

A Drug Administration Decision Support System

Wenqi You*, Alena Simalatsar*, Nicolas Widmer[‡] and Giovanni De Micheli*

**Integrated Systems Laboratory, EPFL, Switzerland 1015.*

Email: {wenqi.you, alena.simalatsar, giovanni.demicheli }@epfl.ch

[‡]*Centre Hospitalier Universitaire Vaudois and University of Lausanne, Switzerland.*

Email: nicolas.widmer@chuv.ch

Abstract—Drug delivery is one of the most common clinical routines in hospitals, and is critical to patients' health and recovery. It includes a decision making process in which a medical doctor decides the amount (dose) and frequency (dose interval) on the basis of a set of available patients' feature data and the doctor's clinical experience (a priori adaptation). This process can be computerized in order to make the prescription procedure in a fast, objective, inexpensive, non-invasive and accurate way. This paper proposes a Drug Administration Decision Support System (DADSS) to help clinicians/patients with the initial dose computing. The system is based on a Support Vector Machine (SVM) algorithm for estimation of the potential drug concentration in the blood of a patient, from which a best combination of dose and dose interval is selected at the level of a DSS. The addition of the RANdom SAMple Consensus (RANSAC) technique enhances the prediction accuracy by selecting inliers for SVM modeling. Experiments are performed for the drug *imatinib* case study which shows more than 40% improvement in the prediction accuracy compared with previous works. An important extension to the patient features' data is also proposed in this paper.

Keywords—Decision Support System, Support Vector Machine, RANSAC algorithm.

I. INTRODUCTION

As 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]. When one develops a CDSS, the two main problems that need to be addressed are: (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 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 [4], Prodigy [5], EON [6], GLIF [7],

SAGE [8], Guide [9], GLARE [10] and Asbru [11], [12]. These tools approach the problem of generalization of medical guidelines (GLs) 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 GLs representation presented until now.

In this paper we focus on a specific class of health care procedures, i.e. the administration of a suitable dose to a patient based on the prediction of the blood drug concentration. The current practice of drug administration for the first dose delivery process is mostly based on a patient's disease symptoms and a clinician's empirical knowledge (a priori adaptation). However, in some cases, e.g. when treating HIV, cancers, etc, the effective therapeutic concentration range is narrow, therefore the experience-based approach creates a risk of over- or under-dosing a patient. Both over- and under-dosing may be harmful. Hence, Therapeutic Drug Monitoring (TDM), an approach aiming at individually optimizing dose regimen of drug delivery based on the measurement of the drug concentration in blood [13], is more and more demanded in treatment procedures (either as a routine or as a rescue method). Therefore, TDM presents a new personalized approach in medicine. Usually, in the process of TDM, blood measurements are carried out regularly to check whether the prescribed dose is adequate to keep a proper drug concentration in the human body, that is within its therapeutic range. However, this procedure suffers from some disadvantages such as being slow, invasive, expensive, etc. Therefore, our work is focused on the drug's dose computations based on patients' personal features (i.e. a priori adaptation), aimed to make the TDM process faster.

The key contribution of this paper is a personalized Drug Administration Decision Support System (DADSS) aimed at assisting medical doctors in the daily dose computing procedure. The system gives a personalized suggestion on the drug's dose and dose administration interval with respect to different patients' features. It is based on an enhanced version of an SVM-based algorithm [14] that has already shown great potential in drug concentration prediction. DADSS consists of four phases: data inputs, data pre-processing, analysis and clinical decision making. The core of the pre-process phase is based on the RANdom SAMple Consensus

(RANSAC) filtering of the initial data which is essential to enhance the accuracy of the concentration predictions. The analysis phase is represented by the Support Vector Machine (SVM) algorithm that is carrying out the computation of the results which are then analyzed in the decision phase. The inference mechanism is built on both the knowledge base (library) and the input data using machine learning algorithms. DADSS can be considered as a solid brick for any general DSS aimed at assisting medical doctors when applying the TDM approach. For instance, it can close the verification loop of the TAT-based medical protocol representation [15] by bridging the modeling gap between treatment applied to a patient and its body reaction to the treatment.

The paper is organized as follows. Section II introduces the related work in Decision Support System. Section III proposes our DSS targeting drug dosing computations. Section IV discusses a survey on the *imatinib* case study. Section V draws the conclusion of this study.

