1866

Age Grading Malaria Transmitting Mosquitoes Using Feed Forward Artificial Neural Networks

*Masabho P. Milali[1,2], Maggy T. Sikulu-Lord[3],Benjamin Durette[1] Samson S. Kiware[1,2],Floyd Dowell[4],George F. Corliss[1] and Richard J. Povinelli[1]
[1]Marquette University USA, [2]Ifakara Health Institute Tanzania, [3]QIMR Berghofer Medical Research Institute Australia, [4]USDA, Agricultural Research Service, Center for Grain and Animal Health Research

∗masabho.milali@marquette.edu/ pmasabho@ihi.or.tz

> Introduction and Method

Near infrared spectroscopy (NIRS) classifies lab-reared and semi-field raised mosquitoes into < or ≥ 7 days old with an average accuracy of 80%, achieved by training a regression model using partial least squares (PLS) and interpreted as a binary classifier.

Study objectives:

- We explored whether using an artificial neural network (ANN) analysis instead of PLS regression improves the current accuracy of NIRS models for age-grading malaria transmitting mosquitoes.
- We also explored if directly training a binary classifier instead of training a regression model and interpreting it as a binary classifier improves the accuracy.
- We used two-tail t-test to test the hypothesis that there is significant difference in accuracies between ANN and PLS trained model, and one-tail t-test to test the hypothesis that ANN trained model scores higher accuracies than PLS trained model.

Materials:

- A total of 786 and 870 NIR spectra collected from laboratory reared *An. gambiae* and semi-field raised *An. arabiensis*, respectively, were used and pre-processed according to previously published protocols.
- LabSpec 5000 NIR spectrometer with an integrated light source (ASD Inc., Longmont, CO), was used to collect spectra.

Model training:

- Trained both PLS regresser and binary classifier on ten-PLS components using ten-fold cross validation.
- For the ANN model, we trained a feed-forward ANN with one hidden layer, ten neurons using Levenberg-Marquardt (damped least-squares) as an optimization method. We used linear (purelin) and logistic regression functions as transfer functions when training ANN regresser and classifier, respectively.

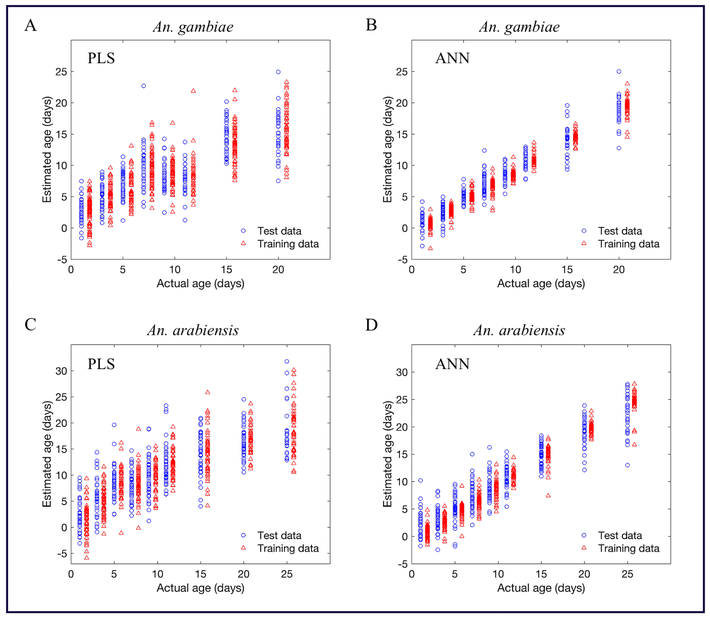
Financial support:
We thank Grand Challenge Canada, The Marquette University GasDay and Graduate school for funding the study.

Acknowledgment:
We thank Fredros O. Okumu, Sheila Ogoma and Marta F. Maia for participating in grant writing and managing of the project produced the data used in this study.

