Build End-to-End ML Pipeline for Warfarin Dosing Prediction

Navyasree Chenchu & Shri Gouri Pinjarla
Saint Louis University
Instructor: Dr. Jie Hou

May 12, 2025

Introduction

The medication Warfarin is a very important anticoagulant for preventing and treating thromboembolic complications, including atrial fibrillation, cardiac valve replacements, and joint replacements. Despite the long clinical use of warfarin, dosing in patients remains a challenging job for doctors because of its narrow therapeutic window and large inter-patient response variation. The difficulty comes from genetic and medical as well as patient-specific factors playing complex interactions in appropriate dosage determination. Warfarin toxic doses have been associated with life-threatening complications, such as bleeding and thrombosis, which makes it important to improve dose prediction accuracy for patient safety. Conventional dosing routines typically involve trial and error titration of the drug, with periodic check-ups to monitor the patient's reaction to alterations and adjust doses as needed.

This practice is likely to be characterized by increased medical costs and greater adverse effects. The use of machine learning algorithms to calculate the optimal therapeutic dose from clinical and pharmacogenetic information builds an effective strategy to reduce risks of inappropriate dosing and avoid unnecessary health costs. Here, the objective is to develop an ML pipeline to predict the warfarin therapeutic dose at treatment initiation with the incorporation of patient-specific International Warfarin Pharmacogenetics Consortium clinical and genetic data. Using a dataset of more than 5700 warfarin-treated patients, where data on the participants involved demographic data, medical histories, and genetic data, determinants so warfarin dose requirement can be completely investigated.

Methods

Data Acquisition and Preprocessing

Data source of the project was the website of PharmGKB and material therefrom holds an adequate dataset related to information about patients regarding warfarin dose choice. Based on selected features, the analysis was carried out particularly using demographic variables (Gender, Age and Race) and clinical (Height, Weight, Diabetes) and drug variables (Simvastatin and Amiodarone). The Genetic part of the dataset (Cyp2C9 and VKORC1 genotypes, which are of the highest significance to warfarin metabolism) was also

considered. Some essential steps were performed as part of the data preprocessing phase. The dataset contained missing data points for: Diabetes, Simvastatin and Amiodarone; which were addressed through an imputation process. Most common value per column was utilized to fill missing values to keep the data set intact. Categorical variables were given numeric value through one-hot encoding of Gender, Race, and Cyp2C9 genotypes. Continuous variables such as Height and Weight required standardization through scaling techniques for ensuring consistency in feature scales, which was required for a highly efficient machine learning system. After preprocessing the dataset, it was divided into a training set-sized 80% of data and a test set-sized 20% of data with an emphasis on fair evaluation of the performance of the models on new data.

Modeling Approaches

Two different machine learning approaches were employed to address this problem: regression and classification. Regression was used to estimate the Continuous Therapeutic Dose of Warfarin and the classification was used to categorize patients into clinically relevant dose classes. Those that required over 30mg/week were classified as high dose, and those requiring 30mg/week or less as low dose. In predicting of Therapeutic Dose of warfarin, Random Forest Regressor was utilized since there is ability to detect intricate feature interactions, and utilization of this model prevents over-fitting. To categorize patients into high and low dose group, the choice was the Random Forest Classifier since it is adequate to binary problems and works well on imbalanced data. Training was conducted on-train data for both the models and a cross-validation was used for tuning of hyperparameters. In the process of regression, the objective was minimization of errors in predicting the exact warfarin dose and so on, whereas the process of classification focused on effective patient population through the intermediary of dose requirements.

Model Evaluation

The performance of the models was measured against a list of metrics. The regression model's performance was measured in terms of the following key metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE) and R-squared (R^2). respond The MAE and MSE estimate prediction errors, whereas the R^2 score represents the percentage of the variance of the target

variable explained by the model. Negative value of R^2 indicates that the model is worse than the mean predictions of the target. To verify how well the classification model performs, Accuracy, Precision, Recall, and F1-Score were used to evaluate its accuracy in classifying patients having either high or low dose class. The AUC and the ROC Curve were used to take into account the relationship between the false positive and the true positive depending on different thresholds.

Hyperparameter Tuning

In the quest for better model performance GridSearchCV was applied to obtain the best hyperparameters for the regression model. Hyperparameter tuning took into consideration the number of estimators (`n_estimators`), the maximum depth of the trees (`max_depth`). Grid search cross-validation demonstrated that Random Forest Regressor gained its optimal performance after changing `n_estimators` to 100 and `max_depth` to 10. Holding the identified optimal values, the model was re-trained so as to ensure that the best configuration was used.

Results

Regression Model Results

In the quest for better model performance, GridSearchCV was applied to obtain the best hyperparameters for the regression model. Hyperparameter tuning took into consideration the number of estimators (`n_estimators`), the maximum depth of the trees (`max_depth`). Grid search cross-validation demonstrated that Random Forest Regressor gained its optimal performance after changing `n_estimators` to 100 and `max_depth` to 10. Holding the identified optimal values, the model was re-trained so as to ensure that the best configuration was used.

Classification Model Results

The Random Forest Classifier was applied to predict whether patients require a high dose (>30mg/week) or a low dose ($\leq$ 30mg/week). The classification model achieved an accuracy of 71.43%, with a precision of 0.71 for class 1

(high dose) and 0.0 for class 0 (low dose). The Recall was 1.00 for class 1 and 0.00 for class 0. These results indicate a significant imbalance in the predicted classes, with the model predominantly predicting high dose patients. The poor performance in predicting low dose patients suggests that the model may benefit from techniques to address class imbalance, such as resampling methods (e.g., SMOTE) or class weighting in the Random Forest algorithm.