II. RELATED WORK

In this section, we first discuss the current mechanisms used in Clinical Decision Support System. Then, we talk about some the existing DSSs of other domains based on SVM.

A. Clinical Decision Support System

In the literature, there exist many definitions of a Decision Support System (DSS) according to various purposes, within which Clinical DSSs (CDSSs) form a special type of DSSs that provide clinicians with medical guidelines of best practices in patient care according to clinical knowledge.

In the inference mechanism, computer algorithms are applied to process the clinical data. In [16], the inference mechanisms are divided into four categories: logic, procedural, structured representations, and graph/network.

- Logic: It describes the medical knowledge using ‘true’ or ‘false’ statements. These statements can be linked to Boolean operators such as ‘and’, ‘or’, and ‘and not’ to form sentences.
- Procedural: It aids the diagnostic and medical decisions by offering a ‘process’ instead of declarative facts. In practice, due to the uncertainty in the medical domain, it is recommended to consider both the logic signs and the procedural knowledge. Hence, Bayes’ rule [17], etc. have been applied to design the medical DSS.
- Structured Representations: This method emphasizes in grouping the knowledge into well defined pieces with higher levels of organizations. It can be used to store both declarative and procedural medical knowledge with or without certainty.
- Graph/Network: This includes Bayesian networks, decision trees, Artificial Neural Networks (ANNs), etc. All of them have their advantages in different aspects.

Bayesian networks are powerful in analyzing the conditional independencies to predict the output; decision trees are simple to understand and model; and ANNs have a strong ability to ‘learn’ a model from the observed data.

Each of the categories has its own specific application domain.

In this paper, the proposed DSS gives a dose recommendations similar to the Graph/Network category. It is based on a drug concentration prediction method using Support Vector Machine, a machine learning algorithm which has successfully proved the ability to solve many classification or regression problems.

B. Support Vector Machine based Decision Support System

Support Vector Machines (SVMs) were invented by Vapnik in 1979 and applied to classification and regression problems in 1995 [18]. They use a nonlinear mapping to transform the original training data into a higher dimensional space, within which they search for the linear optimal separating hyperplane, or ‘decision boundary’, to separate the two classes. Compared with other methods such as Neural Networks, Decision Trees, Adaptive Boosting, the advantages of using SVM are [19], [14]:

- More accurate results due to SVM’s ability to model complex nonlinear decision boundaries;
- Less possible to run into overfitting problems than other methods;
- Wide application areas including prediction, classification and regression.

Therefore, 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 [19], the authors use an SVM and an Artificial Neural Network (ANN) as bases for their heart diseases 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 [20], 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 [21], electric power information systems [22], 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 [23] 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..

III. SYSTEM DESCRIPTION

In this section, we introduce a global view of the Drug Administration Decision Support System (DADSS) and computational functions of the system.

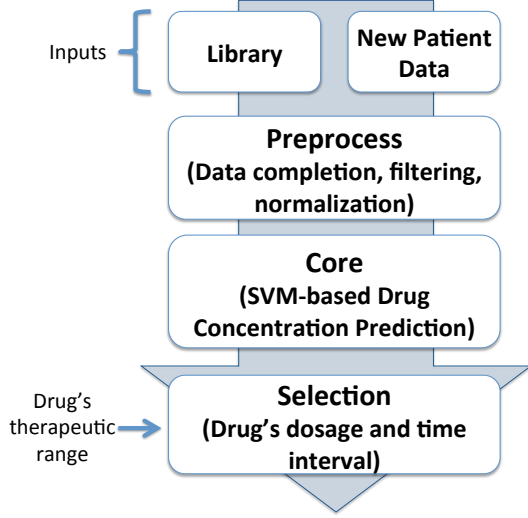


Figure 1. Flowchart of the Drug Administration Decision Support System

A. Drug Administration Decision Support System

The proposed system is composed of four main modules as shown in Figure 1: *Inputs*, *Preprocess*, *Core* and *Selection* parts. Input data are *Library* which stores the previous patients' feature data and *New Patient* features. *Library* is used to build a mathematical model which links the drug concentration to patients' features, while the *New Patient*'s data estimate the drug concentration used to generate the final decisions on dose computation.

