

Personalized Modeling for Drug Concentration Prediction Using Support Vector Machine

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—Building a personalized model to describe the drug concentration inside the human body for each patient is highly important to the clinical practice and demanding to the modeling tools. Instead of using traditional explicit methods, in this paper we propose a machine learning approach to describe the relation between the drug concentration and patients' features. Machine learning has been largely applied to analyze data in various domains, but it is still new to personalized medicine, especially dose individualization. We focus mainly on the prediction of the drug concentrations as well as the analysis of different features' influence. Models are built based on Support Vector Machine and the prediction results are compared with the traditional analytical models.

I. INTRODUCTION

Current clinical pharmacology practice relies mostly upon the basis of clinical monitoring of patients to choose the adequate dosage of medicines. In such clinical cases, the dosage of medicines are decided based on doctors' experiences or on symptoms scales, which are not suitable to all kinds of drug. 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 the mission is to learn a therapeutic range of a drug to support the clinical monitoring properly. This domain is known as *population pharmacokinetics* (PK).

Hence it is critical to provide a methodology that can aid clinicians to decide the drug dosage amount in a fast, less costly, less invasive way and that is dedicated to personalized modeling for each patient. In literature, there are two main approaches applied to data analysis in the field of pharmacokinetics [1]. One of them uses mathematically understandable techniques to compute an overview of a drug disposition, usually by getting a simple AUC (Area Under Curve of plasma concentration time). This method is usually applied to analysis when few patients data are available. However, in reality, such AUC could not be accurate if the data cannot support a modeling approach. The other kind of method extracts pharmacokinetic information from a sample of the populations by using highly-sophisticated techniques, in which some model structures may be ignored due to insufficient data, e.g. the absorption process after an oral dosage. But most of the applications are for a mass evaluation instead of individualization.

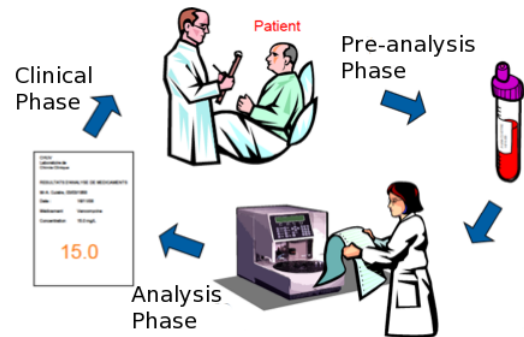


Fig. 1. The Procedure of Therapeutic Drug Monitoring

Structural modeling approach is widely used in many well-designed studies [2] and new statistical approaches have also been introduced to pharmacokinetic analysis, such as Bayesian approaches [3]. However those methods mostly process only variables with real values, while binary-valued variables, such as gender, create strong discontinuities in the models and are in general not supported by the methods. Moreover, due to the explicit analytical model, it is often difficult to add or remove a parameter. Thus in our approach, we emphasize on building the personalized model of drug concentration for each patient by applying Support Vector Machine technique. This paper differs significantly from our previous contribution [4] in the scope and type of result obtained with our work. Here we mainly explore the potentiality of applying Support Vector Machine to personalized modeling of the drug concentration, in order to give an easy visual support for our clinicians.

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 and applying SVM to predict drug concentration. Section 4 shows the analyzations and comparisons of the influences after using different features and parameters. We also demonstrate the quality of our algorithm experimentally in the second part of Section 4 while Section 5 concludes this study.

II. OVERVIEW OF MACHINE LEARNING TECHNIQUES

The main purpose of Machine Learning, a field evolved from the broad field of Artificial Intelligence, is to mimic intelligent abilities of humans by machines. It has been applied in

the domain of classification problems in computer vision and pattern recognition in the past few decades [5]. It also enables the computer to evolve the responses based on the input data, with a certain capability of noise handling schemes. However, till now very little literature could be found on the utilization of machine learning to achieve the dose individualization of personalized medicine.

