to new input data. Here, $\boldsymbol{\phi}(x)$ maps the input samples to a higher-dimensional feature space by applying a nonlinear function in the original space; w and b stand for the weights of the feature space and offset, respectively. To apply the LS-SVM algorithm for the drug concentration prediction, we assume that there are N patient samples in the library, some of which can be obtained from the same patient, in the form of $(x_i, y_i) \in \{(x_1, y_1), \cdots, (x_N, y_N)\}$, where y_i denotes the drug concentration values, and x is a vector of d patient features, e.g., age, gender, and body weight [7].

A loss function $\mathcal{L}(y, f(x)) = (y - f(x))^2$ [32] is used to estimate the deviations between the predicted values and the measured ones. To minimize this loss function and meanwhile to prevent overfitting, SVM adopts the following objective function:

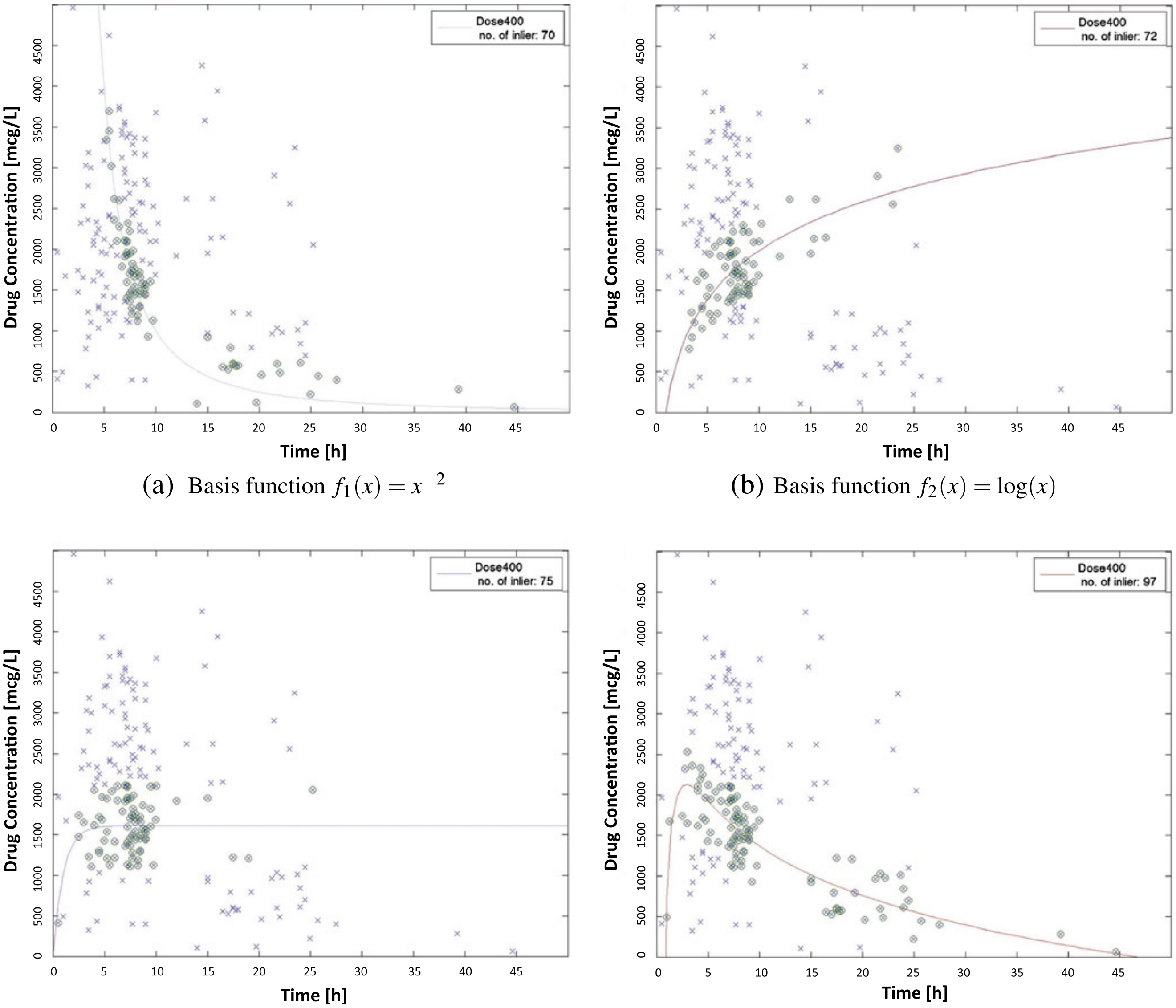
$$\min_{w,b} \frac{1}{2}\|w\|^2 + C_0 \sum_{i=1}^{N} [y_i - w \cdot \boldsymbol{\phi}(x_i) - b]^2 \tag{4}$$

where the constant C_0 determines the tradeoff between overfitting to the function and the amount up to which deviations between the predicted and measured values are tolerated. Note that this objective function has a root-of-sum-of-square fitting error and a regularization term, which is also a standard procedure for the training of *multilayer perceptrons* and is related to ridge regression [46, 47]. Applying the Lagrangian analysis to solve the optimization problem of objective function [45, 47], we see that the optimal w can always be expressed by the following:

$$w = \sum_{i=1}^{N} \alpha_i \boldsymbol{\phi}(x_i) \tag{5}$$

Plugging w into Eq. 4, we can estimate α and b by solving the linear system:

$$\underbrace{\begin{bmatrix} \mathbf{K} + \frac{1}{C_0} I & \mathbf{1} \\ \mathbf{1}^T & 0 \end{bmatrix}}_{H} \begin{bmatrix} \alpha \\ b \end{bmatrix} = \begin{bmatrix} y \\ 0 \end{bmatrix} \tag{6}$$