The *Preprocess* module prepares the input data for the *Core* module. First of all, it checks the completion of patient's features. When a set of features of a new patient is available only partially, the system replaces the missing data by an average value to the corresponding feature data in *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:

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

Afterwards, an effective outlier-detection technique, Random Sample Consensus (RANSAC) (see Section III-C), is applied in order to enhance the accuracy of the prediction results.

The *Core* module runs an SVM-based drug concentration model using the preprocessed *Library* data and predicts the concentration values for a new patient. Instead of solving a convex *quadratic programming problem* (QP) as a theoretical SVM solver, DADSS simply uses a Least Square SVM (LS-SVM) classifier to give a solution by solving a set of linear equations [24]. In Section III-B, we give more details about this algorithm.

The *Selection* module chooses the best dose and dose

intervals according to the given therapeutic ranges (possibly one for the peak and another for the trough concentration). However, in practice, most drugs have only one therapeutic range available, which usually refers to the trough concentration. DADSS proposes different solutions accordingly.

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

To apply LS-SVM algorithm, we assume that there are N patient samples in *Library*, some of which can be obtained from the same patient, in the form of $(x_i, y_i) \in \{(x_1, y_1), \dots, (x_N, y_N)\}$, where y_i denotes the drug concentration values and x is a vector of d patient features, e.g. age, gender, body weight, etc [14]. The goal of SVM is to find a linear function $f(x) = w \cdot \phi(x) + b$ which approximates the relationship between the dataset points and can estimate output y according to a new input patient data. Here, $\phi(x)$ maps the input samples to a higher-dimensional feature space by applying a non-linear function in the original space; w and b stand for the weights of the feature space and offset respectively.

A loss function $\mathcal{L}(y, f(x)) = (y - f(x))^2$ 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:

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

where the constant C_0 determines the tradeoff (relative importance) 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 an RSS (Root of Sum of Square) square fitting error and a regularization term, which is also a standard procedure for the training of Multi-Layer Perceptron's (MLP) and is related to ridge regression [25], [26]. Applying Lagrangian analysis to solve the optimization problem of objective function [24], [26], we see that the optimal w can always be expressed by:

$$w = \sum_{i=1}^N \alpha_i \phi(x_i) \quad (3)$$

Plugging w into Equation (2), we can estimate α and b by solving the linear system:

$$\underbrace{\begin{bmatrix} \mathbf{K} + \frac{1}{C_0} \mathbf{I} & \mathbf{1} \\ \mathbf{1}^T & 0 \end{bmatrix}}_H \begin{bmatrix} \alpha \\ b \end{bmatrix} = \begin{bmatrix} y \\ 0 \end{bmatrix} \quad (4)$$

where \mathbf{K} is the kernel matrix defined by $K_{ab} = \phi(x_a)^T \phi(x_b)$. The use of the kernel matrix greatly helps reducing the computational complexity without explicitly computing $\phi(x)$, making use of the fact that the SVM algorithm depends only on dot products between sample patterns. Hence, after

Algorithm 1 RANSAC algorithm, where $data$ is a set of observations, $model$ is a model that can be fitted to 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: $bestmodel$

```

 $bestinliers \leftarrow \emptyset$ 
for  $i = 1 \rightarrow M$  do
   $possibleinliers \leftarrow \text{SampleUniformly}(data, K)$ 
   $possiblemodel \leftarrow \text{Fit}(model, possibleinliers)$ 
   $inliers \leftarrow \emptyset$ 
  for all  $point \in data$  do
    if  $\text{Distance}(point, model) < Th$  then
       $inliers \leftarrow inliers \cup \{point\}$ 
    end if
  end for
  if  $|inliers| > |bestinliers|$  then
     $bestinliers \leftarrow inliers$ 
  end if
end for
return  $bestmodel \leftarrow \text{Fit}(model, bestinliers)$ 

```

defining the kernel function, the least-square optimization problem could be solved simply by inverting the first term H in the left-hand side of (4).

Once we obtain the value of α and b , the output concentration of the new patient y could be then estimated through the prediction function: $f(x) = \sum_{i=1}^N \alpha_i \mathbf{K}(x_i, x) + b$.

C. RANdom SAMple Consensus Algorithm (RANSAC)

SVM-based algorithm for drug concentration prediction has already shown a great potential in increasing the accuracy of the prediction [14]. However, when looking at the training data, we have realized that some of the measurements have values that lie far from the analytically predicted value and form a set of points that we call outliers. When assuming that the actual measurements are 100% precise the only justification for the existence of outliers lie in the fact that some patient profile parameters were not taken into account. Outliers highly affect the accuracy of the SVM-based algorithm prediction of the average drug concentration value. Therefore, first of all we introduce the RANSAC algorithm that filters out the outliers and can increase the prediction precision by more than 40%.

