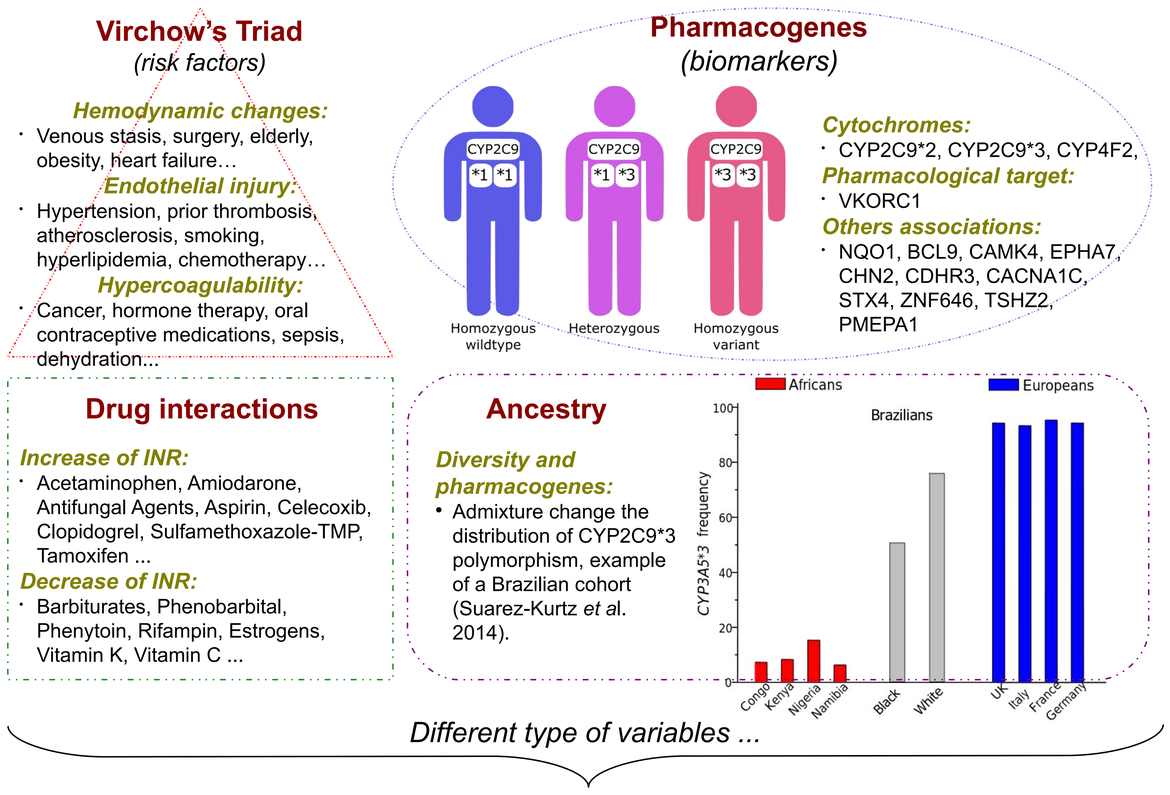
WARFARIN DOSE PREDICTION THROUGH A USER INTERFACE USING CLINICAL, DEMOGRAPHIC AND PHARMACOGENETIC DATA

Jennifer Montoya Neyra[1], Paulo Caleb J. L. Santos[3] and Júlia M.P. Soler[1,2]
[1]Bioinformatics Graduate Program, University of São Paulo.
[2]Statistics Graduate Program, University of São Paulo.
[3]Laboratory of Genetics and Molecular Cardiology, Heart Institute (InCor), FM-USP.