(a) Basis function $f_1(x) = x^{-2}$

(b) Basis function $f_2(x) = \log(x)$

(c) Basis function $f_3(x) = 1 - e^{-x}$

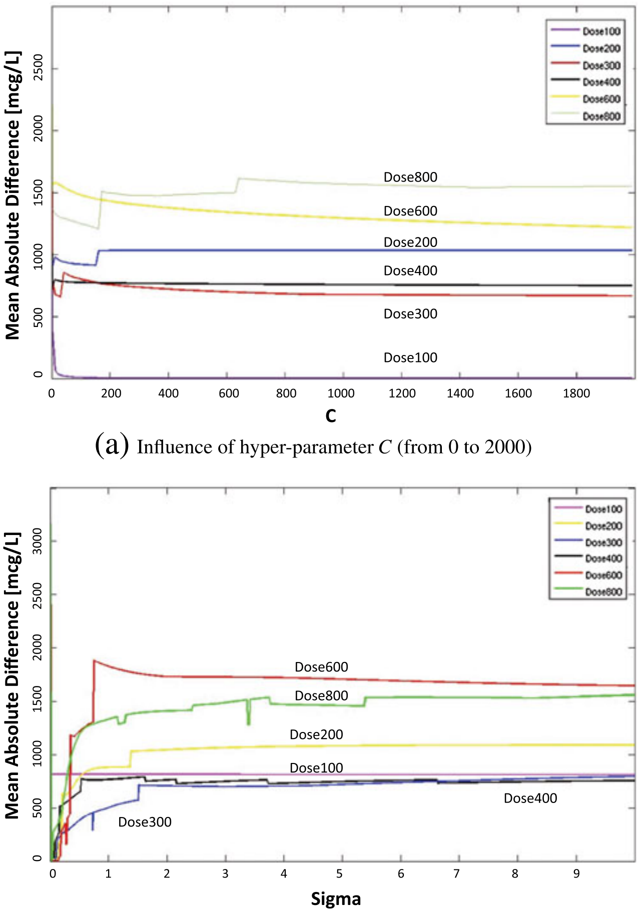
(d) Applying all basis functions in (a), (b) and (c)

Fig. 4 Estimations of different basis functions to the dose group of 400 mg (with the biggest number of data samples). **a** Basis function $f_1(x) = x^{-2}$. **b** Basis function $f_2(x) = \log(x)$. **c** Basis function $f_3(x) = 1 - e^{-x}$. **d** Applying all basis functions in **a**, **b**, and **c**. *Red curve*, estimated RANSAC curve using the given basis function(s); *blue cross point*, data samples with dose = 400 mg; *green circle point*, selected inliers by RANSAC

where $\mathbf{K}$ is the kernel matrix defined by $K_{ab} = \boldsymbol{\phi}(x_a)^T \boldsymbol{\phi}(x_b)$. The use of the kernel matrix greatly helps in reducing the computational complexity without explicitly computing $\boldsymbol{\phi}(x)$, making use of the fact that the SVM algorithm depends only on dot products between sample patterns. Hence, after defining the kernel function, the least square optimization problem can be solved simply by inverting the first term H in the left-hand side of Eq. 6. Once we obtain the value of α and b, the output concentration of the new patient value y can be then estimated through the prediction function: $f(x) = \sum_{i=1}^{N} \alpha_i \mathbf{K}(x_i, x) + b$.

The effectiveness and the accuracy of SVM highly depends on the choice of the kernel function. In our system, we select the Gaussian distribution, a common choice with a single hyper-parameter kernel width $\boldsymbol{\sigma}$ as the kernel function of the SVM algorithm. Hence, Eq. 6 has two parameters to be estimated—C_0 and $\boldsymbol{\sigma}$—the best combination of which is found by a grid search with exponentially growing sequences, e.g., $C_0 \in \{10^{-2}, 10^{-1}, \cdots, 10^2, 10^3\}$, and $\boldsymbol{\sigma} \in \{10^{-3}, \cdots, 10\}$, through cross validation. An L-fold cross validation is a commonly used method to estimate the parameters of a model over each observation value [48]. It randomly partitions the original training sample into L subsamples, each of which is treated as the "validation data" in the training phase, and the remaining $L - 1$ subsamples are used as training data. The cross-validation process is then repeated L times, or folds, to compute the values of C_0 and $\boldsymbol{\sigma}$ with each of the L subsamples used exactly once as the validation data. As different dose groups have different numbers of data points, and after preprocessed by RANSAC, the remaining number of training data points, N_{s_i}, are further reduced. Therefore, we choose $L = \min(10, N_{s_i})$, where s_i denotes the dose group i. We choose C_0 and $\boldsymbol{\sigma}$ to be the one of the L results having the least *mean absolute difference* (MAD) between the predicted values and the validation data.