The RANSAC [27] 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:

$$M = \frac{\log(1 - p)}{\log(1 - (1 - u)^K)}. \quad (5)$$

The model of the RANSAC algorithm is a linear combination of several basis functions, such as x , x^{-1} , x^2 , $\exp(x)$, $\log(x)$, etc.. 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. In Section IV, a detailed introduction of an approach to choose the basis functions is given with respect to the threshold values.

D. Extension of Patient Features

Although the RANSAC algorithm enhances the SVM prediction results by more than 40%, the number of used *Library* samples in the *Core* module decreases, due to the use of only the ‘inliers’. However, with a proper threshold in RANSAC, we believe that the reason of causing the outliers is the insufficiency in the patient features currently considered in the clinical practice. Thus we extend our algorithm such that it can account for all kinds of standard (such as age, weight) and non-standard (not used in the current clinical practice) patient features. In Figure 2, we classify the dataset features into three types: *Patient Profile*, *Physical Measurements*, and *External Parameters*. 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 are collected by using clinical tools, questionnaires, etc..

Unfortunately, in the case study in Section IV, when training our algorithm we depend on the available databases that were created to support existing analytical methods which account only for a limited number of features. Due to a high computational complexity of the currently used analytical methods, it would require an enormous effort to enforce all the extra features to increase at least one new feature, especially taking into account that it is not clear what are the other important parameters. Applying the SVM-based algorithm that accounts for new types of features allows one to study a large number (several dozens) of new parameters and determines which are the other important ones. Choosing new valuable parameters is critical to enhance the accuracy of the drug concentration prediction as well as can be potentially used to enhance the existing analytical methods. Therefore, we encourage the extension of existing drug concentration measurements databases with new patient profile parameters in clinical practice.

E. Selection Rules for Dose Computation

The proposed Drug Administration Decision Support System (DADSS) is based on the concentration prediction

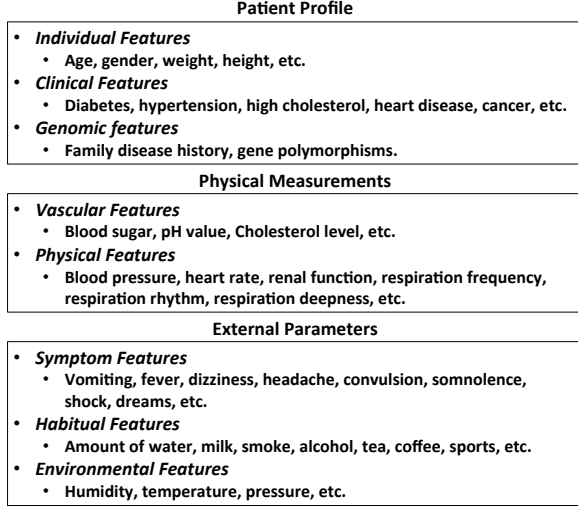


Figure 2. List of Three Groups of the Input Patients' Feature Data

results of the new patient given by the *Core* module. The final output of DADSS can be a recommended dose D^* and/or the dose interval τ^* . Currently we account for discrete sets of $D_j \in \{100, 200, \dots, 2000\}$ mg and $\tau \in \{1, 2, \dots, 24\}$ h, because those values are more convenient for oral tablets (as in our *imatinib* case study). For each value of D , DADSS plots a Concentration-Time curve, as in Figure 3, and chooses the best dose D^* and τ^* according to the given therapeutic ranges.

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. Ideal peak and trough's concentration values are assumed to be previously defined (thanks to former clinical trials), namely C_p^* and C_t^* respectively. If they are not available, the mean values of the higher and lower bounds of the corresponding ranges will be used as the ideal values. However, for most drugs including *imatinib*, only one therapeutic range (generally referred to the trough concentration) exists. DADSS can be adapted to only one therapeutic range as well. Let us assume that the drug concentration decreases monotonically after reaching its peak, since there are lower and upper bounds on the trough's therapeutic range, DADSS defines the output τ^* with a range (τ_h, τ_l) shown in Fig. 3 on the curve (c).

