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

June 8, 2020

Dear Editors, *PLOS ONE*

We submit this article entitled, "Modeling *Plasmodium falciparum* Diagnostic Test Sensitivity using Machine Learning with Histidine-Rich Protein 2 Variants" for your consideration. This original work, completed by Colby T. Ford, Gezahegn Alemayehu, Kayla Blackburn, Karen Lopez, Cheikh Cambel Dieng, Eugenia Lo, Lemu Golassa, and Daniel Janies is not under consideration for publication elsewhere.

Malaria is one of the largest and most durable health threats in the world. Over a decade ago, previous researchers modeled the sensitivity of rapid diagnostic tests using simple statistical techniques. Rapid diagnostic tests allow for quick testing of patients using only a small blood sample. However, as *Plasmodium falciparum* (the parasite that causes malaria) has mutated, rapid diagnostic test sensitivity has since diminished.

In this work, we use machine learning to model test sensitivity using specific types of genetic repeats found in the isolate sequences. We show that the previously identified types are no longer all that useful but have identified additional types that may be candidates for future iterations of rapid diagnostic tests. The utility of each type of genetic repeat was derived by measuring feature importances in the machine learning models, showcasing the innovation in using model explanability to direct real-world applications.

This article is the result of extensive work and partnership between Addis Ababa University in Ethiopia and the University of North Carolina at Charlotte in the United States, including population-level surveys of infection using rapid diagnostic tests, analysis of the pathogen through multiple observational modalities, and large-scale machine learning to help direct future testing. In the current pandemic situation, we expect that that these methodologies and analyses will be of broad interest.

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 School of Data Science
The University of North Carolina at Charlotte