Figure 5 shows the influence of the hyper-parameters C and $\boldsymbol{\sigma}$ on the prediction accuracy in different dose groups of the training data, when the RANSAC algorithm is not applied. The results show that the MAD values decrease with the growth of the C value within (10, 200) and start to increase after $C > 200$ due to overfitting. On the other hand, smaller kernel width $\boldsymbol{\sigma} < 1$ decreases the MAD values.



(a) Influence of hyper-parameter C (from 0 to 2000)

(b) Influence of hyper-parameter σ (from 0.001 to 10)

Fig. 5 **a** Influence of hyper-parameter C (from 0 to 2000). **b** Influence of hyper-parameter σ (from 0.001 to 10). Influence of hyper-parameters C and σ on the prediction accuracy (mean absolute difference values) with respect to different dose groups

Table 4 shows the accuracy results of the SVM and RANSAC algorithms over different dose groups. The predictions are made according to the following five methods:

- Method 1: training separately each dose group without RANSAC
- Method 2: training the whole library without RANSAC
- Method 3: training separately each dose group with RANSAC on training data
- Method 4: training the whole library with RANSAC on the training data
- Method 5: training separately each dose group with RANSAC on both training and testing data, with different thresholds to each dose group
- PK: the pharmacokinetic method [6]
- $\text{PK}^{\text{RANSAC}}$: the pharmacokinetic method with RANSAC on the testing data

The data in this table indicate that separating the dose group gives better prediction accuracy, especially in the groups where there are very few data samples. It also shows that using RANSAC on the testing database can enhance the prediction accuracy, which indicates an insufficiency in the patient features in current clinical practice.

4.2.2 SVM-based Parametrization (ParaSVM)

Instead of directly computing the drug concentration at a given time, clinicians are more interested in modeling the concentration–time curve, or DCT curve, for each patient in order to visually check whether the concentration drops within the therapeutic range at the trough value. There are mainly two ways to obtain the DCT curve:

- Compute the point-wise concentration values and build the DCT curve by linear interpolation;
- Compute the parameters used to construct the DCT curve (Eq. 3), by curve fitting interpolation.

The first method implicitly models the relationship of patient features and the drug concentrations. It relies on the point-wise prediction for the concentrations over time using the SVM algorithm to take into account new patient features. However, it is difficult to capture a global structure explicitly; especially when new measured concentration data is provided, it does not provide any means to adapt the global curve structure according to this given concentration value. The traditional PK model [6] provides an explicit structural information of the concentration curve. However, it considers a limited number of patient features apart from the fact that some of these features require several blood concentration measurements to be determined accurately,

Table 4 Comparisons of the prediction accuracy over the five methods and the pharmacokinetic (PK) method (MAD values with units in microgram per liter)

Dose group (mg)	100	200	300	400	600	800
No. of training/no. of testing	3 / 1	11 / 1	18 / 7	193 / 176	10 / 13	16t / 10
Method 1	26.25	178.05	1,799.01	983.16	771.41	1,868.66
Method 2	1,024.47	1,318.78	1,785.94	1,033.71	1,274.10	1,493.37
Method 3	26.25	209.95	1,841.65	860.11	660.04	1,011.15
Method 4	935.64	838.54	1,732.18	827.41	461.31	1,059.87
PK	311.4	446.3	1,474.4	819.3	630.1	1,200.4
Method 5 (threshold (μg/L))	0.6 (210)	42.8 (260)	51.6 (20)	90.5 (80)	480.8 (120)	82.8 (110)
$\text{PK}^{\text{RANSAC}}$	311.4	446.3	322.7	493.4	627.7	1,831.1

i.e., the drug absorption/elimination rate. Hence, we combined the "implicit" SVM method and the "explicit" analytical way to model the parameters of the basis functions used to construct the DCT curve [10].

In the case of modeling N patient samples as described in Section 4.2.1, the form of patient samples becomes $(x_i, y_i^1, \cdots, y_i^j, \cdots, y_i^{NP})$, where i is the ID of a sample $i \in \{1, 2 \cdots, N\}$, y_i^j denotes the j-th parameter value of this patient, and NP is the number of parameters which is equal to three in our case. The goal is to find NP linear functions $y^j = w^j \cdot \boldsymbol{\phi}^j(x) + b_j$ to describe the relationship between the dataset points and estimate the parameter value y according to a new input dataset. Similarly, we try to minimize the following modified objective function:

$$\min_{w,b} \frac{1}{2}||w||^2 + C_0 \sum_{j=1}^{NP} \sum_{i=1}^{N} [y_i^j - w^j \cdot \boldsymbol{\phi}^j(x_i) - b^j]^2, \quad (7)$$

the right part of which are the mean squared error functions for all the parameters. Similar to Section 4.2.1, after applying the Lagrangian analysis to solve the optimization problem of objective function, we obtain w as follows:

$$w^j = \sum_{i=1}^{N} \alpha_i^j \boldsymbol{\phi}^j(x_i) \quad (8)$$

Combining Eqs. 7 and 8, we can obtain a linear system:

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

where each entry of the kernel matrix $\mathbf{K}^j$ is defined to be $K_{ab}^j = \boldsymbol{\phi}^j(x_a)^T \boldsymbol{\phi}^j(x_b)$. Therefore, the prediction function becomes $y^j = \sum_{i=1}^{N} \alpha_i \mathbf{K}^j(x_i, x) + b^j$.

