Original Research Article

Personalized drug concentration predictions with machine learning: an exploratory study

Danish Shakeel[1], Shakeel Ahmad Mir[2]*

[1]Department of Computer Sciences and Engineering, University Institute of Engineering, Chandigarh University, Mohali, Punjab, India
[2]Department of Clinical Pharmacology, Sher-I-Kashmir Institute of Medical Sciences, Srinagar, Kashmir, India

*Correspondence:
Dr. Shakeel Ahmad Mir,
Email: drshakeelahmadmir@gmail.com

ABSTRACT

Background: The dose individualization by therapeutic drug monitoring (TDM) can be improved if population-based reference ranges are available, as there is large inter- and intrapatient variability. If these ranges are not available, dose individualization may not be optimal. Machine learning can help achieve accurate drug dose settings and predict the resultant levels.

Methods: Two random forest models, a multi-class classifier to predict dose and a regression model to predict blood drug level were trained on 320 patients' data, consisting of their age, sex, dose and blood drug level. The classifier consisted of 1000 estimators (decision trees) and the regression model consisted of 1300 estimators. The model was evaluated on randomly split test set having 10% of the total dataset size. The regression model was compared against k-Nearest neighbor and linear regression models. The classifier was evaluated using accuracy, precision, and F1 Score; the regression model was evaluated using R2, Root mean squared error, and mean absolute error.

Results: The classifier had an out-of-sample accuracy of 68.75%, average precision of 0.7567, and an average F1 score of 0.6907. The regression model had an out-of-sample R2 value of 0.2183, root mean squared value of 3.7359, and a mean absolute error of 2.5156. These values signify an average classification performance, and a below-average regression performance due to small dataset.

Conclusions: It is possible for machine learning algorithms to be used in therapeutic drug monitoring. With a well-structured, rich, and large dataset, a very accurate model can be built.

Keywords: Drug concentration, Dose, Machine learning, Therapeutic drug monitoring

INTRODUCTION

Therapeutic drug monitoring (TDM) is the process of measuring drug concentrations in patients and using these measurements to individualize dosing regimens.[1] It is generally used for monitoring the drugs with a narrow safety margin and wide pharmacokinetic variability.[2] In TDM it is assumed that a definable relationship exists between dose and blood drug concentration, and between the blood drug concentration and drug effects.[3]

The drug concentrations measured in TDM are compared to predefined therapeutic ranges that are considered to reflect the optimum efficacy and safety of the drug.[4]

Therapeutic ranges used in TDM are mere recommendations based on the clinical response of a small group of patients taking the drug.[5] Most drugs are usually assigned a single therapeutic range for all indications, at all ages, and regardless of co-medication and co-morbidities. Therapeutic ranges should be population and

indication-specific.[4] There is not a generally accepted method to estimate these ranges and there is also no consensus whether plasma or serum should be preferred.[6]

Genetic polymorphism is associated with marked inter-individual variability in drug response and toxicity. Genetic heterogeneity has been reported among the Indian populations too.[7] It has been observed that 20.56% North Indian population are poor drug metabolizers for some specific drug metabolizing enzymes.[8] In ethnic Kashmiri population CYP2C9*3 has been found the most frequent mutant allele.[9] Due to these genetic variations, the reference ranges of drugs used in TDM, established elsewhere may not be fit for this population. The availability of local population-based therapeutic ranges may improve the validity of TDM results. TDM facilities are not available in every hospital and the clinicians face problems while tailoring the dose or targeting a specific drug level to maximize the therapeutic effect and minimize the toxicity.

Machine learning refers to the scientific algorithms and statistical models that machines learn from experience.[10] It involves different approaches such as decision tree, neural networks, and support vector machines, which are usually used for predictive models. Machine learning algorithms can be trained to predict blood drug concentrations. We attempted to develop a machine learning model to predict blood drug levels achieved with a specific dose and vice-versa. The model can help clinicians to make dose adjustments according to patient characteristics in the absence of proper TDM facilities and to find local therapeutic ranges.

METHODS

Dataset

TDM data for phenytoin was used to build and evaluate this model, as it exhibits non-linear kinetics. The dataset was obtained from the database of the therapeutic drug monitoring service of SKIMS Hospital Srinagar, Kashmir. The dataset contains anonymised TDM analyses of 320 patients who had received phenytoin alone for more than four weeks. The study was conducted from October, 2019 to February, 2020.

Inclusion criteria

Patients receiving phenytoin monotherapy for >4 weeks, suffering from well-established epilepsy, no co-medications, fully adherent patient, trough sample were included.

Exclusion criteria

Patients receiving combination antiepileptic drug therapy, co-medication with the drugs interfering with phenytoin kinetics, non-adherent patients, samples drawn for evaluating peak levels were excluded.