INTRODUCTION

Warfarin is an anticoagulant administrated orally, composed by a racemic mixture of R and S enantiomers, indicated in the prophylaxis and treatment of venous thrombosis, pulmonary embolisms, thromboembolic complications from atrial fibrillation or cardiac valve replacement, to reduce the mortality in cases of recurrent myocardial infarction, stroke and systemic embolization. Warfarin has a narrow therapeutic window, which tends to produce serious adverse effects such as bleeding, skin necrosis or even death. Its pharmacological action can be affected by different factors such as age, ancestry, diet, smoking, comorbidities, genetic mutations (pharmacogenes) that encode cytochromes (CYP2C9, CYP4F2), the enzyme epoxide-reductase of vitamin K (VKORC1), among others.

Thus, several pharmacogenetic algorithms have been developed to predict a safe dose of this drug, always with the main objective that these models are applicable to different people with different ancestralities maintaining accurate predictions. Contributing to this overall objective, we work on a user interface (UI) designed to assist in warfarin therapy by predicting a best therapeutic dose, calculated from the data entered into the UI. It will be able to predict more accurate doses for patients diagnosed with atrial fibrillation, stroke, thrombosis or heart valve prosthesis in whom it is desired to maintain an international normalized ratio (INR) between two and three, using their clinical, demographic and pharmacogenetic data from patients of the Brazilian Heart Institute (InCor - USP). of the Medical Faculty of the University of São Paulo (Santos *et al.*, 2015).



METHODS

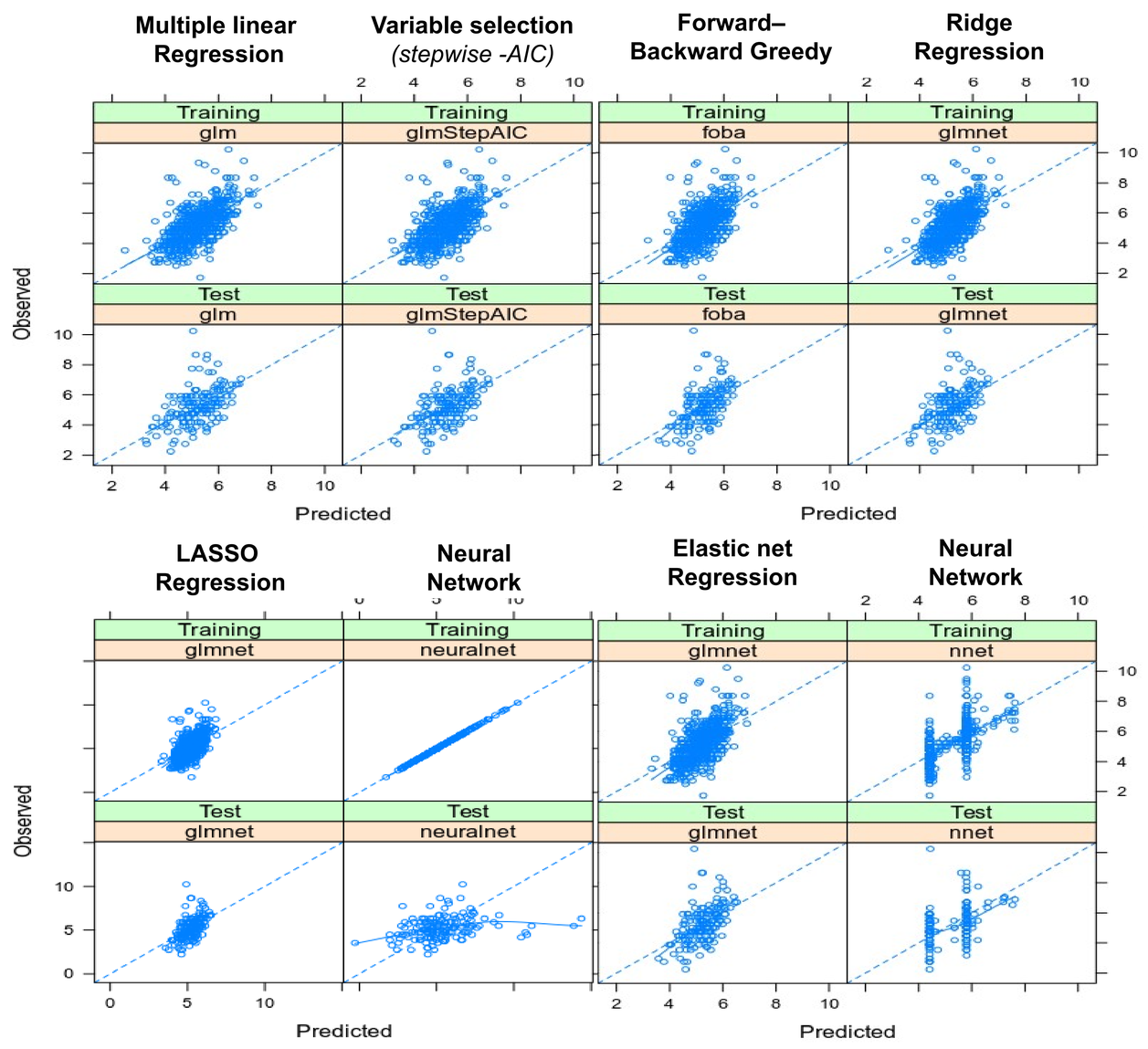
Data set: The data set with the information of 766 individuals, which reached international normalized ratio (INR) values between 2 and 3 when receiving a maintenance dose of warfarin was divided in the training (614 individuals) and test (152 individuals) subgroups. This data include clinical, demographic, pharmaceutical information and variant genotypes of the cytochrome P450 2C9 (CYP2C9), vitamin K epoxy reductase (VKORC1), leukotriene B(4) omega-hydroxylase 1 (CYP4F2) and NAD(P)H dehydrogenase (quinone) 1 (NQO1) pharmacogenes.

