RESEARCH ARTICLE

Age grading *An. gambiae* and *An. arabiensis* using near infrared spectra and artificial neural networks

Masabho P. Milali[1,2]*, Maggy T. Sikulu-Lord[3], Samson S. Kiware[1,2], Floyd E. Dowell[4], George F. Corliss[5], Richard J. Povinelli[5]

1 Ifakara Health Institute, Environmental Health and Ecological Sciences Thematic Group, Ifakara, Tanzania, **2** Department of Mathematics, Statistics and Computer Science, Marquette University, Milwaukee, Wisconsin, United States of America, **3** Queensland Alliance of Agriculture and Food Innovation, The University of Queensland, Brisbane, Queensland, Australia, **4** USDA, Agricultural Research Service, Center for Grain and Animal Health Research, Manhattan, Kansas, United States of America, **5** Department of Electrical and Computer Engineering, Marquette University, Milwaukee, Wisconsin, United States of America

* pmasabho@ihi.or.tz

Citation: Milali MP, Sikulu-Lord MT, Kiware SS, Dowell FE, Corliss GF, Povinelli RJ (2019) Age grading *An. gambiae* and *An. arabiensis* using near infrared spectra and artificial neural networks. PLoS ONE 14(8): e0209451. https://doi.org/10.1371/journal.pone.0209451

Editor: Olle Terenius, Swedish University of Agricultural Sciences, SWEDEN

Received: November 11, 2018

Accepted: July 29, 2019

Published: August 14, 2019

Data Availability Statement: All relevant data are within the paper and its Supporting Information files. Data is also freely available online at: https://github.com/masabho/Artificial-neural-network.

Funding: This study was funded by Grand Challenges Canada Stars for Global Health funded by the government of Canada grant 043901 awarded to MTSL and Marquette University Graduate School, for studentship awarded to MPM.

Abstract

Background

Near infrared spectroscopy (NIRS) is currently complementing techniques to age-grade mosquitoes. NIRS classifies lab-reared and semi-field raised mosquitoes into < or $\geq$ 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.

Methods and findings

We explore 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 explore if directly training a binary classifier instead of training a regression model and interpreting it as a binary classifier improves the accuracy. A total of 786 and 870 NIR spectra collected from laboratory reared *An. gambiae* and *An. arabiensis*, respectively, were used and pre-processed according to previously published protocols. The ANN regression model scored root mean squared error (RMSE) of 1.6 ± 0.2 for *An. gambiae* and 2.8 ± 0.2 for *An. arabiensis*; whereas the PLS regression model scored RMSE of 3.7 ± 0.2 for *An. gambiae*, and 4.5 ± 0.1 for *An. arabiensis*. When we interpreted regression models as binary classifiers, the accuracy of the ANN regression model was 93.7 ± 1.0% for *An. gambiae*, and 90.2 ± 1.7% for *An. arabiensis*; while PLS regression model scored the accuracy of 83.9 ± 2.3% for *An. gambiae*, and 80.3 ± 2.1% for *An. arabiensis*. We also find that a directly trained binary classifier yields higher age estimation accuracy than a regression model interpreted as a binary classifier. A directly trained ANN binary classifier scored an accuracy of 99.4 ± 1.0 for *An. gambiae* and 99.0 ± 0.6% for *An. arabiensis*; while a directly trained PLS binary classifier scored 93.6 ± 1.2% for *An. gambiae* and 88.7 ± 1.1% for *An. arabiensis*. We further tested the reproducibility of these results on different independent mosquito datasets. ANNs scored higher estimation accuracies than when the

Competing interests: The authors have declared that no competing interests exist.

same age models are trained using PLS. Regardless of the model architecture, directly trained binary classifiers scored higher accuracies on classifying age of mosquitoes than regression models translated as binary classifiers.

Conclusion

We recommend training models to estimate age of *An. arabiensis* and *An. gambiae* using ANN model architectures (especially for datasets with at least 70 mosquitoes per age group) and direct training of binary classifier instead of training a regression model and interpreting it as a binary classifier.