The *Selection* module chooses the best dose (D^*) and range of dose interval (τ_h, τ_l) out of a cross product of $D_j \in \{100, 200, \dots, 2000\}$ mg and $\tau \in \{1, 2, \dots, 24\}$ h.

- Case 1: If both peak and trough therapeutic ranges are available. D^* is determined as:

$$\operatorname{argmin}_{D_j} (|C_{j_{max}} - C_p^*|), \quad (6)$$

where $C_{j_{max}}$ stands for the peak concentration value within 24 hours after taking the dose D_j . This indicates

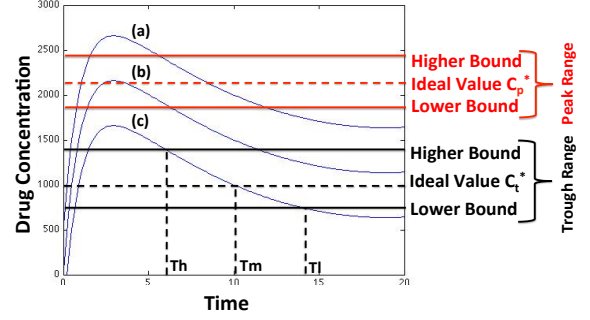


Figure 3. An Example of the Drug Concentration Curves Intersecting with the Peak and Trough Therapeutic Ranges. Curve (a)'s dose = 800mg, Curve (b)'s dose = 600mg, Curve (c)'s dose = 400mg.

the smallest difference between the ideal peak concentration value and the peak values estimated by the *Core* module corresponding to each D_j .

- Case 2: If only the trough therapeutic range is available. D^* is determined as:

$$\operatorname{argmin}_{D_j} (|C_{j_{24}} - C_t^*|), \quad (7)$$

where $C_{j_{24}}$ stands for the concentration values estimated at 24 hours after giving a dose D_j .

Having chosen the best dose D^* in both cases, DADSS computes the best $\tau^* = (\tau_h, \tau_l)$. Since we want to keep the trough drug concentration value within the trough's therapeutic range, τ_h and τ_l are computed according to the higher and lower bounds of the trough's therapeutic range respectively.

IV. CASE STUDY

Imatinib [28], a drug used to treat chronic myeloid leukemia and gastrointestinal stromal tumors, is the drug considered in our case study. Until now, only a trough's therapeutic range of this drug has been proposed and is presently being validated in a randomized clinical study in leukemia patients (I-COME; ISRCTN31181395). The trough range has a lower bound at 750mg/L, upper bound at 1500mg/L and target value at 1000mg/L [28], [29]. To apply our algorithm, we use the training data which include 54 patients and 252 samples in total. The testing data, in turn, contains 65 patients presenting gastrointestinal stromal tumors with 209 samples. The results are compared with the general population pharmacokinetic (PK) model [30] and the SVM-based model without RANSAC in the *Preprocess* module [14]. The set of input features of patient profile data includes: Gender, Age, Body Weight, and Type of pathology.

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

Table I
RANSAC BASIS FUNCTION ANALYSIS WITH RESPECT TO DIFFERENT THRESHOLDS. TH: THRESHOLD WITH UNIT [MG/L]. SCORE 0 STANDS FOR 'UNUSED' AND SCORE 1 FOR 'IN USE'.

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
1000	0	1	0	0	0	0	1	1	0
1500	0	1	0	0	0	0	0	0	0

of these basis functions. However, in our experiments, we find that not all the listed basis functions are utilized to estimate the inliers and the outliers. Table I shows the utilization results ('1' for 'used', '0' for 'unused') 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 chosen basis functions: $f(x) = \{x^{-2}, x, x^3, \log(x), \cos(x), (1 - \exp(-x)), \exp(x)\}$.

Besides an outlier-free input data, the effectiveness and the accuracy of the *Core* module highly depends on the choice of the kernel function. In our system, we select Gaussian distribution, a common choice with a single parameter σ , as the kernel function of the SVM algorithm. Hence, Equation 4 has two parameters to be estimated, C_0 and σ , the best combination of which is found by a grid-search with exponentially growing sequences, e.g. $\{C_0, \sigma\} \in \{10^{-2}, 10^{-1}, \dots, 10^3, 10^4\}$, through a 10-fold cross validation. An L -fold cross validation is a commonly-used method to estimate the parameters of a model over each observation value [31]. It randomly partitions the original training sample into L subsamples, of which a single subsample 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 σ with each of the L subsamples used exactly once as the validation data. We choose C_0 and σ to be the one of the L results having the least Mean Squared Error (MSE) between the predicted values and the 'validation data'.