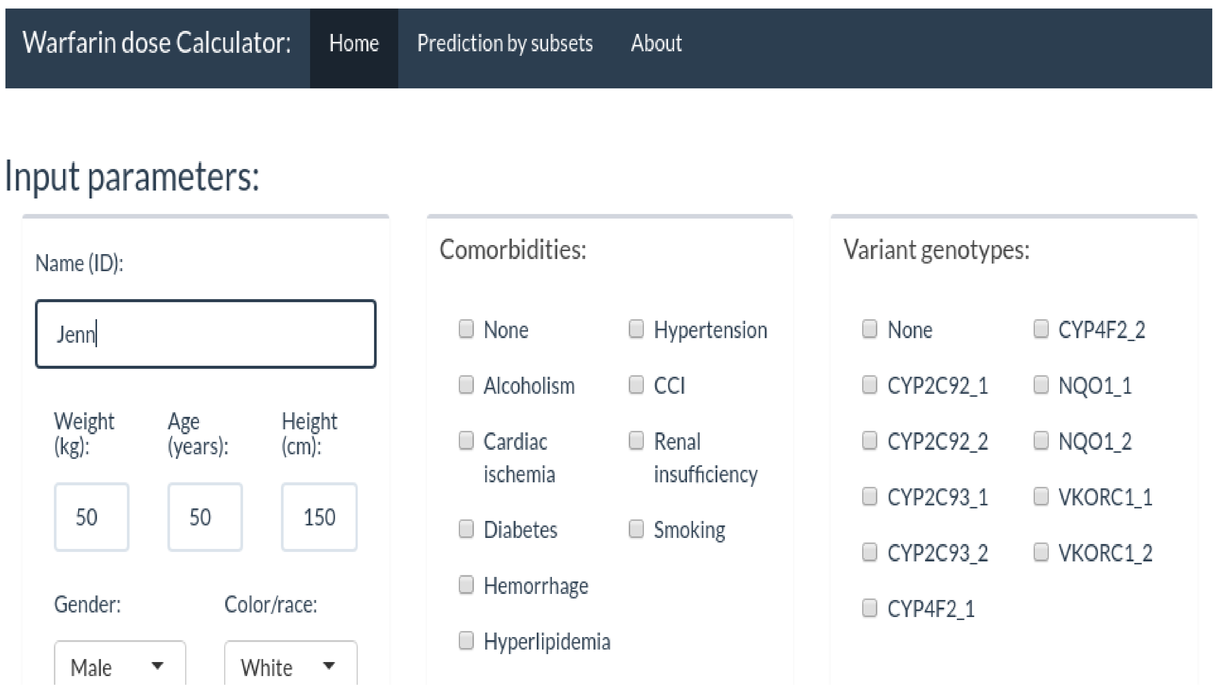
Prediction models: The fitting process and model performance analysis were developed in the R language, using the caret, MASS, foba, neuralnet and nnet packages. The following models were evaluated: the International Warfarin Pharmacogenetics Consortium (IWPC) algorithm, multiple linear regression, regression using regularizers (Lasso regression, Ridge regression), Elastic net regression, variable selection using information criteria (AIC), Forward–Backward Greedy algorithm (Foba) and a simple neural network model that consists of 3 hidden layers of 100 neurons each. To evaluate the accuracy of the models, the mean absolute error (MAE), root-mean-square error (RMSE) and R-squared were calculated.