Introduction

Estimating the age of mosquitoes is one of the indicators used by entomologists for estimating vectorial capacity [1] and the effectiveness of an existing mosquito control intervention. Malaria is a vector-borne parasitic disease transmitted to people by mosquitoes of the genus *Anopheles*. The disease killed approximately 445,000 people in 2016 [2]. Mosquitoes contribute to malaria transmission by hosting and allowing the development to maturity of the malaria-causing *Plasmodium* parasite [3]. Depending on environmental temperature, *Plasmodium* takes 10–14 days in an *Anopheles* mosquito to develop fully enough to be transmitted to humans [3]. Therefore, knowing the age of a mosquito provides an indication of whether a mosquito is capable of transmitting malaria.

Knowing the age of a mosquito population is also important when evaluating the effectiveness of a mosquito control intervention. Commonly used vector control interventions such as insecticide treated nets (ITNs) and indoor residual spraying (IRS) reduce the abundance and the lifespan of a mosquito population to a level that does not support *Plasmodium* parasite development to maturity [4, 5]. Monitoring and evaluation of ITNs and IRS involves determining the age and species composition of the mosquito population before and after intervention. The presence of a small number of old mosquitoes in an area with an (ITNs or IRS) intervention indicates that the intervention is working. On the other hand, if there are more old mosquitoes, the intervention is not working effectively.

The current techniques used to estimate mosquito age are based on a combination of ovary dissecting and conventional microscopy to determine their egg laying history. Those found to have laid eggs are assumed to be older than those found to not have laid eggs [6]. This assumption can be misleading, as mosquitoes can be old but have not laid eggs and can be young (at least three days old) and have laid eggs. Dissection is laborious, difficult, and limited to only few experts. As a result, we need a new approach to address these limitations.

Different techniques such as a change in abundance of cuticular hydrocarbons [7, 8], transcriptional profiles [9, 10], and proteomics [11, 12] have been developed to age grade *Anopheles* mosquitoes. However, these techniques are still in early development stages and are limited to analyzing a small number of samples due to high analysis costs involved.

Near infrared spectroscopy (NIRS) is a complementary method to the current mosquito age grading techniques [13, 14]. NIRS is a high throughput technique, which measures the amount of the near infrared energy absorbed by samples. NIRS has been applied to identify species of insects infecting stored grains [15]; to age grade houseflies [16], stored-grain pests [17], and biting midges [18]; to differentiate between species and subspecies of termites [19]; to estimate the age and to identify species of morphologically indistinguishable laboratory reared and semi-field raised *Anopheles gambiae* and *Anopheles arabiensis* mosquitoes [13, 14, 20–23]; to estimate the age of *Aedes aegypti* mosquitoes [24]; and to detect and identify two strains of *Wolbachia pipientis* (wMelPop and wMel) in male and female laboratory-reared *Aedes aegypti* mosquitoes [25].

The current state-of-the-art of the accuracy of NIRS to classify the age of lab-reared *An. gambiae* and *An. arabiensis* is an average of 80% [13, 14, 20–23]. This accuracy is based on a trained regression model using partial least squares (PLS) and interpreted as a binary classifier to classify mosquitoes into two age groups ($<$ 7 days and $\geq$ 7 days).

In this paper, using a set of spectra collected from lab-reared and field collected *An. gambiae* and *An. arabiensis*, we explored ways to improve the reported accuracy of a PLS model for estimating age of mosquito vectors of infectious diseases. Selection of a method to train a model is one of the important factors influencing the accuracy of the model [26]. Studies [27–30] compared the accuracies of artificial neural network (ANN) and PLS regression models for predicting respiratory ventilation; explored the application of ANN and PLS to predict the changes of anthocyanins, ascorbic acid, total phenols, flavonoids, and antioxidant activity during storage of red bayberry juice; determined glucose multivariation in whole blood using partial least-squares and artificial neural networks based on mid-infrared spectroscopy; and compared modeling of nonlinear systems with artificial neural networks and partial least squares, concluding that ANN models generally perform better than PLS models. Therefore, using ANN [29–31] and PLS, we trained regression age models and compared results.

Since previous studies [13, 14, 20–23] trained a regression model and interpreted it as a binary classifier ($<$ 7 d and $\geq$ 7 d), the interpretation process may introduce errors and compromise the accuracy of the model. We further trained ANN and PLS binary classifiers and compared their accuracies with the ANN and PLS regression models translated as binary classifiers.