The dataset has four variables age, sex, dose and level.

The dataset was handled using a comma separated values (CSV) file.

Modelling

This data product shall be able to predict (forecast) the level using age, sex, and dose; and it shall be able to predict (classify) dose class using age, sex, and level. To achieve this, authors build two separate models, both Random Forest.
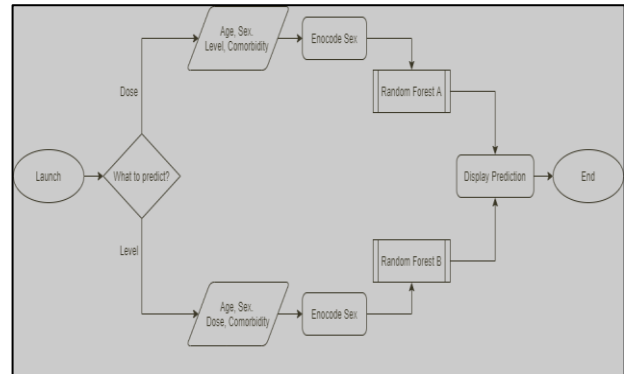

Figure 1: Overview of the architecture.

Predicting dose

Authors treat the problem of predicting the dosage as a multi-class classification problem. The feature variables-age, sex, and level are used to predict the target variable, i.e. dose class, which can be considered as a categorical feature.

Authors use a Random Forest, with the properties i.e. bootstraping, n_estimators (number of constituent decision trees) = 1000, Gini criterion.

The model is optimized using Grid Search Algorithm over n_estimators ranging from 500 to 2000.

Since the prediction is in the form of a class, e.g., (100,150), it is important to consider co-morbidities, which are not inherently included in this dataset. For a subject with co-morbidities or with an age of 60 or greater, authors usually suggest a lower dose, which is achieved in this model by rounding it off to the lower limit.[1,2] The subject is assigned the higher limit otherwise.

Predicting level

The problem of predicting the level can be seen as a forecasting or regression problem, with age, sex, and dose as the features, and level as the target variable. Since, there exists no definitive linear relationships among the features and target variables, a linear regression model will poorly

fit the data, and will not be able to correctly forecast or predict, therefore- authors use a random forest.

The random forest model has the following properties:

- *Bootstraping*
- *n_estimators* (number of constituent decision trees) = 1300
- *Gini* criterion

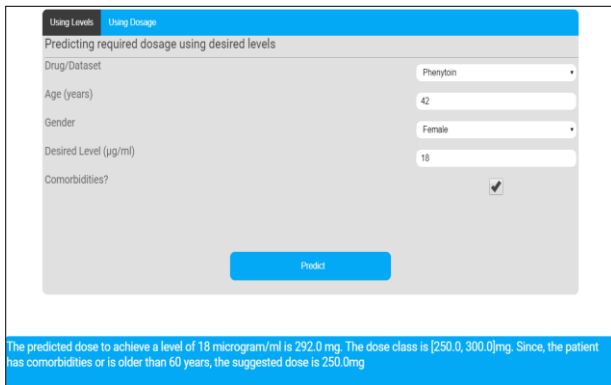The model is optimized using Grid Search Algorithm over n_estimators ranging from 1000 to 2500.

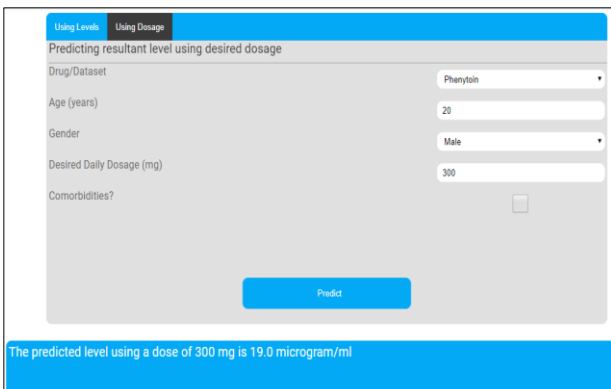**Figure 2: Predicting the dose to achieve 18 µg/ml in a 42-year-old female subject with comorbidities.**

Figure 3: Predicting the level achieved by a daily dose of 300 mg in a 20-year-old male with no comorbidities.

Evaluation

Both of the models were evaluated on different metrics, since one of the models is a classifier, and the other is a forecasting or regression model.

For evaluating the classifier, authors have used Accuracy-which is the measure of how many correct predictions this model has made over the total predictions; Precision-which is the measure of how many correct predictions has this model made over all the predictions made for the class; and F1 score, which is the harmonic mean of precision and recall, mathematically:

$$2 * \frac{precision*recall}{precision+recall}.$$

For evaluating the regression model, authors have used R^2 (R-squared), also called coefficient of determination, Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). The model has been compared with a k-Nearest Neighbour model, and a Multiple Linear Regression model.