Out of all the machine learning techniques, there are four important and representative ones, *Decision Trees* (DT), *Neural Networks* (NN), *Boosting* or *Adaptive Boosting* and *Support Vector Machine* (SVM). After being extended to solving the regression problems, these techniques became popular in various other domains such as image superresolution [6], object tracking [7], etc.. Among the four, Decision Trees is the simplest in understanding and thus the fastest approach, but it does not predict as precise as the other three, especially for regression where we need to give a prediction on continuous numbers [8]. Neural Networks is the oldest technique of the four and is inspired by neurobiological knowledge, but it is often regarded as a black box due to its model's high complexity [9]. Boosting Algorithm is a meta-algorithm used in conjunction with weak classifiers, and is sensitive to outliers or noisy data [10]. Support Vector Machine utilizes a hyperplane in a high- or infinite-dimensional space to do classification or regression[11]. It is convenient due to both its clear mathematic understanding and its control of the overfitting problem.

In this paper we use the SVM technique to model our system because of its appropriate complexity, efficiency and strength in data regularization [12]. We focus on analyze the influence of the clinical features to the drug concentrations. The experimental results show that drug concentration predicted by SVM-based approach is similar to the measured values as the results predicted by the PK model, depending on the input features that the algorithm used. Furthermore it is able to build a personalized model for each patient.

III. METHODOLOGY

As has been concluded in our previous work [4], the current practice in the clinical settings is mostly applying General Pharmacokinetic Model or Population Pharmacokinetic Model. In this session, we will briefly describe these two models and give the introduction to the model which is based on Least Square Support Vector Machine.

A. General Pharmacokinetic Model

Drugs that bind with a high affinity to the clinical PK features tend to behave differently due to its sensitivity of a slight variation in the modeling. In the simplest situation, the drug concentration after a single *intravenous* (IV) bolus dose is considered. This procedure has an immediate and concentrated drug effect. In a 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).

$$\frac{dC}{dt} = -k_{el} \cdot C \quad (1)$$

where $C = \frac{\text{dose}}{V} \cdot e^{-k_{el} \cdot t}$, V is the volume of distribution and k_{el} is the elimination rate of the drug inside a body. [1]

In the case of oral dosage, one more component is considered to be i.e. the mechanism of absorption from the *gastrointestinal* (GI) tract to the arteriovenous system. In this way, the concentration of the drug for a single dose is calculated by (2):

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

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

B. Population Pharmacokinetic Model

In the real clinical scenarios, clinicians are more interested by the multiple dose regimens. After m dosages, the general equation for the drug concentration could be written as:

$$C = \frac{F \cdot \text{dose} \cdot k_a}{V \cdot (k_a - k_{el})} \cdot \left\{ \left[\frac{1 - e^{-m \cdot k_{el} \cdot \tau}}{1 - e^{-k_{el} \cdot \tau}} \right] \cdot e^{-k_{el} \cdot t} - \left[\frac{1 - e^{-m \cdot k_a \cdot \tau}}{1 - e^{-k_a \cdot \tau}} \right] \cdot e^{-k_a \cdot t} \right\} \quad (3)$$

where τ stands for the dosage interval and t for the time when the drug concentration is measured. [1]

In the practical clinical settings, even though a pharmacokinetic model may require many parameters, only one or two data points may be available. In such cases, we perform a nonlinear regression through the whole samples of the available patient population, using approach such as NONMEM [13]. This approach could be compared to a Bayesian analysis performed to obtain estimations of the patient pharmacokinetic parameter values by minimizing:

$$\sum_{i=1}^{N_1} \frac{(C_{\text{obs}_i} - C_{\text{calc}_i})^2}{\text{variance}_i} + \sum_{j=1}^{N_2} \frac{(P_{\text{pop}_j} - P_{\text{calc}_j})^2}{\text{variance}_j} \quad (4)$$

where C is drug concentration, P is used to indicate some other parameters, N_1 and N_2 are the number of data points and parameters respectively[14].