We find that training of both regression and binary classification models using an artificial neural network architectures yields higher accuracies than when the corresponding models are trained using partial least squares model architectures. Also, regardless of the architecture of the model, training a binary classifier yields higher age class estimation accuracy than a regression model interpreted as a binary classifier.

We then tested the reproducibility of our results by applying similar analyses on different mosquito data sets from other published studies [20, 24, 32–34], whose data are freely available for other studies to use.

Materials and methods

Ethics approval

Permission for blood feeding laboratory-reared mosquitoes was obtained from the Ifakara Health Institute (IHI) Review Board, under Ethical clearance No. IHRDC/EC4/CL.N96/2004. Oral consent was obtained from each adult volunteer involved in the study. The volunteers were given the right to refuse to participate or to withdraw from the experiment at any time.

Mosquito and spectra collection

We used spectra of *Anopheles gambiae* (IFA-GA) mosquitoes collected at 1, 3, 5, 7, 9, 11, 15, and 20 days and *An. arabiensis* (IFA-ARA) collected at 1, 3, 5, 7, 9, 11, 15, 20 and 25 days post emergence from the Ifakara Health Institute insectary. While *An. arabiensis* were reared in a semi-field system (SFS) at ambient conditions, *An. gambiae* were reared in a room made of bricks at controlled conditions. Adult mosquitoes were often provided with a human blood meal in a week and 10% glucose solution daily. Using a LabSpec 5000 NIR spectrometer with an integrated light source (ASD Inc., Longmont, CO), we followed the protocol supplied by Mayagaya and colleagues to collect spectra [13]. Prior to spectra collection, as opposed to killing by chloroform, mosquitoes were killed by freezing for 20 minutes and left to re-equilibrate to room temperature for approximately 30 minutes. A total of 786 *An*. *gambiae* and 870 *An*. *arabiensis* were scanned with at least 70 mosquitoes from each age group.

Model training

We first trained ANN and PLS regression models, scored and compared their accuracies as regressors and then as binary classifiers. We further trained binary classifiers and compared the accuracies with regressors interpreted as binary classifiers. We used a two-tail t-test to test the hypothesis that there is significant difference in accuracies between ANN and PLS trained model, a one-tail t-test to test the hypothesis that an ANN trained model scores higher accuracies than a PLS trained model.

In each species, we separately processed spectra according to Mayagaya et al., randomized, and divided processed spectra into two groups. The first group contained 70% of the total spectra and was used for training models. The second group had 30% of the total spectra and was used for out-of-sample testing.

We trained a PLS ten-component model using ten-fold cross validation [35]. Even though a range of six to ten PLS components were used in previous studies [13, 14, 20–22], we used ten PLS components after plotting the percentage of variance explained in the dependent variable against the number of PLS components (S1 Fig in the supporting information). For both species, there is not much change in the percentage variance explained in the dependent variables beyond ten components.

For the ANN model, we trained a feed-forward ANN with one hidden layer, ten neurons, and a linear transfer function (purelin) using Levenberg-Marquardt (damped least-squares) optimization [36]. We used actual mosquito ages as labels during training of both PLS and ANN regression models. We determined whether the trained models are over-fit by applying trained models (PLS and ANN) to estimate ages of mosquitoes on both training (in sample) and test (out-of-sample) data sets. Normally, if the model is not over-fit, the accuracy of the model is consistent between training and test sets [37].

The accuracies of the models were determined by computing their root mean squared error (RMSE) [38–40]. We evaluated the influence of the model architecture on the model accuracy by comparing their accuracies.

When interpreting the regression models as binary classifiers, mosquitoes with an estimated age < 7 days were considered as less than seven days old, and those ≥ 7 were considered older than or equal to seven days old. Using Eqs 1, 2 and 3, we computed and compared sensitivity, specificity, and accuracy between the PLS and ANN regression models interpreted as binary classifiers. Sensitivity of the model is the ability to classify mosquitoes correctly, which are older than or equal to seven days old (assumed to be positively related to malaria transmission), and specificity is the ability of the model to classify mosquitoes correctly which are less than seven days old (assumed to be negatively related to malaria transmission) [41–43].

$$\text{Sensitivity} = \frac{\text{Number of mosquitoes correctly predicted as} \geq 7 \text{ days old (TP)}}{\text{Total number of mosquitoes} \geq 7 \text{ days old (P)}} \tag{1}$$

$$\text{Specificity} = \frac{\text{Number of mosquitoes correctly predicted} < 7 \text{ days old (TN)}}{\text{Total number mosquitoes} < 7 \text{ days old (N)}} \tag{2}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{P} + \text{N}} \tag{3}$$