4.3 Selection Module

The *selection* module chooses the best dose amount and dose interval according to the given therapeutic ranges (possibly one for the peak and another for the trough concentration). However, in practice, some drugs have only one therapeutic range available, such as in the imatinib case study. For a general purpose, the DADSS system proposes both solutions accordingly.

4.3.1 Computation Rules

As shown in Fig. 2, predictions from both the SVM-based point-wise prediction and the ParaSVM-based DCT curve prediction will be the inputs to the selection module to compute the ideal dose and the time interval. We consider for discrete sets of candidate doses $D_j \in \{100, 200, \cdots, 2{,}000\}$mg and candidate time intervals $\boldsymbol{\tau} \in \{1, 2, \cdots, 24\}$h. The final output of DADSS can be a recommended dose amount D^* and/or the dose interval $\boldsymbol{\tau}^*$.

As indicated before, there could be two therapeutic ranges defined for each drug: peak and trough drug concentration ranges. Our system enables the recommendations based on both. Figure 1a shows an example of selecting a proper curve based on the peak concentration range. The system chooses the curve whose peak concentration response is the closest to the ideal value (PkBD_m), as indicated in the Eq. 10.

$$\underset{D_j}{\text{argmin}}(|C_{j\max} - C_{PkBD_m}|), \quad (10)$$

where $C_{j\max}$ stands for the peak concentration value within 24 h after taking the dose D_j. This indicates the smallest difference between the ideal peak concentration value and the peak values estimated by the *core* module corresponding to each D_j. Thus, curve 3 is picked up in this example. Similarly, Fig. 1b shows the example of selecting a dose interval based on the trough concentration range, and thus, the system computes the T_h and T_l with respect to the intersections between curve 3 and the trough range (TrBD_{up} and TrBD_{low}). Tm, which is the time that corresponds to the ideal trough concentration value, is recommended to be the time of next dose intake.

When only the trough concentration range is available, the system first selects the dose amount whose corresponding concentration value at 24 h is the closest to the ideal trough concentration value, as shown in Eq. 11.

$$\underset{D_j}{\operatorname{argmin}}(|C_{j_{24}} - C_{TrBD_m}|), \tag{11}$$

where $C_{j_{24}}$ stands for the concentration values estimated at 24 h after giving a dose D_j. Then it computes the dose interval $\boldsymbol{\tau}^* = (\boldsymbol{\tau}_h, \boldsymbol{\tau}_l)$ of this curve. Since we want to keep the trough drug concentration value within the trough therapeutic range, $\boldsymbol{\tau}_h$ and $\boldsymbol{\tau}_l$ are computed according to the higher and lower bounds of the trough therapeutic range, respectively.

4.3.2 Case Study of Imatinib

Table 5 describes some examples of how the decisions about imatinib's dose and intake interval are made for five randomly selected patients. For each pair of the cross product of the dose D_i and the dose interval $\boldsymbol{\tau}_i$ for a new patient, the core module computes the corresponding drug concentration value. The selection module first removes the candidate doses whose predicted resulting drug concentration at time 24 h ($C_{i_{24}}$) are higher than the upper bound of the trough therapeutic range as well as the ones whose predicted peak concentration value is lower than the trough lower bound. Furthermore, to choose the best dose, our system computes the absolute difference between each $C_{i_{24}}$ value and the ideal value of the trough therapeutic range, and selects the dose with respect to the smallest difference, as shown in Eq. 11. For example, for patient 1, we obtained the set of $C_{i_{24}} = [890.6, 1{,}032.5, 1{,}152.5, 1{,}239.1]$ µg/L that corresponds to the set of $D_i = [200, 400, 600, 800]$ mg. The ideal value is 1,000 µg/L, therefore $C_{i_{24}} = 1{,}032.5$ µg/L has the smallest difference, and thus the curve whose $D^* = 400$ mg is chosen. Hereafter, the system obtains the range of the dose interval ($\boldsymbol{\tau}^* = (\boldsymbol{\tau}_h, \boldsymbol{\tau}_l)$) according to the lower and upper bounds of the trough therapeutic range. However, in the real clinical practice, doctors tend to give a common dose amount (400 mg) and time interval (24 h) to every adult patient at the a priori stage (at the start of treating a new patient). Following the clinical protocol in the later stages, doctors update the dose amount and the time interval with respect to patients' responses to the treatment [26].