The data is randomly split into training and test set, with a ratio of 0.1. The training set contains 288 samples, and the test set contains 32 samples. Since the dataset is very small, the models have been evaluated on both- the training set (called: In-sample evaluation), and on the test set (called: Out-of-sample evaluation).

Deployment

As a proof-of-concept, the models were deployed as a custom-built software (webapp) on the cloud. The software was built using Python, Flask, HTML, CSS, and JavaScript; and can be accessed using the following link: http://drug-administration-ml.herokuapp.com/.

RESULTS

In this study 320 anonymised TDM analyses were included. 76.2% (n=244) were from male patients and 23.8% (n=76) from female patients (Table 1).

Table 1: Demographics.

Gender	Frequency	%	Valid percent	Cumulative percent
Male	244	76.2	76.2	76.2
Female	76	23.8	23.8	100.0
Total	320	100.0	100.0	

5.6% patients (n=18) received 100 mg, 0.3% (n=1) 150 mg, 20.9% (n=67) 200 mg, 6.9% (n=22) 250 mg, 52.8% (n=169) 300 mg, 5.9% (n=19) 350 mg and 7.5% (n=24) 400 mg daily dose of phenytoin (Table 2). The mean age was 38.39±15.45 years. The mean dose was 274.37±69.59 mg. The mean measured level was 16.28±3.97 µg/ml (Table 3).

Table 2: Phenytoin daily dose (mg).

Dose	Frequency	%	Valid percent	Cumulative percent
100	18	5.6	5.6	5.6
150	1	0.3	0.3	5.9
200	67	20.9	20.9	26.9
250	22	6.9	6.9	33.8
300	169	52.8	52.8	86.6
350	19	5.9	5.9	92.5
400	24	7.5	7.5	100.0
Total	320	100.0	100.0	

Table 3: Mean and standard deviation of variables.

Variable	Descriptive statistics				
	N	Min.	Max.	Mean	Std. deviation
Age	320	10	80	38.39	15.457
Sex	320				
Dose	320	100	400	274.37	69.591
Level	320	0.5	28.0	16.288	3.9738

This classifier has 96.25% in-sample accuracy, and 68.75% out-of-sample accuracy. Authors also observe that the in-sample average precision over all the classes of this classifies is 0.9630, and the in-sample average F1 score is 0.9620. The out-of-sample accuracy of the classifier is 68.75%, the average precision is 0.7567, and the F1 score is 0.6907 (Table 4).

Table 4: Classifier evaluation.

Model	Sample	Accuracy	Precision (average)	F1 score (average)
Random Forest	In-sample	0.9625	0.9630	0.9620
	Out-of-sample	0.6875	0.7567	0.6907

This regression model has a fairly average in-sample R^2 score of 0.56, which is 34.32% better than the KNN model and 301.15 % better than the baseline Linear Regression model. The model has a below average out-of-sample R^2 score of 0.2183, a root mean squared error of 3.7359, and mean absolute error of 2.5156 (Table 5).

Table 5: Regression model evaluation.

Model	Sample	R-squared	RMSE	MAE
Random forest	In-sample	0.5600	2.6328	1.6019
	Out-of-sample	0.2183	3.7359	2.5156
KNN	In-sample	0.4169	3.0295	1.9472
	Out-of-sample	0.2107	3.7540	2.6062
Linear regression	In-sample	0.1396	3.6802	2.8253
	Out-of-sample	0.1362	3.9272	2.9028

DISCUSSION

The TDM results are interpreted by comparison with reference intervals. The data for setting reference intervals may come from literature, manufacturers, data mining or other laboratories.[11]

Current clinical practice depends mainly on the traditional PK models to work out the drug concentration values. These models consider only a few patient features and also assume that the patient is receiving only one drug. On the other hand Machine learning models take into consideration as many patient features as possible, can consider co-medications, and do not require to be explicitly programmed. Therefore, the machine learning approach can be considered appropriate in the drug concentration prediction.[12]

A machine learning model is as good as the training data. With limited data, authors attained an accuracy of 68.75% (classification) and R-squared value of 0.2183 (regression) in this study. Properly collected, well-structured, and large dataset will make the model drastically better, and extremely accurate, hence fit for actual drug monitoring purposes.