Training a regression model and interpreting it as a binary classifier can compromise the accuracy of the model as a classifier. This is because, while training a regression model forces the model to learn differences between actual ages of mosquitoes, direct training of a binary classifier forces the model to learn similarities between mosquitoes of the same class and only differences between two classes. Therefore, we directly trained binary classification models using ANN and PLS architectures and compare the accuracies with the ANN and PLS regression models interpreted as binary classifiers. In both species, we divided processed spectra (786 spectra for *An. gambiae* and 870 spectra for *An. arabiensis*) into two groups; < 7 days old and ≥ 7 days old. The spectra in a group with mosquitoes < 7 days old were labeled 0, 1 for those in a group with mosquitoes ≥ 7 days old, and the two groups were merged. The spectra were randomized and divided into training (N = 508 for both species) and test (N = 278 for *An. gambiae* and N = 362 for *An. arabiensis*) sets. We trained a PLS ten-component model using ten-fold cross-validation [35] and a one hidden layer, ten neuron feed-forward ANN using logistic regression as a transfer function and Levenberg-Marquardt (damped least-squares) optimization for training [36, 44]. During interpretation of these models, mosquitoes < 0.5 were considered as < 7 days old and ≥ 0.5 as ≥ 7 days old. Using Eqs 1, 2 and 3, for each species, we computed specificity, sensitivity, and accuracy of the trained PLS and ANN binary classifiers and compared to the PLS and ANN regressors interpreted as the binary classifiers. We repeated the process of random splitting the dataset into training and test sets; training, testing and scoring the accuracies of trained models ten times and compare the average results, a process known as Monte Carlo cross-validation [45–47].

To test reproducibility of our results, we further applied similar analysis on different data sets of mosquitoes already used in other publications but freely available for re-use [20, 24, 32–34] (S2 Fig in the supporting information). S1 and S2 Tables in the supporting information, respectively, summarize key information and number of mosquitoes per age group in each data set. Details on these data sets can be found in their respective publications.

Despite differences in characteristics (i.e., different killing methods, different scanning instruments and different sources of mosquitoes) of mosquitoes in our datasets (IFA-ARA and IFA-GA) and datasets 1–8 (S1 Table), we use datasets 7–8 and datasets 1–4 as independent test sets to test models trained on IFA-ARA and IFA-GA, respectively, (S3 Fig in the supporting information).

Here, we compare how ANN and PLS models extrapolate on datasets whose samples have different characteristics than the samples used to train them.

Results

Both PLS and ANN regression models consistently estimated the age of *An. gambiae* and *An. arabiensis* in the training and test data sets, showing that the models were likely not over-fit on these datasets during training (S4 and S5 Figs in the supporting information). Figs 1 and 2, Tables 1 and 2, and S3 Table in the supporting information present the performances of PLS and ANN regression models when estimating actual age of *An. gambiae* and *An. arabiensis* in the test data set and when their outputs are interpreted into two age classes, showing significant differences in accuracies of the two models (PLS vs ANN models). ANN regression model scores significantly higher accuracy than the PLS regression model. S4 and S5 Tables in the

