UNC CHARLOTTE

College of Computing and Informatics

Department of Bioinformatics and Genomics

9201 University City Blvd, Charlotte, NC 28223-0001
t/ 704.687.2201 www.uncc.edu

November 28, 2019

Dear Editor, *F1000*

We submit this article entitled, "Ensemble Machine Learning Modeling for the Prediction of Artemisinin Resistance in Malaria" for your consideration. This original work, completed by Colby T. Ford and Daniel Janies, has not been submitted for publication elsewhere.

Resulting from our participation in the 2019 Malaria DREAM Challenge, this work explores two use cases:

1) Predict Artemisinin drug resistance levels of *Plasmodium falciparum* isolates, as quantified by the IC_{50}.
2) Predict the parasite clearance rate of malaria parasite isolates based on *in vitro* transcriptional profiles.

We develop machine learning models using novel methods for transforming isolate data and handling the tens of thousands of variables that result from these data transformation exercises. This is demonstrated by using massively parallel processing of the data vectorization for use in scalable machine learning. In addition, we show the utility of ensemble machine learning modeling for highly effective predictions of both goals of this challenge.

Moreover, we identify genes that help to predict resistance rates using a mimic black box explainer. This quantifies the feature importance of the genes as identified by the ensemble modeling exercise. By understanding these important genes, we can direct future development efforts in abating artemisinin resistance by pinpointing relevant genetic mutations in malaria.

On behalf of all authors, we thank you for considering our manuscript for publication.

Sincerely,

Colby T. Ford, Ph.D.
Department of Bioinformatics and Genomics
and Data Science Initiative
The University of North Carolina at Charlotte