Model	Hyperparameter	Optimal value
IWPC	-	-
Linear regression	-	-
Ridge	alpha	0
	lambda	0.219929
LASSO	alpha	1
	lambda	0.366233
Stepwise AIC	-	-
Foba	k	9
	lambda	0.126486
Elastic net	alpha	1
	lambda	0.035854
Neural network (*neuralnet*)	neurons (Layer 1)	100
	learning rate	0.001
	activation	identity
Neural network (*nnet*)	size	8
	decay	2.3645e-06

RESULTS

The best adjusted model was the Ridge Regression with Variable Selection, which obtained the best performance when analyzing both the training group (MAE = 7.54, RMSE = 0.993, R-squared = 0.296) and the evaluation group (MAE = 0.766, RMSE = 1.07, R-squared = 0.282).





CONCLUSION

The UI is still in development, but we have great expectations about its applicability and usefulness for patients who require it. Considering the sensitivity of the models in relation to the nature of the data, the UI will offer more accurate results, as long as the models can be trained with a larger number of patients.

REFERENCES

- P. C. J. L. Santos, L. R. Marcatto, N. E. Duarte, R. A. Gadi Soares, C. M. Cassaro Strunz,M. Scanavacca, J. E. Krieger, and A. C. Pereira. Development of a pharmacogenetic-based warfarin dosing algorithm and its performance in Brazilian patients: highlighting the importance of population-specific calibration. Pharmacogenomics, 16(8):865–876, jun 2015.
- F. Pedregosa, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer,R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, F. Pedregosa, G. Varoquaux, A. Gramfort, B. Thirion, P. Prettenhofer, J. Vanderplas, M. Brucher, M. et al., Scikit-learn: Machine Learning in Python Gaël Varoquaux. Journal of Machine Learning Research, 2011.
- T. Klein, R. Altman, N. Eriksson, B. Gage, S. Kimmel, M. Lee, N. Limdi, D. Page, D. Roden, M. Wagner, M. Caldwell, and J. Johnson. Europe PMC Funders Group Estimation of the Warfarin Dose with Clinical and Pharmacogenetic Data. N Engl J Med, 360(8):753–764, 2009.
- M. Pimorade. Pharmacogenetics and pharmacogenomics. Br J Clin Pharmacol , 52(4):345–347, 2001.
- Z. Ma, P. Wang, Z. Gao, R. Wang, and K. Khalighi. Ensemble of machine learning algorithms using the stacked generalization approach to estimate the warfarin dose. PLOS ONE, 13(10):e0205872, oct 2018.
- Y. Tao, Y. Zhang, and B. Jiang. Evolutionary learning-based modeling for warfarin dose prediction in Chinese. In Proceedings of the Genetic and Evolutionary Computation Conference Companion on - GECCO '17 , pages 1380–1386, New York, New York, USA, 2017. ACM Press.