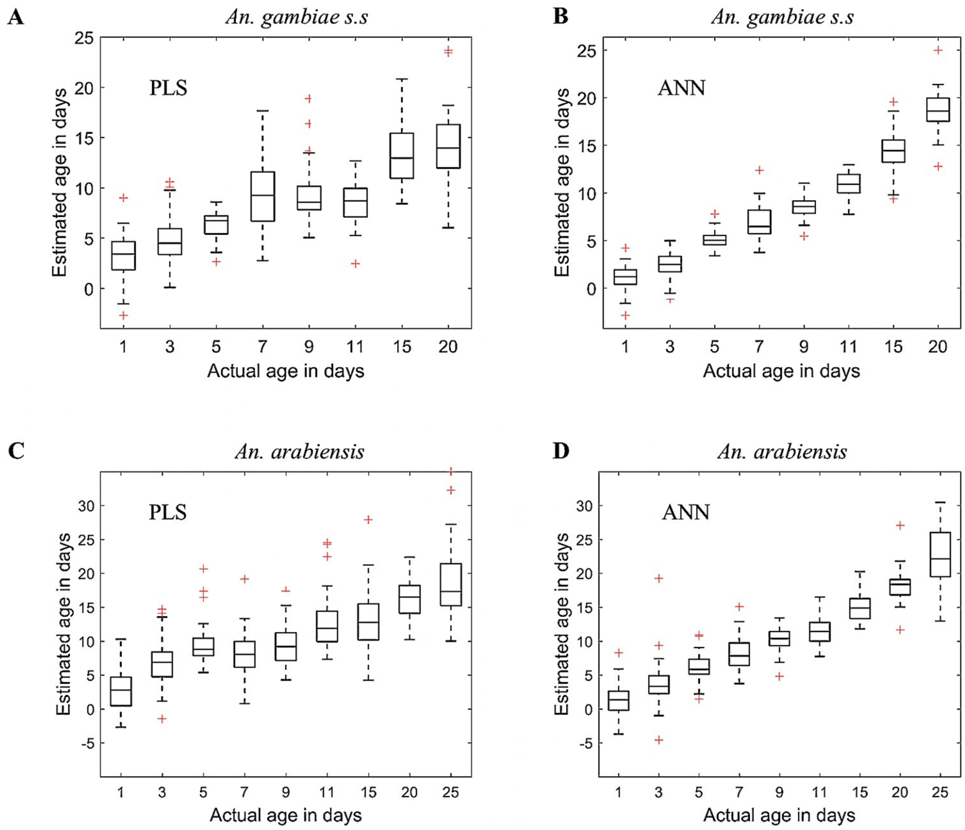

Fig 1. Box plots when PLS (A and C) and ANN (B and D) were applied to estimate the actual age of out of the sample *An. gambiae* (A and B) and *An. arabiensis* (C and D), respectively.

https://doi.org/10.1371/journal.pone.0209451.g001

supporting information represent results when the same analysis was extended to different datasets of *An. arabiensis*, *An. gambiae* s.s, *Aedes aegypti* (infected and non-infected with Wolbachia) and *Aedes albopictus* already used in other publications, showing reproducibility of the results presented in Table 1 (ANN performing better than PLS model).

S6 Fig in the supporting information represents consistency in accuracy of PLS (A and C) and ANN (B and D) directly trained binary classifiers on estimating both training and test data sets, showing that the models were likely not over-fitted during training. Figs 3 and 4 and Table 3 present the results when directly trained PLS (A and C) and ANN (B and D) binary classifiers were applied to classify ages of *An. gambiae* (A and B) and *An. arabiensis* (C and D) in test sets (out-of-sample testing), showing ANN binary classifier scores higher accuracy than the PLS binary classifier. The results further show that in both species, irrespective of the architecture used to train the model, direct training of the binary classifier scores significantly higher accuracy, specificity, and sensitivity than the regression model translated as a binary classifier (S6 Table in the supporting information). This observation was not only true to our dataset but also observed when the same analysis was applied to different datasets of mosquitoes already used in other publications [20, 24, 25, 32, 33] (S7 and S8 Tables in the supporting information).

S9 Table in the supporting information presents results when our models trained on IFA-ARA and IFA-GA were tested on an independent dataset, showing that the ANN model generally performing better than the PLS model.