C. Least Square Support Vector Machine (LS-SVM)

Support Vector Machine has been shown to be very powerful at separating highly intertwined data. In SVM Classification, a mapping is from the input data X onto a binary output $y = \pm 1$, while SVM Regression extends to allow a mapping f to a real value output which suits the case in our study:

$$\begin{aligned} f : X &\rightarrow \mathbb{R} \\ x &\rightarrow y = f(x) \end{aligned} \quad (5)$$

Least Square SVM (LS-SVM) is selected here, because 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]. Processing the same

patient data as in [4], we assume N patients, each having been examined n_i times. Every sample takes into account d features. Hence the training data are given as:

$$\{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$$

where $N = \sum_{i=1}^N n_i$, y is the drug concentration and x the space of input patterns \mathbb{R}^d . The target is to find a linear function $f(x) = w \cdot \phi(x) + b$, which approximates the relationship between the points in the data set and can be used to estimate the output y with respect to a new input patient data. Here, we use a quadratic loss function $\mathcal{L}(y, f(x)) = (y - f(x))^2$ to describe the training errors.

In order to find the optimal solution, the loss function has to be minimized. In the same time, overfitting problem is also taken into consideration by keeping the norm $\|w\|^2$ small. Thus the regression problem can be rewritten as optimizing the following objective function:

$$\min_{w,b} \frac{1}{2} \|w\|^2 + C_0 \sum_{i=1}^N [y_i - w \cdot \phi(x_i) - b]^2 \quad (6)$$

where the constant C_0 determines the tradeoff between the overfitting and the amount of the training error. It is always a question to decide which C_0 best balances the two factors, so here we simply adopt a cross-validation method. The optimal w can be decided by $w = \sum_{i=1}^N \alpha_i \phi(x_i)$, where α and b are found by solving the linear system:

$$\begin{bmatrix} \mathbf{K} + \frac{1}{C_0} I & \mathbf{1} \\ \mathbf{1}^T & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ b \end{bmatrix} = \begin{bmatrix} y \\ 0 \end{bmatrix} \quad (7)$$

in which $K_{ij} = \phi(x_i)^T \phi(x_j)$ is called kernel matrix. [15] The use of the kernel matrix greatly helps reduce 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 [16]. Hence, after defining the kernel function, the least-square optimization problem could be solved simply by inverting the first term in the left-hand side of (7). And the output could be estimated via the following prediction function:

$$f(x) = \sum_{i=1}^N \alpha_i \mathbf{K}(x_i, x) + b \quad (8)$$

In the comparison with traditional PK model, we will see how each feature affects the the prediction results. Furthermore, in the personalized modeling section, we choose a random sample from a random patient in the testing set, and the model is built via generated features, such as dosage amount, age, etc..

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 [17]. The training data set consists of 54 patients and 252 samples (obtained at time of a previous clinical trial [17]). In Subsection A, the test patient data is randomly selected from

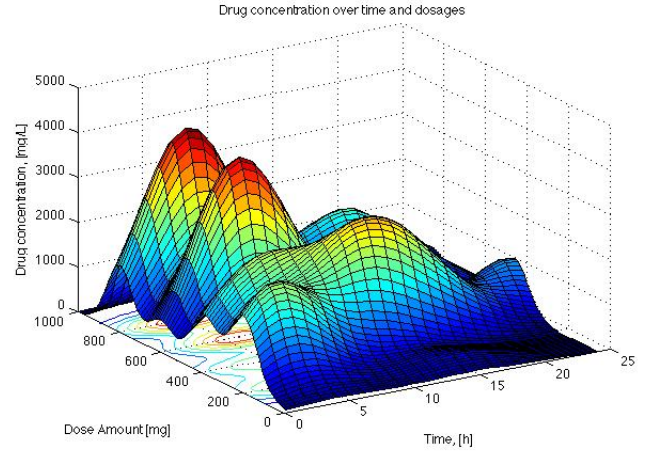


Fig. 2. Drug Concentration Modeling on One Sample Patient over Time and Dosages, x -axis: Time [h], y -axis: Dosage [mg], z -axis: Drug Concentration[mcg/L]