In some other studies conducted within the previous few years on this relatively new concept, similar results are found. Imai in a study on vancomycin dosing using machine learning, found that machine learning is useful in drug dose setting.[13] Wenki You in a study on algorithmic approach to personalized drug concentration predictions has presented various machine learning algorithms to solve the problems in drug concentration predictions. The researcher found that RANSAC algorithm produces a more reasonable concentration curve.[12] Hu et al, in a study on prediction of digoxin dosage found that machine learning models can accurately predict initial digoxin dose and help clinicians in safe use of digoxin.[14] Yao et al, found that machine learning algorithms are the best alternatives to predict serum digoxin concentrations whenever blood samples from newborn infants for therapeutic drug monitoring are not available.[15]

Goicoechea et al, has found that machine learning models can be helpful in interpretation of plasma concentrations of antiretrovirals in therapeutic drug monitoring.[16] Zhang et al, opine that machine learning techniques have the potential to improve the development and validation of predictive modeling in critical care research.[17]

CONCLUSION

Authors studied the concepts of some machine learning algorithms and how they can be applied to the blood drug concentration prediction. It is possible for machine learning algorithms to be used in therapeutic drug monitoring especially where TDM facilities are limited. With a well-structured, rich, and large dataset, a very accurate model can be built.

REFERENCES

1. Walson PD. Therapeutic drug monitoring in special populations. Clin Chem. 1998;44(2):415-9.
2. Kang JS, Lee MH. Overview of therapeutic drug monitoring. Korean J Int Medi. 2009;24(1):1.

3. Gross AS. Best practice in therapeutic drug monitoring. Br J Clini Pharmacol. 2001;52(S1):5-9.
4. Cooney L, Loke YK, Golder S, Kirkham J, Jorgensen A, Sinha I, Hawcutt D. Overview of systematic reviews of therapeutic ranges: methodologies and recommendations for practice. BMC Medi Res Methodol. 2017;17(1):84.
5. Lucas C, Donovan P. Medications:'Just a repeat': When drug monitoring is indicated. Austr Family Physic. 2013;42(1/2):18.
6. Hiemke C, Baumann P, Bergemann N, Conca A, Dietmaier O, Egberts K, et al. AGNP consensus guidelines for therapeutic drug monitoring in psychiatry: update 2011. Pharmacopsychiatry. 2011;21(06):195-235.
7. Umamaheswaran G, Kumar DK, Adithan C. Distribution of genetic polymorphisms of genes encoding drug metabolizing enzymes & drug transporters-a review with Indian perspective. The Ind J Medi Res. 2014;139(1):27.
8. Varshney E, Saha N, Tandon M, Shrivastava V, Ali S. Prevalence of poor and rapid metabolizers of drugs metabolized by CYP2B6 in North Indian population residing in Indian national capital territory. Springerplus. 2012;1(1):1-7.
9. Kousar S, Wafai ZA, Wani MA, Jan TR, Andrabi KI. Clinical relevance of genetic polymorphism in CYP2C9 gene to pharmacodynamics and pharmacokinetics of phenytoin in epileptic patients: validatory pharmacogenomic approach to pharmacovigilance. Int J Clini Pharmacol Therap. 2015;53(7):504-16.
10. Kavakiotis I, Tsave O, Salifoglou A, Maglaveras N, Vlahavas I, Chouvarda I. Machine learning and data mining methods in diabetes research. Computat Structural Biotechnol J. 2017;15:104-16.
11. Jones G, Barker A. Reference intervals. Clini Biochem Rev. 2008;29(Suppl 1):S93.
12. W. You. An Algorithmic Approach to Personalized Drug Concentration Predictions. Doctoral dissertation, 2014. Available at: https://pdfs.semanticscholar. org/c003/0746b9f344baa795cce838bc5a238e7cf9b8. pdf. Accessed on 3 January 2020.
13. Imai S, Takekuma Y, Miyai T, Sugawara M. A New Algorithm Optimized for Initial Dose Settings of Vancomycin Using Machine Learning. Biolog Pharmac Bull. 2020;43(1):188-93.
14. Hu YH, Tai CT, Tsai CF, Huang MW. Improvement of Adequate Digoxin Dosage: An Application of Machine Learning Approach. J Healthcare Engineer. 2018;2018.
15. Yao SH, Tsai HT, Lin WL, Chen YC, Chou C, Lin HW. Predicting the serum digoxin concentrations of infants in the neonatal intensive care unit through an artificial neural network. BMC Pediatrics. 2019;19(1):1-1.
16. Goicoechea M, Vidal A, Capparelli E, Rigby A, Kemper C, Diamond C, et al. A computer-based system to aid in the interpretation of plasma concentrations of antiretrovirals for therapeutic drug monitoring. Antiviral Therapy. 2007;12(1):55.
17. Zhang Z, Ho KM, Hong Y. Machine learning for the prediction of volume responsiveness in patients with oliguric acute kidney injury in critical care. Crit Care. 2019;23(1):112.