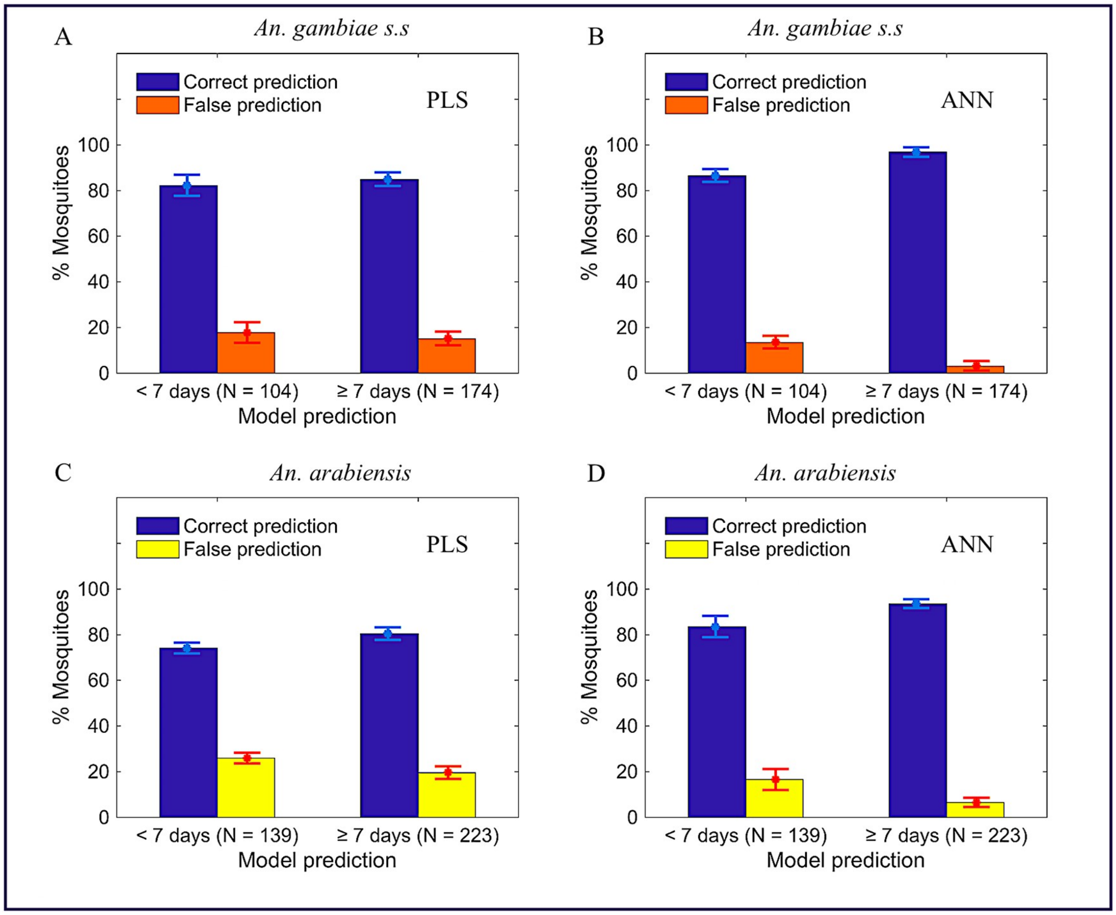

Fig 2. Number of *An. gambiae s.s* (A and B) and *An. arabiensis* (C and D) in two age classes (less than or greater/equal seven days) when PLS (A and C) and ANN (B and D) regression models, respectively, interpreted as binary classifiers.

https://doi.org/10.1371/journal.pone.0209451.g002

Discussion

This study aimed at improving the current state of the art accuracies of the models trained using near infrared spectra to estimate the age of *An. gambiae* and *An. arabiensis*. Previous studies [13, 14, 20–23] trained a regression model using partial least squares (PLS) and interpreted it as a binary classifier (< 7 d and ≥ 7 d) with an accuracy around 80%.

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

Species	Model estimation	Metric	Model architecture		P-value (two tail)	P-value (one tail)
			PLS	ANN		
An. gambiae	Actual age	RMSE	3.7 ± 0.2	1.6 ± 0.2	< 0.001	< 0.001
	Age class	Accuracy (%)	83.9 ± 2.3	93.7± 1.0	< 0.001	< 0.001
		Sensitivity (%)	89.0 ± 2.1	92.5 ± 1.6	0.005	0.047
		Specificity (%)	75.8 ± 5.2	95.6 ± 1.8	< 0.001	< 0.001
An. arabiensis	Actual age	RMSE	4.5 ± 0.1	2.8 ± 0.2	< 0.001	< 0.001
	Age class	Accuracy (%)	80.3 ± 2.1	90.2 ± 1.7	< 0.001	< 0.001
		Sensitivity (%)	90.5 ± 1.9	91.7 ± 3.3	0.58	0.60
		Specificity (%)	60.3 ± 4.2	88.4 ± 3.9	< 0.001	< 0.001

https://doi.org/10.1371/journal.pone.0209451.t001

Table 2. Mean actual age estimation of mosquitoes in out of the sample test sets by ANN and PLS regression models. Column "N" represents the number of mosquitoes in each age group.