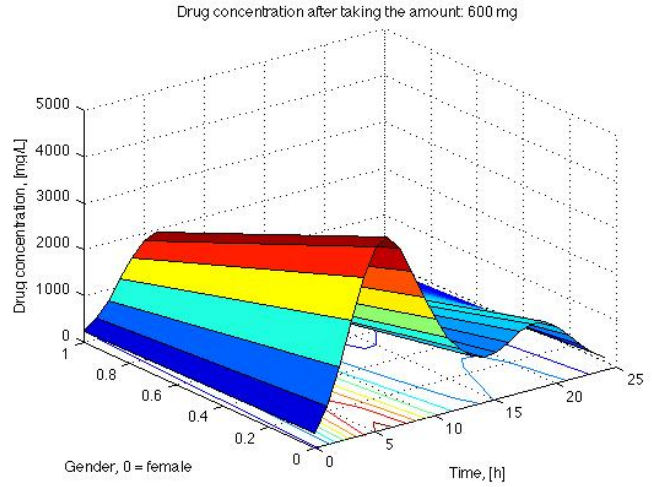


Fig. 3. Drug Concentration Modeling over Time and Gender after Taking 600mg Oral Dosage (0 indicates female), x -axis: Time [h], y -axis: Gender, z -axis: Drug Concentration[mcg/L]

the testing sets and generated with interpolations, while the whole testing library is used in Subsection B including 65 patients and 209 samples (patients followed latter on a routine basis, in the context of an amendment of the initial trial), and we compare the prediction results of LS-SVM system to the ones using a general population pharmacokinetic model.

A. Personalized Modeling

We consider 5 personalized parameters here: {Time, Dosage Amount, Gender, Age, Body Weight}. ‘Time’ parameter denotes the measuring time after the previous dosage has been taken, which is usually less than 24 hours, and ‘Dosage Amount’ is defined to be a single oral dose varying from 0 to 1000mg. ‘Gender’, ‘Age’ and ‘Body Weight’ are personalized parameters and will be used to analyze the drug concentration over time given the other parameters remain unchanged.

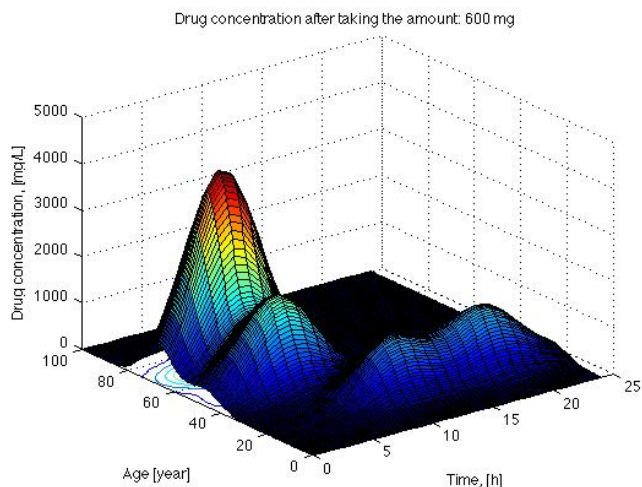


Fig. 4. Drug Concentration Modeling over Time and Ages from 0 to 100 after Taking 600mg Oral Dosage, x -axis: Time [h], y -axis: Age [year], z -axis: Drug Concentration[mcg/L]