A comparison of the prediction accuracy of our core algorithm is shown in Figure 4 and Figure 5. Figure 4 shows absolute values of the measured and the predicted drug concentrations for five new patients (selected randomly from our testing datasets). In these five sample cases, DADSS system achieves the highest accuracy (the predicted concentration values are closest to the measured ones) among the three approaches, while Pharmacokinetic (PK) method deviates much further in most cases. Figure 5 compares the absolute differences between the measured and the predicted concentrations. Histograms (a) and (b) of Figure 5 show that the PK mode and SVM without RANSAC [14] have several samples with large differences ($\geq 1000\text{mg/L}$) in their

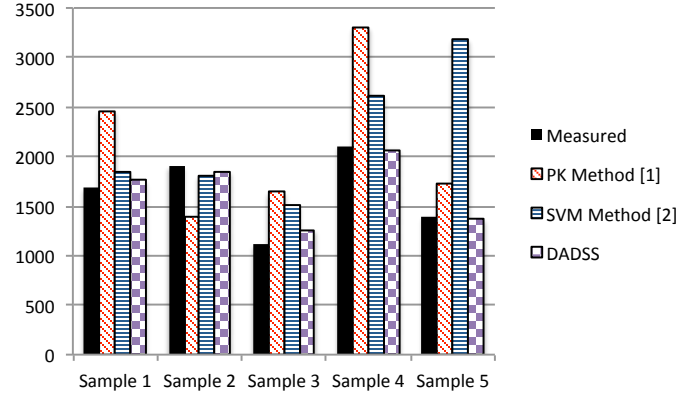


Figure 4. Comparisons of Absolute Values of the Measured Concentrations and the Predicted Ones.

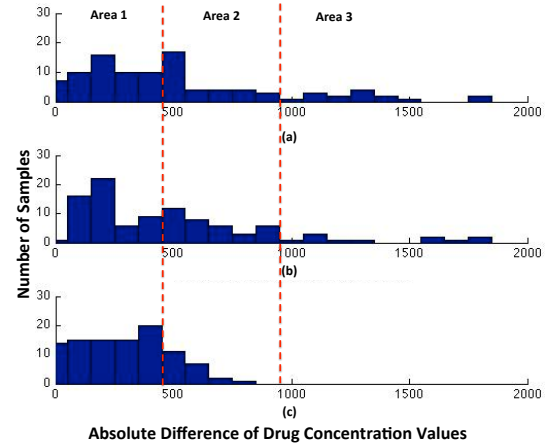


Figure 5. Histogram Comparisons of Absolute Difference between Measured Values and Predicted Values. ((a): PK-based algorithm, (b): SVM-based algorithm without preprocessing [14], (c): SVM-based algorithm with preprocessing.

concentration prediction results, which was caused by the outliers in the dataset samples. Statistically, 71% of the predicted drug concentration values will lay within the range of 500mg/L from the real measured values (summed number in 'Area 1' over the total number of patients in the test dataset, as shown in Figure 5). While when applying the PK model and SVM without RANSAC [14] this value will reach only 52% and 53% respectively. On average, the proposed system improves the mean prediction accuracy by about 44.7% over the PK model and 42.6% over SVM without RANSAC [14].

Table II describes some examples of how the decisions about *imatinib*'s dose and dose interval are made for 5 randomly selected patients. For each pair of the cross product of the dose D_i and the dose interval τ_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

Table II
5 SAMPLE RECOMMENDATIONS FROM DADSS. M: MALE, F: FEMALE,
G: GASTROINTESTINAL STROMAL TUMORS

No.	Patient Profile Features				Recommendations		
	Gender	Age	Body Weight	Disease	D^*	τ_h	τ_l
1	M	82	56kg	G	400mg	13h	24h
2	F	58	53kg	G	500mg	15h	24h
3	F	62	54kg	G	700mg	16h	24h
4	M	58	100kg	G	800mg	18h	24h
5	M	47	73kg	G	500mg	14h	24h