Table 5 5 Sample recommendations from DADSS

	Patient profile features			Recommendations		
No.	Gender	Age	Body weight (kg)	D^* (mg)	$\boldsymbol{\tau}_h(h)$	$\boldsymbol{\tau}_l$ (h)
1	M	82	56	400	13	24
2	F	58	53	500	15	24
3	F	62	54	700	16	24
4	M	58	100	800	18	24
5	M	47	73	500	14	24

4.4 Adaptation Module

In the simplest clinical routine, once a new patient is admitted, a first dose is determined based on a population value estimated from the library data. This method does not guarantee that the new patient and the library datasets have similar conditions, SVM-based first dose estimation uses these patient's features and then predicts the drug concentrations at a specific time or the concentration curves. This is called the a priori adaptation. We show a way to build the DCT curve for multiple doses after the initial a priori adaptation taking into consideration the residual drug concentrations in patient's blood (see Section 4.4.1). To further refine the predicted DCT curve to be as close to the real measurement as possible, a blood test is sometimes taken to control the drug concentration to be within the therapeutic range (Section 4.4.2). This is known as the a posteriori adaptation. As presented in the previous sections, insufficient types of patient features and measurement errors are two possible factors of current predicted concentration values being sometimes largely deviated from the measured values. Here, we consider updating the predicted DCT curve with a given measured concentration value to be efficient to personalize the prediction results, as the feedback adaptation loop in Fig. 2.

4.4.1 Multidose Estimation

Knowing how the concentration value varies with time after multiple doses is important to clinicians and patients in order to monitor a long-term therapeutic procedure. In the a priori adaptation, the multi-dose DCT curve can be obtained simply by recomputing over the days of the updated one-dose concentration curve, taking into account the residual concentration value of the previous day, since the drug sometimes takes several days, i.e., 5 to 7 days, to be cleared out from the human body.

Figure 6 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, while the first period of the DCT curve is assumed to be the starting day. 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. 6, the residual concentration affects both the peak and the trough concentration values in the beginning of the treatment and starts to be steady after several days (3 days in this example).

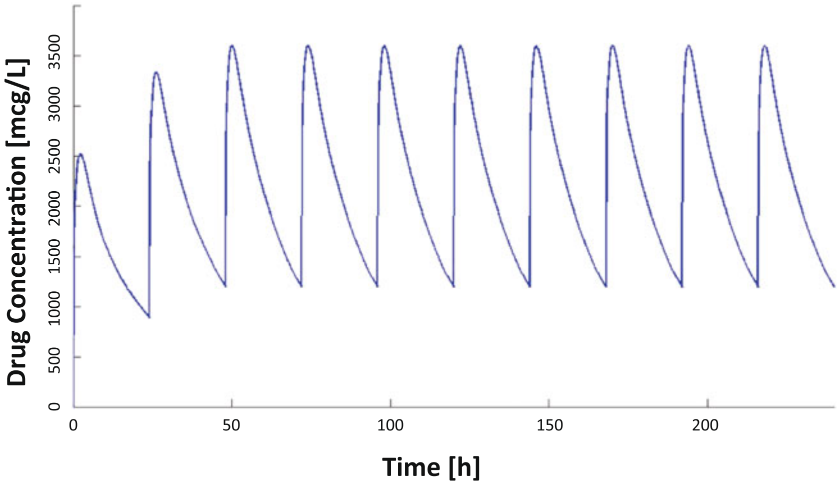


Fig. 6 Example of multiple dose estimation for the drug concentration to time curve over 10 days of the drug imatinib. X-axis, time (in hour); Y-axis, concentration value (in microgram per liter)

4.4.2 The A Posteriori Adaptation

In the 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 point under certain constraints. This procedure is important in clinical routine to overcome an inaccuracy caused by insufficient feature data collection for concentration prediction. 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 a new measurement is provided, we first predict the basis function's parameters using ParaSVM and then search within a certain radius ΔD around each parameter value with a step δd 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:

$$\frac{\partial f_{\text{concentration}}}{\partial t}|_{t=0} > 0; \tag{12}$$

- After several hours, the concentration value reaches the peak value and starts to decrease:

$$\frac{\partial f_{\text{concentration}}}{\partial t}|_{t=T_p} < 0, \tag{13}$$

where T_p is a time point after the peak value, i.e., we set it as $T_p = 24$ h.

- 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 ($t = 24$ h, since imatinib is usually administrated once a day), should be within a certain range (R), i.e., < 50 µg/L:

$$|f^{t=0}_{\text{concentration}} - f^{t=24}_{\text{concentration}}| < \mathbf{R}, \tag{14}$$

- The DCT curve whose shape is the closest to the curve previously predicted by ParaSVM will be chosen:

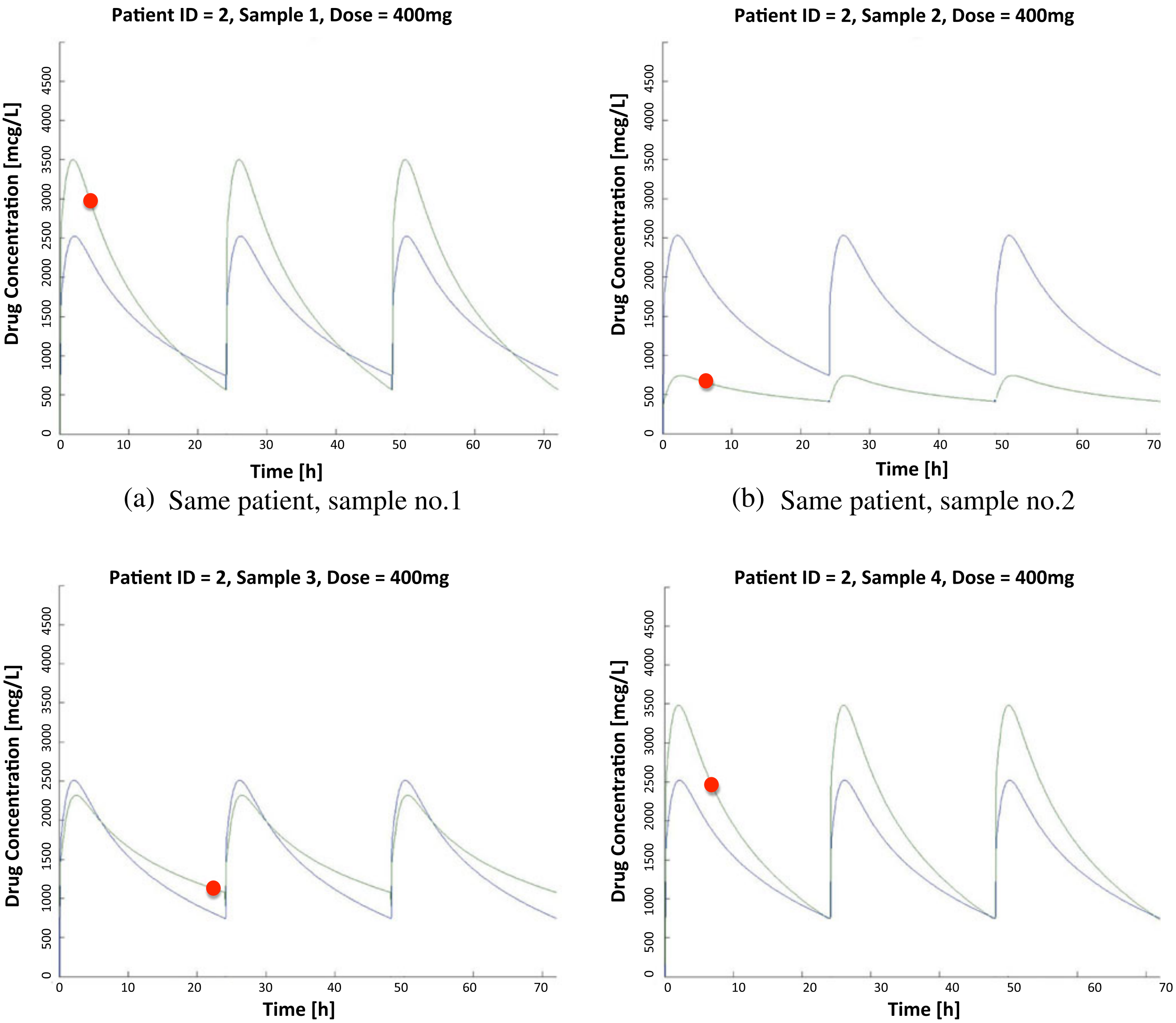
$$\min_{g_r} \sum_{j=0,\cdots,N_s} (g_r^{t=j} - g^{t=j})^2, \tag{15}$$

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.

Figure 7 shows the a posteriori adaptation for a sample patient with the same dose amount $D = 400$ mg, but in different dose periods (we assume that the patient has reached the steady state). As the measured concentration values vary a lot even though the measuring time is relatively similar between 2 days, it indicates that a potential intra-variation has happened to this patient. Based on the above adaptation rules, the DCT curves are adjusted accordingly.

5 Conclusion

In this paper, we have presented an advanced DADSS system with a feedback loop using the ParaSVM method. The system has also been tested and evaluated for different dose groups to overcome the drawbacks caused by the unbalanced sample distribution, which has improved the prediction accuracy in each dose group. Detailed information of choosing the basis functions, RANSAC threshold, and hyper-parameters has been introduced. The results show that a general setting of $C = 1{,}000$ and $\sigma < 1$ can be applied to all the dose groups. The difference between the a priori case analysis and a steady-state analysis is discussed.



(a) Same patient, sample no.1

(b) Same patient, sample no.2

(c) Same patient, sample no.3

(d) Same patient, sample no.4

Fig. 7 **a** Same patient, sample no. 1. **b** Same patient, sample no. 2. **c** Same patient, sample no. 3. **d** Same patient, sample no. 4. Examples of parametrically refined DCT curves over 3 days in a *steady* state of a same patient being sampled in different days with a same dose amount 400 mg. *Blue curve*, DCT curve computed by SVM-based parametric approach; *green curve* (passing through a *red dot*), DCT curve adjusted with one measurement; *red dot*, measured concentrations

As to the former, a residual concentration from the previous dose period should be taken into consideration when we construct the drug concentration to time curve for multiple doses. As for the latter, the algorithm considers the residual drug concentration influence and gives the prediction as in a steady-state case. Though the ParaSVM algorithm does not improve the drug concentration's prediction accuracy, it provides the ability of being adjustable when a new measured concentration value is available. The system is also able to deal with a very large number of features which are currently not presented in the training database. Therefore, we encourage the extension of patient feature datasets.