References:

- Mayagaya VS, Michel K, Benedict MQ, Killeen GF, Wirtz RA, Ferguson HM, Dowell FE. Non-destructive Determination of Age and Species of *Anopheles gambiae* sl Using Near-infrared Spectroscopy. Am J Trop Med Hyg. 2009;81(4):622-30.
- Lin M, Groves W, Freivalds A, Lee E, Harper M. Comparison of Artificial Neural Network (ANN) and Partial Least Squares (PLS) Regression Models for Predicting Respiratory Ventilation: An Exploratory Study. Eur J Appl Physiol. 2012 May;112(5):1603-11

> Results - Regressers



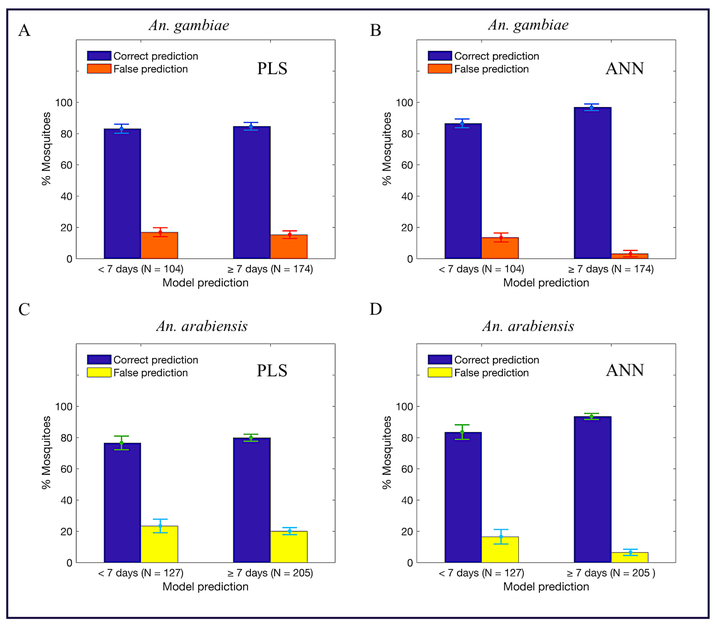


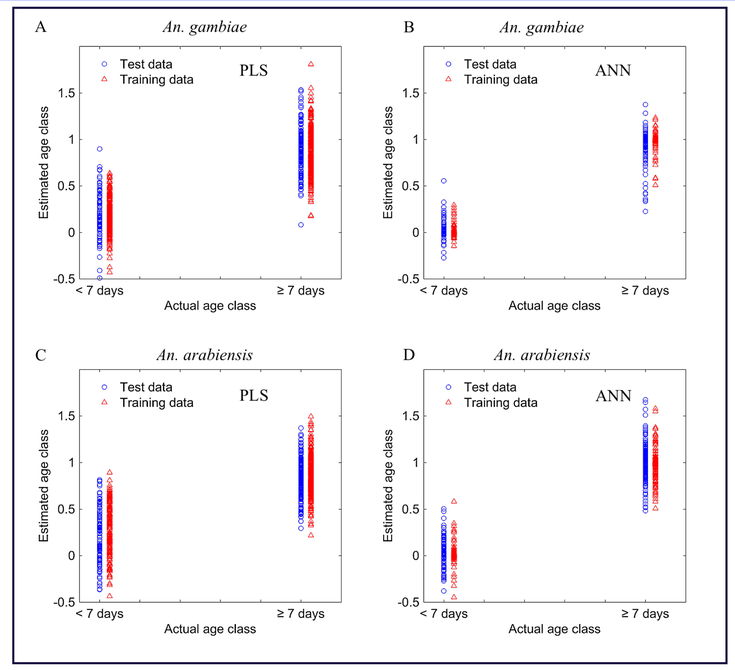
Table 1: Performance analysis of PLS and ANN regression models on estimating age of *An. gambiae* and *An. arabiensis*. Results from ten Monte Carlo cross validation.