Best Hyperparameters

Through hyperparameter tuning using GridSearchCV, the optimal parameters for the Random Forest Regressor were determined to be:

- **n_estimators**: 100
- **max_depth**: 10

These settings provided the best performance based on cross-validation results.

Discussion

The performance of the model, as indicated by the negative R^2 score, indicates that significant improvement is necessary in the current model. An R^2 with a negative value indicates that the model is failing to capture the underlying patterns of the data and that its forecasts are worse than simply forecasting the mean value of the target variable. Based on the results, it is not possible to say that the model is able to adequately control the variability associated with estimation of the optimal warfarin dose. One motivation to address this limitation is better feature engineering, where more in-depth analysis of the relationships between available features and outcome being predicted is performed. A hopeful approach to feature engineering is the transformation of continuous parameters such as age, height and weight in such a way that it more accurately reflects their combined effects. Without other major characteristics such as the lifestyle choices of patients or specific genetic markers or environmental factors that impact how warfarin is metabolized, the model could perform better.

These extra features will allow the model's accuracy in predictions to improve. It could also be done by trying out some more high-tech methods,

i.e., XGBoost or Neural Networks in the classification algorithm. Such models as XGBoost and Neural Networks are very good at detecting complex, non-linear patterns between features and targets, and they may be able to deliver greater predictive accuracy than traditional techniques (e.g Random Forests). The gains from having high-dose and low-dose classes tilted against the performance of the classification model resulting in it predicting high-dose patients more than low-dose patients. This imbalance is problematic as the model had trouble predicting low-dose patients, which is important as the ability to make adequate predictions is essential to any model being useful to clinical practice. It might help balance the imbalance by techniques such as SMOTE to generate synthetic examples of the minority class or reduce the numbers of the majority class. Adjusting the class weights of the Random Forest Classifier is a different method to focus on the minority class and consequently improve the model's accuracy to predict both high and low dose groups.

Web Application

Using Gradio, a web application was created to make the model more accessible, which would allow patients to enter their information and receive predictions on the needed warfarin dose. The user interface is designed to ease user experience making input of patient data by medical professionals straightforward even if the user lacks technical competence. Clinical as well as genetic information such age, weight, sex, INR and Cyp2C9 genotypes can be conveniently entered by users through the application. Once received, the application uses trained algorithms of the machine learning to calculate and present a predicted warfarin dose estimation. With an intuitive and simple platform, the web application allows healthcare providers to make decisions about warfarin dosing more easily. Clinicians can use the software on their systems and it is accessible for use at `https://huggingface.co/spaces/ShriGouri12/Warfarin_dosage`. The current version is local, but cloud/clinical server deployment is planned so it can be used in a wider regard in the future. Doctors in various facilities may use the web application to be able to perform better at managing patient care and reducing mistakes when administering warfarin. At the end of the day, this web application is aimed at demonstrating the ways in which machine learning can be used successfully in healthcare, to make advanced data analysis tools easily available

to medical practitioners.

Conclusion

This project illustrated that machine learning techniques can be employed to unveil prediction of the suitable dose of warfarin for patients through the combination of clinical and genetic information. The regression model's performance was poor: its R^2 was negative; therefore, it shows the need for improvements in the construction of a model and feature selection so as to make reliable predictions. However, the classification model provided a worthwhile initial reference for distinguishing between high and low warfarin treatment needs for patients, but with imperfections due to inequitable distribution of data. The fact that the model classified a greater number of patients into high dose requirement category overshadowed the ability to accurately predict low dose needs. CNN training with a convolutional neural network was a fairly effective solution for the CNN models since it demonstrated good performance and needed fewer hyperparameter tuning attempts compared to the classical GCV, which was always inferior in AUC. By implementing these models into actual healthcare settings, institutions will reduce the number of mismanaged warfarin cases leading to better patient safety and avoidable health care visits. This is exemplified by the design of a Gradio web app that allows clinicians to enter data on patients and get dose predictions: a good example of the usefulness of machine learning in the practice of healthcare. As the model is made more advanced and sophisticated in the usage of advanced techniques and larger feature sets, it can be utilized as an important clinical tool for delivering precise and personalized warfarin dosage recommendations. Adding such models to clinical decision support systems would make healthcare delivery easy so as to reduce the risk of dosing mistakes, improve outcomes, and reduce overall healthcare costs. Further development in such a project as well as improvements on model accuracy and practical implementation have been expected to make machine learning a critical part of precision medicine.

Zoom Link

- https://slu.zoom.us/rec/share/0qRqYjJN6GAXG8IRGl3v8aK36JM7gxZIuHVQPpXTr5JrvbEKl7Y.2X1GE7BFlNeMxYZe
- Passcode: PFH!9vY&

Gradio Link

- https://huggingface.co/spaces/ShriGouri12/Warfarin_dosage

Google Colab Link

- https://colab.research.google.com/drive/1PoBrtqWKWFT3gRIJmLl1dIDVbyDWtggL?usp=sharing

References

- International Warfarin Pharmacogenetics Consortium. "Estimation of the warfarin dose with clinical and pharmacogenetic data." *New England Journal of Medicine*, 360.8 (2009): 753-764.
- https://www.hindawi.com/journals/cmmm/2015/560108
- https://www.nature.com/articles/s41598-023-49831-6
- https://pmc.ncbi.nlm.nih.gov/articles/PMC11487003/