

Age Grading Malaria Transmitting Mosquitoes Using Feed Forward

Artificial Neural Networks

*Masabho P. Milali, Benjamin Durette, Greg Merkel, Maggy T. Sikulu-Lord, Samson S. Kiware, Floyd Dowell, George F. Corliss and Richard J. Povinelli

Near infrared spectrometry (NIRS) is currently complementing techniques to age-grade mosquitoes. NIRS can classify lab-reared and semi-field raised mosquitoes into $<$ or ≥ 7 days old age groups with an average accuracy of 80%. This accuracy was achieved by training a regression model using partial least squares (PLS) and interpreting the model as a binary classifier. In this study, we explore whether using an artificial neural network (ANN) analysis instead of PLS improves the current accuracy of NIRS models on age-grading malaria transmitting mosquitoes. We also determined if directly training of a binary classifier instead of training a regression model and interpreting it as a binary classifier improves the accuracy. A total of 786 spectra of laboratory reared *Anopheles gambiae* were used and pre-processed according to previously published protocols.

Based on ten replicates, we find that training both regression and binary classification age models using ANN yields models with higher estimation accuracies than when the same models are trained using partial least squares. While ANN regression model scored root mean squared error (RMSE) and mean absolute percentage error (MAPE) of 1.6 ± 0.2 and 30.6 ± 4.7 , respectively, PLS regression model scored 3.7 ± 0.2 and 67.57 ± 6 , respectively. When we interpreted regression models as binary classifiers, the accuracy, specificity and sensitivity of ANN regression model were $93.7 \pm 1.0\%$, $95.6 \pm 1.8\%$, and $92.5 \pm 1.6\%$, respectively while PLS regression model scored $83.9 \pm 2.3\%$, $75.8 \pm 5.2\%$ and $89 \pm 2.1\%$, respectively. We also find that a directly trained binary classifier yields higher age estimation accuracy than a regression model interpreted as a binary classifier. While a directly trained ANN binary classifier scored an accuracy, sensitivity and specificity of $99.4 \pm 1.0\%$, $99.3 \pm 1.3\%$ and $99.5 \pm 0.7\%$, a directly trained PLS binary classifier scored $93.6 \pm 1.2\%$, $94.4 \pm 1.6\%$ and $92.4 \pm 1.9\%$, respectively. These results suggest the use of ANN models for age grading *An. gambiae*, however validation on larger data set and on other species is required.