Species	Model interpretation	Metric	Model architecture		P-value (two tail)	P-value (one tail)
			PLS	ANN		
An. gambiae	Actual age estimation	RMSE	3.7 ± 0.2	1.6 ± 0.2	3.9×10^{-9}	1.6×10^{-11}
		MAE	2.9 ± 0.2	1.2 ± 0.1	5.5×10^{-10}	7.5×10^{-12}
	Age class estimation	Accuracy (%)	83.9 ± 2.3	93.7 ± 1.0	3.6×10^{-7}	2.3×10^{-07}
		Sensitivity (%)	89.0 ± 2.1	92.5 ± 1.6	0.047	0.4696
		Specificity (%)	75.8 ± 5.2	95.6 ± 1.8	3.7×10^{-11}	1.1×10^{-06}
An. arabiensis	Actual age estimation	RMSE	4.5 ± 0.1	2.8 ± 0.2	1.7×10^{-9}	5.9×10^{-08}
		MAE	3.5 ± 0.1	2.1 ± 0.2	1.4×10^{-9}	1.4×10^{-08}
	Age class estimation	Accuracy (%)	80.3 ± 2.1	90.2 ± 1.7	1.4×10^{-7}	2.4×10^{-08}
		Sensitivity (%)	90.5 ± 1.9	91.7 ± 3.3	0.58	0.60
		Specificity (%)	60.3 ± 4.2	88.4 ± 3.9	1.7×10^{-7}	1.2×10^{-06}

Project supported by:
Grand Challenges Canada®
Grands Défis Canada

GASDAY
MARQUETTE UNIVERSITY

> Results - Classifiers



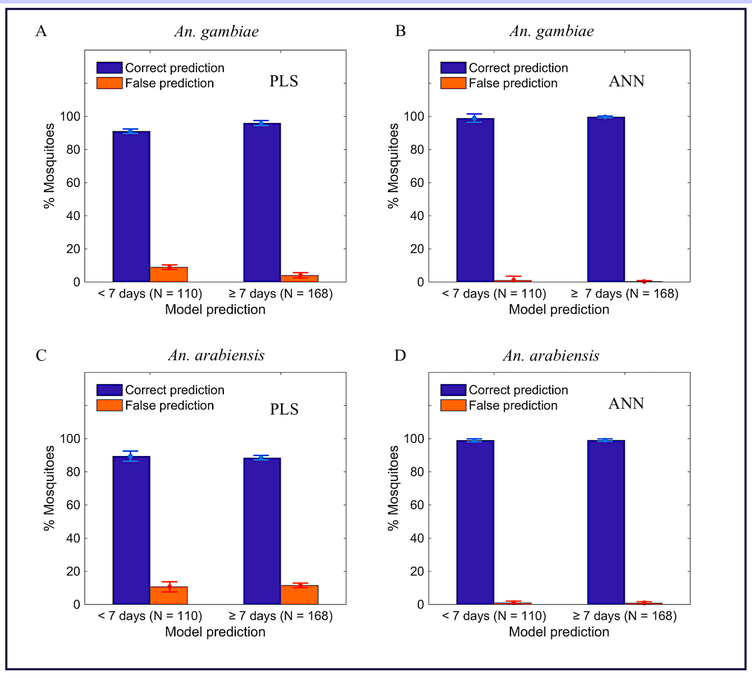


Table 2: Comparison of the accuracy of ANN and PLS classification models on ten replicates

Species	Metric	Model architecture		P-value (two-tail)	P-value (one-tail)
		PLS	ANN		
An. gambiae	Accuracy (%)	93.6 ± 1.2	99.4 ± 1.0	2.4×10^{-19}	1.2×10^{-19}
	Sensitivity (%)	94.4 ± 1.6	99.3 ± 1.4	1.6×10^{-04}	2.0×10^{-05}
	Specificity (%)	92.4 ± 1.9	99.5 ± 0.7	2.2×10^{-06}	6.0×10^{-05}
An. arabiensis	Accuracy (%)	88.7 ± 1.1	99.0 ± 0.6	1.5×10^{-21}	7.6×10^{-22}
	Sensitivity (%)	95.4 ± 1.4	99.5 ± 0.5	4.5×10^{-05}	2.3×10^{-05}
	Specificity (%)	75.2 ± 3.4	98.3 ± 1.3	4.0×10^{-09}	2.0×10^{-09}

> Conclusion

We recommend training of age models using artificial neural network and training of binary classifier instead of training regression model and interpret it as binary classifier.