Acknowledgments The authors would like to thank Carlotta Guiducci from EPFL for the help in manuscript revision and T. Buclin and V. Gotta from CHUV Hospital of Lausanne for the precious suggestions on clinical data modeling and provision with sufficient data. 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 NanoTera.ch initiative, evaluated by the Swiss National Science Foundation.

References

1. Chiu, D.Y., Chang, C.C., Evens, M.W., Chern, J.C., Hier, D.B., Trace, D.A., Naeymi-Rad, F. (1994). A complete, hypermedia medical decision analysis support system. In *Proceedings of IEEE seventh symposium on computer-based medical systems* (pp. 16–21).
2. Keen, P.G.W., & Morton, M.S.S. (1978). *Decision support systems: an organizational perspective*. Reading: Addison-Wesley.
3. Kong, G., Xu, D.-L., Yang, J.-B. (2008). Clinical decision support systems: a review on knowledge representation and inference under uncertainties. *International Journal of Computational Intelligence Systems*, *1*(2), 159–167.
4. Brown, R.F. (1980). Compartmental system analysis: state of the art. *IEEE Transactions on Biomedical Engineering*, *27*(1), 1–11.
5. Blau, G., & Orcun, S. (2009). A Bayesian pharmacometric approach for personalized medicine—a proof of concept study with simulation data. In *Proceedings of the 2009 winter simulation conference* (pp. 1969–1976).
6. Widmer, N., Decosterd, L., Csajka, C., Leyvraz, S., Duchosal, M.A., Rosselet, A., Rochat, B., Eap, C.B., Henry, H., Biollaz, J., Buclin, T. (2006). Population pharmacokinetics of imatinib and the role of α_1-acid glycoprotein. *British Journal of Clinical Pharmacology*, *62*(1), 97–112.
7. You, W., Widmer, N., De Micheli, G. (2011). Example-based support vector machine for drug concentration analysis. In *Engineering in medicine and biology society (EMBC), 2011 annual international conference of the IEEE* (pp. 153–157).
8. You, W., Simalatsar, A., Widmer, N., De Micheli, G. (2012). A drug administration decision support system. In *Bioinformatics and biomedicine workshops, 2012 IEEE international conference, October* (pp. 122–129).
9. You, W., Simalatsar, A., De Micheli, G. (2012). RANSAC-based enhancement in drug concentration prediction using support vector machine. In *Proceedings of the international workshop on innovative simulation for healthcare (IWISH)* (pp. 21–25).
10. You, W., Simalatsar, A., De Micheli, G. (2013). Parameterized svm for personalized drug concentration prediction. In *Engineering in medicine and biology society (EMBC), 2013 annual international conference of the IEEE*.
11. Fitzmaurice, D.A., Hobbs, F.D., Murray, E.T., Bradley, C.P., Holder, R. (1996). Evaluation of computerized decision support for oral anticoagulation management based in primary care. *The British Journal of General Practice*, *46*(410), 533–535.
12. Fihn, S., McDonell, M., Vermes, D., Henikoff, J., Martin, D., Callahan, C., Kent, D., White, R. (1994). A computerized intervention to improve timing of outpatient follow-up: a multicenter randomized trial in patients treated with warfarin. National consortium of anticoagulation clinics. *Journal of General Internal Medicine*, *9*(3), 131–139.
13. Pller, L., Wright, D., Rowlands, M. (1993). Prospective comparative study of computer programs used for management of warfarin. *Journal Clinical Pathology*, *46*(4), 299–303.
14. Ryff-de, L.A., Engler, H., Ntzi, E., Berger, M., Berger, W. (1992). Clinical application of two computerized diabetes management systems: comparison with the log-book method. *Diabetes Research*, *19*(3), 97–105.
15. Hickling, K., Begg, E., Moore, M. (1989). A prospective randomised trial comparing individualised pharmacokinetic dosage prediction for aminoglycosides with prediction based on estimated creatinine clearance in critically ill patients. *Intensive Care Medicine*, *15*(4), 233–7.
16. Burton, M., Ash, C., Hill, D.J., Handy, T., Shepherd, M., Vasko, M. (1991). A controlled trial of the cost benefit of computerized Bayesian aminoglycoside administration. *Clinical Pharmacology and Therapeutics*, *49*(6), 685–94.
17. Sutton, D.R., Taylor, P., Earle, K. (2006). Evaluation of PROforma as a language for implementing medical guidelines in a practical context. *BMC Medical Informatics and Decision Making*, *6*, 20+.
18. Johnson, P., Tu, S., Booth, N., Sugden, B., Purves, I. (2000). Using scenarios in chronic disease management guidelines for primary care. In *Proceedings of American medical informatics association (AMIA) symposium* (pp. 389–393). PubMed Central PMCID.
19. Tu, S.W., & Musen, M.A. (2001). Modeling data and knowledge in the EON guideline architecture. *Medinfo*, *10*(1), 280–284.
20. Boxwala, A., Peleg, M., Tu, S., Ogunyemi, O., Zeng, Q., Wang, D., Patel, V., Greenes, R., Shortliffe, E. (2004). GLIF3: a representation format for sharable computer-interpretable clinical practice guidelines. *Journal of Biomedical Informatics*, *37*, 147–161.
21. Tu, S.W., Campbell, J.R., Glasgow, J., Nyman, M.A., McClure, R., McClay, J., Parker, C., Hrabak, K.M., Berg, D., Weida, T., Mansfield, J.G., Musen, M.A., Abarbanel, R.M. (2007). The SAGE guideline model: achievements and overview. *Journal of the American Medical Informatics Association*, *14*, 589–598.
22. Ciccarese, P., Caff, I.E., Boiocchi, L., Quaglini, S., Stefanelli, M. (2004). A guideline management system. *Studies in Health Technology and Informatics*, *107*(1), 28–32.
23. Terenziani, P., Montani, S., Bottrighi, A., Torchio, M., Molino, G., Correndo, G. (2004). The GLARE approach to clinical guidelines: main features. *Studies in Health Technology and Informatics*, *101*, 162–166. doi:10.3233/978-1-60750-944-8-162.
24. Duftschmid, G., Miksch, S., Shahar, Y., Johnson, P. (1998). Multi-level verification of clinical protocols. In *Proceedings of the workshop on validation verification of knowledge-based systems* (pp. 1–4).
25. Miksch, S., Shahar, Y., Johnson, P. (1997). Asbru: a task-specific, intention-based, and time-oriented language for representing skeletal plans. In *UK, open university* (pp. 1–9).
26. Simalatsar, A., & De Micheli, G. (2012). Tat-based formal representation of medical guidelines: imatinib case-study. In *Proceedings of engineering in medicine and biology society (EMBC)*. San Diego.
27. Alpaydin, E. (2004). *Introduction to machine learning (adaptive computation and machine learning)*. Cambridge: MIT Press.
28. Ni, K.S., & Nguyen, T.Q. (2007). Image superresolution using support vector regress. *IEEE Transaction on Image Processing*, *16*, 1596–1610.
29. Zhu, G., Liang, D., Liu, Y., Huang, Q., Gao, W. (2005). Improving particle filter with support vector regression for efficient visual tracking. In *IEEE ICIP 2005* (pp. 422–425).
30. Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J. (1984). *Classification and regression trees*. Monterey: Wadsworth and Brooks/Cole Advanced Books and Software.
31. Jeatrakul, P., & Wong, K.W. (2009). Comparing the performance of different neural networks for binary classification problems. In *IEEE SNLP 2009* (pp. 111–115).
32. Gunn, S.R. (1998). *Support vector machines for classification and regression* (pp. 29–43). Technical Report, University of Southampton.
33. Freund, Y., & Schapire, R.E. (1999). A short introduction to boosting. *Journal of Japanese Society for AI*, *5*(14), 771–780.
34. Auria, L., & Moro, R.A. (2008). Support vector machines (SVM) as a technique for solvency analysis. In *DIW Berlin*. Discussion papers.
35. Cortes, C., & Vapnik, V.N. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297.
36. Gudadhe, M., Wankhade, K., Dongre, S. (2010). Decision support system for heart disease based on support vector machine and