24h ($C_{i_{24}}$) are higher than the upper bound of the trough's therapeutic range as well as the ones whose predicted peak concentration value is lower than the trough's 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's therapeutic range, and selects the dose with respect to the smallest difference, as shown in Equation 7. For example, for patient 1, we obtained the set of $C_{i_{24}} = [890.6, 1032.5, 1152.5, 1239.1]mg/L$ that corresponds to the set of $D_i = [200, 400, 600, 800]mg$. The ideal value is $1000mg/L$, therefore $C_{i_{24}} = 1032.5mg/L$ has the smallest difference, and thus the curve whose $D^* = 400mg$ is chosen. Hereafter, the system obtains the range of the dose interval ($\tau^* = (\tau_h, \tau_l)$) according to the lower and upper bounds of the trough's therapeutic range.

V. CONCLUSIONS

In this paper we have presented a Drug Administration Decision Support System (DADSS) which aims to help patients and doctors in the drug dose computations for the prescription procedure. The system predicts the drug concentrations using patient's features and computes the best combination of dose and dosing interval for the patient (a priori adaptation). The main part of the system uses Support Vector Machine (SVM)-based algorithm to model the concentration-time curve while being trained by the *Library* database. The RANSAC algorithm is applied in the *Preprocess* module that enhances the prediction accuracy about 42% with respect to SVM without RANSAC technique. Even though this approach deserves formal clinical validation with more data sets, the case study on *imatinib* shows an improvement by 44.7% compared with the general PK method. We analyzed the data of 5 randomly chosen patients and have selected the best fitting dose and boundaries for the dose interval between two consequent drug administrations. The system accounts for a very large number of features which are currently not presented in the training database. Therefore, we encourage the extension of patient feature data sets.

ACKNOWLEDGMENT

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] D. Y. Chiu, C. C. Chang, M. W. Evens, J. C. Chern, D. B. Hier, D. A. Trace, and F. Naeymi-Rad, "A complete, hypermedia medical decision analysis support system," in *Computer-Based Medical Systems, Proceedings of IEEE Seventh Symposium on*, Jun 1994, pp. 16–21.
- [2] P. G. W. Keen and M. S. S. Morton, *Decision Support Systems: An Organizational Perspective*. Addison-Wesley Pub. Co., 1978.
- [3] G. Kong, D.-L. Xu, and J.-B. Yang, "Clinical decision support systems: A review on knowledge representation and inference under uncertainties," *International Journal of Computational Intelligence Systems*, vol. 1, no. 2, pp. 159–167, 2008.
- [4] D. R. Sutton, P. Taylor, and K. Earle, "Evaluation of PROforma as a language for implementing medical guidelines in a practical context," *BMC Medical Informatics and Decision Making*, vol. 6, pp. 20+, Apr. 2006. [Online]. Available: <http://dx.doi.org/10.1186/1472-6947-6-20>
- [5] P. Johnson, S. Tu, N. Booth, B. Sugden, and I. Purves, "Using scenarios in chronic disease management guidelines for primary care," in *Proceedings of American Medical Informatics Association (AMIA) Symposium*. PubMed Central PMCID, 2000, pp. 389–93.
- [6] S. W. Tu and M. A. Musen, "Modeling data and knowledge in the EON guideline architecture," *Medinfo*, vol. 10, no. 1, pp. 280–284, 2001. [Online]. Available: <http://view.ncbi.nlm.nih.gov/pubmed/11604749>
- [7] A. Boxwala, M. Peleg, S. Tu, O. Ogunyemi, Q. Zeng, D. Wang, V. Patel, R. Greenes, and E. Shortliffe, "GLIF3: a representation format for sharable computer-interpretable clinical practice guidelines," *Journal of Biomedical Informatics*, vol. 37, no. 3, pp. 147–161, Jun. 2004.
- [8] S. W. Tu, J. R. Campbell, J. Glasgow, M. A. Nyman, R. McClure, J. McClay, C. Parker, K. M. Hrabak, D. Berg, T. Weida, J. G. Mansfield, M. A. Musen, and R. M. Abarbanel, "The SAGE Guideline Model: Achievements and Overview," *J. of the American Medical Informatics Association*, vol. 14, no. 5, pp. 589–598, Jun. 2007.
- [9] P. Ciccarese, I. E. Caff, L. Boiocchi, S. Quaglini, and M. Stefanelli, "A guideline management system," *Stud Health Technol Inform*, vol. 107, no. Pt 1, pp. 28–32, 2004.
- [10] P. Terenziani, S. Montani, A. Bottrighi, M. Torchio, G. Molino, and G. Correndo, "The glare approach to clinical guidelines: main features," *Studies in Health Technology and Informatics*, vol. 101, 2004.

