Ensemble Machine Learning with Uncertainty Quantification for Pharmacogenetic-Guided Warfarin Dosing Across Diverse Populations

Favour Igwezeke

University of Nigeria, Faculty of Pharmaceutical Sciences, Nsukka

Abstract Text (max 350 words)

Background: Warfarin requires individualized dosing due to 10-fold inter-patient variability driven by genetic and clinical factors. Current pharmacogenetic algorithms achieve modest predictive accuracy ($R^2 \approx 0.47$ in published validation cohorts). Machine learning ensemble methods may improve precision medicine approaches while maintaining interpretability for clinical decision support.

Methods: We trained 8 machine learning algorithms (XGBoost, LightGBM, Random Forest, Gradient Boosting, Ridge, ElasticNet, SVR, MLP) on 4,671 patients from the International Warfarin Pharmacogenetics Consortium dataset, using 45 engineered pharmacogenomic features (CYP2C9/VKORC1 activity scores, drug interactions, anthropometric indices, and interaction terms). Predictions were averaged to create an ensemble. Validation was performed on 1,168 held-out patients with stratified analysis across four ethnic groups (White n=592, Asian n=328, Black/African American n=138, Unknown n=110).

Results: The ensemble achieved $R^2 = 0.476$, MAE = 8.0 mg/week, and clinical accuracy of 47.9% (within ±20% of actual dose). Uncertainty quantification produced 95% prediction intervals with mean width of 13.2 mg/week. Ethnic subgroup analysis revealed variable performance across populations: White patients ($R^2 = 0.412$, MAE = 8.5 mg/week), Asian patients ($R^2 = 0.346$, MAE = 6.3 mg/week), Black/African American patients ($R^2 = 0.277$, MAE = 9.9 mg/week), and Unknown ethnicity ($R^2 = 0.438$, MAE = 7.8 mg/week). Black/African American patients showed lower predictive accuracy despite highest mean dose requirements (39.4 ± 14.2 mg/week). Feature importance analysis identified anthropometric-genetic interactions (Age×BSA: 13.6% importance) as dominant predictors over single genetic markers (<3% individual importance). The framework included automated 4-tier risk stratification with clinical monitoring protocols.

Conclusion: Machine learning ensemble methods achieve competitive predictive performance for warfarin dosing across real-world clinical populations. Performance disparities across ethnic groups particularly lower accuracy in Black/African American patients highlight the importance of diverse training cohorts and population-specific model refinement. Feature importance findings suggest that anthropometric-genetic interactions contribute more substantially to dosing prediction than individual pharmacogenetic markers alone. Prospective clinical validation with stratified outcome assessment (time to therapeutic INR, adverse events) across diverse populations is warranted before clinical deployment.

Keywords

Warfarin, pharmacogenomics, machine learning ensemble, clinical decision support, precision medicine, ethnic disparities, health equity