Author's response to reviews

Title: Fuzzy Association Rule Mining and Classification for the Prediction of Malaria in South Korea

Authors:

Anna L Buczak (anna.buczak@jhuapl.edu)
Benjamin Baugher (benjamin.baugher@jhuapl.edu)
Erhan Guven (erhan.guven@jhuapl.edu)
Liane C Ramac-Thomas (liane.ramac-thomas@jhuapl.edu)
Yevgeniy Elbert (eugene.elbert@jhuapl.edu)
Steven M Babin (steven.babin@jhuapl.edu)
Sheri H Lewis (sheri.lewis@jhuapl.edu)

Version: 5
Date: 15 April 2015

Author's response to reviews:

Associate Editor Comment #1: I enjoyed reading the manuscript and think the method may be of interest to people working in the field of malaria prediction. However, in order to be able to assess this, I think the manuscript needs to address the issues identified below.

1. Please try to reduce number of figures, especially the maps. The first map on regions may be skipped, and the others could be merged in multi-panels.

We have completely removed Figures 1 and 4 that contained the maps.

2. Please try to explain better the model validation, add confidence intervals to spec/sens and other performance measure, and -accordingly- give more emphasis on the difference among models with respect to the performance distributions, not single point estimates.

We have added confidence intervals as requested (lower and upper bound). The model validation (called model testing in machine learning, as validation has a different meaning in machine learning), is performed on a specially set aside test data set. This data set is not used for model development or fine-tuning. In the manuscript we state explicitly which were the time periods used for model development (i.e. training), fine-tuning and testing. The testing is performed by feeding the model with the input data, and computing the metrics using equations (1) – (3). The metrics are computed using all the points in the test data set (6656).

3. Kindly specify who granted access to the data from the Korea Centers for Disease Control and Prevention.

We have addressed this in response to Reviewer One’s comment number 9. The
data are from a publicly accessible KCDC website, which we now reference. We apologize for the oversight in omitting this from the original submission.