- [11] G. Duftschmid, S. Miksch, Y. Shahar, and P. Johnson, "Multi-level verification of clinical protocols," in *Proceedings of the Workshop on Validation Verification of Knowledge-Based Systems*, 1998, pp. 1–4.
- [12] S. Miksch, Y. Shahar, and P. Johnson, "Asbru: a task-specific, intention-based, and time-oriented language for representing skeletal plans," in *UK, OPEN UNIVERSITY*, 1997, pp. 1–9.
- [13] T. Buclin, V. Gotta, A. Fuchs, N. Widmer, and J. Aronson, "An agenda for UK clinical pharmacology monitoring drug therapy," *British Journal of Clinical Pharmacology*, vol. 73, no. 6, pp. 917–23, 2012.
- [14] W. You, N. Widmer, and G. De Micheli, "Example-based support vector machine for drug concentration analysis," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, Aug. 30–Sept. 3 2011, pp. 153–157.
- [15] A. Simalatsar and G. De Micheli, "Tat-based formal representation of medical guidelines: Imatinib case-study," in *proceedings of Engineering in Medicine and Biology Society (EMBC)*, San Diego, USA, Aug. 28–Sept. 1 2012.
- [16] J. Carter, *Clinical Decision Support Systems*. New York: Springer-Verlag, 1999.
- [17] H. R. W. Jr., "Iliad: Moving medical decision-making into new frontiers," *Methods of Information in Medicine*, vol. 1, no. 28, pp. 370–372, 1989.
- [18] C. Cortes and V. N. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [19] M. Gudadhe, K. Wankhade, and S. Dongre, "Decision support system for heart disease based on support vector machine and artificial neural network," in *Computer and Communication Technology 2010*, Sept. 2010, pp. 741–745.
- [20] L. Guo, Y. Wu, W. Yan, X. Shen, and Y. Li, "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 2005*, Jan. 2005, pp. 2413–2416.
- [21] X. Wang, P. Shu, X. Rong, and L. Nie, "A decision support system based on support vector machine for hard landing of civil aircraft," in *Computer Science-Technology and Applications, 2009.*, vol. 2, Dec. 2009, pp. 66–70.
- [22] X. Cheng, Y. Wei, and X. Geng, "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, Aug. 2009, pp. 536–539.
- [23] D. Basak, S. Pal, and D. C. Patranabis, "Support Vector Regression," *Neural Information Processing*, vol. 11, no. 10, pp. 203–224, 2007.
- [24] J. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle, *Least square support vector machine*. World Scientific Pub. Co., Singapore, 2002.
- [25] G. Golub and C. F. Van Loan, *Matrix Computations*. John Hopkins University Press, 1989.
- [26] J. D. Brabanter, "Ls-svm regression modelling and its applications," *PhD thesis*, pp. 25–28, 2004.
- [27] M. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [28] N. Widmer, L. Decosterd, C. Csajka, S. Leyvraz, M. A. Duchosal, A. Rosselet, B. Rochat, C. B. Eap, H. Henry, J. Biollaz, and T. Buclin, "Population pharmacokinetics of imatinib and the role of α_1 -acid glycoprotein," *British Journal of Clinical Pharmacology*, vol. 62, no. 1, pp. 97–112, 2006.
- [29] V. Gotta, N. Widmer, L. Decosterd, C. Csajka, M. Duchosal, Y. Chalandon, D. Heim, M. Gregor, and T. Buclin, "Therapeutic drug monitoring (TDM) of imatinib: Effectiveness of bayesian dose adjustment for CML patients enrolled in the imatinib concentration monitoring (I-COME) study," *ser. Swiss Medical Forum*, 2010, p. 285.
- [30] D. W. A. Bourne, *Mathematical Modeling of Pharmacokinetic Data*, 2nd ed. Technomic Publishing Company, Inc, 1995.
- [31] G. J. McLachlan, K. A. Do, and C. Ambroise, *Analyzing microarray gene expression data*, 2nd ed. John Wiley and Sons, Inc., Hoboken, New Jersey, 2004.