artificial neural network. In *Computer and communication technology 2010* (pp. 741–745).

37. Guo, L., Wu, Y., Yan, W., Shen, X., Li, Y. (2005). Research on medical diagnosis decision support system for acid-base disturbance based on support vector machine. In *Engineering in medicine and biology society 2005 (IEEE-EMBS)* (pp. 2413–2416).
38. Wang, X., Shu, P., Rong, X., Nie, L. (2009). A decision support system based on support vector machine for hard landing of civil aircraft. In *Computer science-technology and applications 2009* (Vol. 2, pp. 66–70).
39. Cheng, X., Wei, Y., Geng, X. (2009). A support vector machines security assessment method based on group decision-marking for electric power information system. In *Information assurance and security 2009* (Vol. 2, pp. 536–539).
40. Basak, D., Pal, S., Patranabis, D.C. (2007). Support vector regression. *Neural Information Processing*, *11*(10), 203–224.
41. Buclin, T., Gotta, V., Fuchs, A., Widmer, N., Aronson, J. (2012). An agenda for UK clinical pharmacology monitoring drug therapy. *British Journal of Clinical Pharmacology*, *73*(6), 917–23.
42. Bourne, D.W.A. (1995). *Mathematical modeling of pharmacokinetic data*, 2nd edn. Lancaster: Technomic.
43. Gotta, V., Widmer, N., Decosterd, L., Csajka, C., Duchosal, M., Chalandon, Y., Heim, D., Gregor, M., Buclin, T. (2010). Therapeutic drug monitoring (TDM) of imatinib: effectiveness of Bayesian dose adjustment for CML patients enrolled in the imatinib concentration monitoring (I-COME) study. In *Swiss medical forum* (p. 285).
44. Fischler, M., & Bolles, R. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, *24*(6), 381–395.
45. Suykens, J., Van Gestel, T., De Brabanter, J., De Moor, B., Vandewalle, J. (2002). *Least square support vector machine*. Singapore: World Scientific.
46. Golub, G., & Van Loan, C.F. (1989). *Matrix computations*. Baltimore: John Hopkins University Press.
47. Brabanter, J.D. (2004). *LS-SVM regression modelling and its applications* (pp. 25–28). PhD thesis.
48. McLachlan, G.J., Do, K.A., Ambroise, C. (2004). *Analyzing microarray gene expression data*, 2nd edn. Hoboken: Wiley.