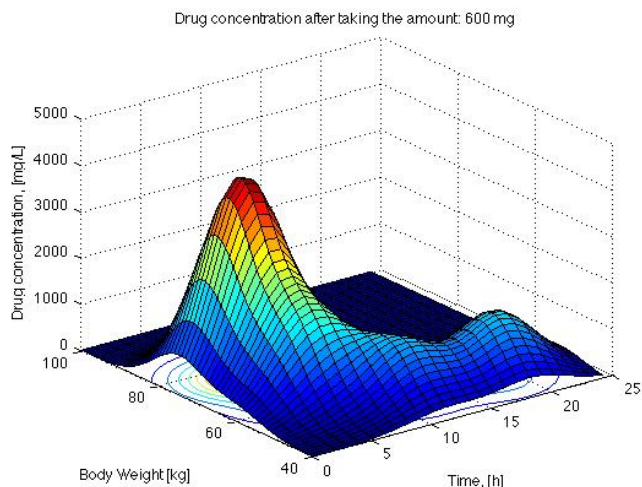


Fig. 5. Drug Concentration Modeling over Time and Body Weight from 40kg to 100kg after Taking 600mg Oral Dosage, x -axis: Time [h], y -axis: Body Weight [kg], z -axis: Drug Concentration[mcg/L]

From Figure 2 to Figure 5, the z -axis is always the predicted drug concentration values. Figure 2 plots the drug concentration predictions over different time and dosages for an individual patient. From this figure we could see that the larger the drug amount is, the earlier the concentration will reach a peak and the longer it takes the concentration to drop. However, after a certain amount, i.e. 600mg in this patient, the arrival time of the concentration peak does not vary too much, but the peak concentration will reach the global maximum. This figure aims to process the case of a single oral dosage. But we could see some of the curve along the time axis has more than one extreme value, which is because of a short interval in between two oral dosages in the training data.

Figure 3 shows the drug concentrations' being affected by

the gender information. The dosage is chosen to be 600mg based on the observation of Figure 2. On 'Gender' axis, 0 indicates female patient and 1 male patient, while all the other parameters are unchanged. Given a same dosage, a female patient can get a higher drug concentration value earlier than a male patient, but the value also drops faster than a male.

Figure 4 describes how the age of one patient influences the drug concentration after taking a 600mg. Although the age varies from 0 to 100, but the training set is lack of the data less than 20 or more than 80 years old. But with 20 to 80 years old, we could observe 2 maximum values in the concentration, one around 65 and the other around 30.

Figure 5 reveals the relationship between the drug concentrations and the values of body weight in one patient. The maximum concentration value is reached when the patient is about 70kg. A double-peak is observed at the body weight being around 50kg, 15 hours after the single dosage. This is because the measurement of the training data has been taken within a short interval between two dosages.

Hence, the LS-SVM method could build a full model for each individual patient so as to aid the clinicians to monitor their patients in a fast, less invasive, less costly way. However, the quality of the training data set is critical in the modeling.

B. Comparisons of Drug Concentration Predictions

As shown in equations (6) and (8), we need to determine several parameters both for LS-SVM method as follows:

- 1) K : the kernel function decides the effectiveness of the SVM method. Here we choose to use the Gaussian Kernel which has a single parameter σ that has to be estimated.
- 2) C_0 : after we select the kernel, the choice of the kernel parameters and the margin factor C_0 determines the performance of SVM. The best combination of C_0 is found via 10-fold cross validation. In our experiments, we use $C_0 \in \{10^{-2}, 10^{-1}, \dots, 10^3, 10^4\}$ and we randomly separate the original training data into 10 subsample groups.
- 3) σ : we need to decide the parameter σ in our Gaussian kernel. In the personalized modeling part, we set the σ to be the mean pairwise distance of the features, while in the comparison part, σ is set similarly as C_0 using 10-fold cross validation. The value of σ is also highly related to the shape of the curve in the Figure 2 to 5.

One disadvantage of using traditional PK model is that it is often unable to consider binary-valued inputs. Neither could it be modified easily to address the importance of each feature. The Support Vector Machine approach could deal with such problems and provides a similar prediction result. As analyzed in our previous work [4], in Table I, the mean absolute difference has been compared between the measured drug concentration and the predicted ones using LS-SVM. Features 'dosage' (Feature A) and 'measuring time after one dose' (Feature B) are thought to be the key features to the results and hence we do not exclude them in any prediction. Using the PK model, the mean value of the differences between the predicted

TABLE I
MEAN ABSOLUTE DIFFERENCES BETWEEN DRUG CONCENTRATION
PREDICTED AND MEASURED BASED ON LS-SVM (FEATURES:
A-DOSAGE, B-MEASURING TIME, C-GENDER, D-AGE, E-BODY
WEIGHT) [UNIT: MCG / L]

M%	100%	70%	50%	30%	10%
A+B	822.3	850.4	858.2	857.0	971.9
A+B+C	835.3	848.0	893.5	856.7	982.1
A+B+C+D	846.5	867.1	834.6	951.4	980.6
A+B+D	853.0	878.2	890.1	904.0	973.4
A+B+E	853.8	837.5	882.2	880.5	995.8
A+B+D+E	868.1	880.6	882.1	860.7	979.0
A+B+C+D+E	903.6	882.7	867.8	855.5	984.2
A+B+C+E	849.1	865.6	854.0	926.4	975.9

concentrations and the real values is 842.1 and the LS-SVM based method can obtain the concentration predictions with similar accuracy.

Moreover, [4] also reveals that the best (lowest) differences for each subset are the ones predicted only by knowing the feature A and the feature B given 100% of the data. 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 often sufficient to obtain comparable results, while the PK model depends on analyzing all the parameters needed in Equation (1), (2) or (3). In addition, not all the features are useful in drug concentration prediction given sufficient training data. However, with a reduced number in the data, some features which does not improve the prediction with 100% of the data turn out to be more important.

V. CONCLUSIONS

Applying Support Vector Machine in drug concentration analysis has been proved to obtain some comparable results as the traditional pharmacokinetic model. It can build personalized model for each patient in drug concentration analysis in order to help give the guidance to the doctors in the dosage prescription. Despite its advantages cited above, the limitation of this approach is that it highly depends on the results to the quality of the initial data obtained from patients, the library. Indeed, no assumption on the theoretical disposition of the drug in the body, according to the classical pharmacokinetic behavior of drugs, is made. This is probably the reason of the unfamiliar shapes of some of the curves observed. With the future work on including a prior of the theoretical distribution over the parameters, we expect an improving results. Under certain conditions, it, however, improves the predictions using fewer input features. This approach however deserves formal clinical validation with more data sets from various drugs undergoing therapeutic drug monitoring.

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] 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.
- [4] Wenqi You, Nicolas Widmer, Giovanni De Micheli, 'Example-based Support Vector Machine for Drug Concentration Analysis', 33rd IEEE EMBS, USA, August 2011, pp. 153-157.
- [5] Ethem Alpaydin, 'Introduction to Machine Learning (Adaptive Computation and Machine Learning)', MIT Press 2004, ISBN 0-262-01211-1.
- [6] Karl S. Ni, Truong Q. Nguyen, 'Image Superresolution Using Support Vector Regress', IEEE Transactions on Image Processing, vol. 16, June 2007, pp. 1596-1610.
- [7] 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.
- [8] 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.
- [9] P. Jeatrakul, K. W. Wong, 'Comparing the Performance of Different Neural Networks for Binary Classification Problems', IEEE SNLP 2009, pp.111-5.
- [10] Yoav Freund, Robert E. Schapire, 'A Short Introduction to Boosting', Journal of Japanese Society for AI, September 1999, 14(5):pp.771-780.
- [11] Steve R. Gunn, 'Support Vector Machines for Classification and Regression', Technical Report, University of Southampton.
- [12] Laura Auria, Rouslan A. Moro, 'Support Vector Machines (SVM) as a Technique for Solvency Analysis', Discussion Papers, DIW Berlin, August 2008.
- [13] 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
- [14] 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.
- [15] J. Suykens, T. V. Gestel, J. D. Brabanter, B. D. Moor, J. Vanderwalle, 'Least Squares Support Vector Machines', World Scientific, 2002.
- [16] Aly Farag, Refaat M. Mohamed, 'Regression Using Support Vector Machines: Basic Foundations', Technical Report, CVIP Laboratory, University of Louisville, 2004.
- [17] 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 α_1 -Acid Glycoprotein', Br J Clin Pharmacol 2006; 62:1 pp.97-112