Actual age	Model Prediction					
	An. arabiensis			*An. gambiae s.s*		
	PLS	N	ANN	PLS	N	ANN
1	1.9 ± 3.2	43	1.3 ± 2.5	2.4 ± 2.8	29	1.0 ± 1.4
3	5.8 ± 3.9	40	3.7 ± 3.5	5.0 ± 2.2	45	2.4 ± 1.3
5	9.3 ± 3.3	39	6.1 ± 2.1	6.5 ± 2.1	35	5.0 ± 0.9
7	8.7 ± 2.9	47	8.1 ± 2.4	10.5 ± 3.3	41	6.9 ± 1.7
9	9.9 ± 3.7	35	10.2 ± 1.7	9.2 ± 2.5	35	8.5 ± 1.2
11	12.2 ± 3.4	45	11.5 ± 1.8	8.7 ± 3.9	29	10.8 ± 1.3
15	13.6 ± 4.3	37	14.9 ± 1.9	13.6 ± 3.3	36	14.3 ± 2.2
20	17.3 ± 3.4	38	18.2 ± 2.4	15.8 ± 3.6	28	18.6 ± 2.3
25	19.9 ± 6.7	38	23.2 ± 6.4			

https://doi.org/10.1371/journal.pone.0209451.t002

Knowing that the selection of a model architecture often influences the model accuracy [26], we trained age regression models using an artificial neural network [29–31, 48, 49] and partial least squares as model architectures and compared the accuracies. ANN models achieved significantly higher accuracies than corresponding PLS regression models. As summarized in Table 1, ANN regression models scored an average RMSE of 1.60 ± 0.18 for *An. gambiae* and 2.81 ± 0.22 for *An. arabiensis*. The PLS regression models scored RMSE of 3.66 ± 0.23 for *An. gambiae* and 4.49 ± 0.09 for *An. arabiensis*. When both ANN and PLS regression models were interpreted as binary classifiers, ANN regression model scored accuracy, sensitivity, and specificity of 93.71 ± 1.03%, 92.54 ± 1.60%, and 95.64 ± 1.82%, respectively, for *An. gambiae*; 90.16 ± 1.70%, 91.68 ± 3.27% and 88.44 ± 3.86%, respectively, for *An. arabiensis*. The PLS regression model scored accuracy, sensitivity, and specificity of 83.85 ± 2.32%, 89.00 ± 2.10%, and 75.82 ± 5.22%, respectively, for *An. gambiae*; 80.30 ± 2.06%, 90.48 ± 1.88%, and 60.25 ± 4.20%, respectively, for *An. arabiensis*.

The interpretation of a regression model as a binary classifier can introduce errors that compromise the accuracy of the model. We directly trained PLS and ANN binary classifiers and compared the accuracies with ANN and PLS regression models interpreted as binary classifiers. Irrespective of the model architecture, directly trained binary classifiers scored significantly higher accuracies than corresponding regression models interpreted as binary classifiers (S6 Table in the supporting information). The explanation of these results could be that training a regression model and interpreting it as a binary classifier involved learning differences between multiple age groups (1, 3, 5, 7, 9, 11, 13, 15, and 20 days old for *An. gambiae* and 1, 3, 5, 7, 9, 11, 13, 15, 20 and 25 days for *An. arabiensis*) of mosquitoes, which can be challenging for two consecutive age groups. In contrast, direct training of the binary classifier involved learning differences existing between only two age groups. During direct training of the binary classifier, the process of dividing spectra into two groups (< 7 or ≥ 7 days) forced a model to learn similarities instead of differences between mosquitoes of the same age class. We also observed that directly trained ANN binary classifier scored higher accuracy than directly trained PLS binary classifier. ANN binary classifier scored an accuracy, sensitivity, and specificity of 99.4 ± 1.0%, 99.3 ± 1.4%, and 99.5 ± 0.7%, respectively, for *An. gambiae*; 99.0 ± 0.6%, 99.5 ± 0.5%, and 98.3 ± 1.3%, respectively, for *An. arabiensis*. The PLS binary classifier scored 93.6 ± 1.2%, 94.4 ± 1.6%, and 92.5 ± 1.9% for *An. gambiae*; 88.7 ± 1.1%, 95.5 ± 1.4%, and 75.2 ± 3.5% for *An. arabiensis* (Table 3).

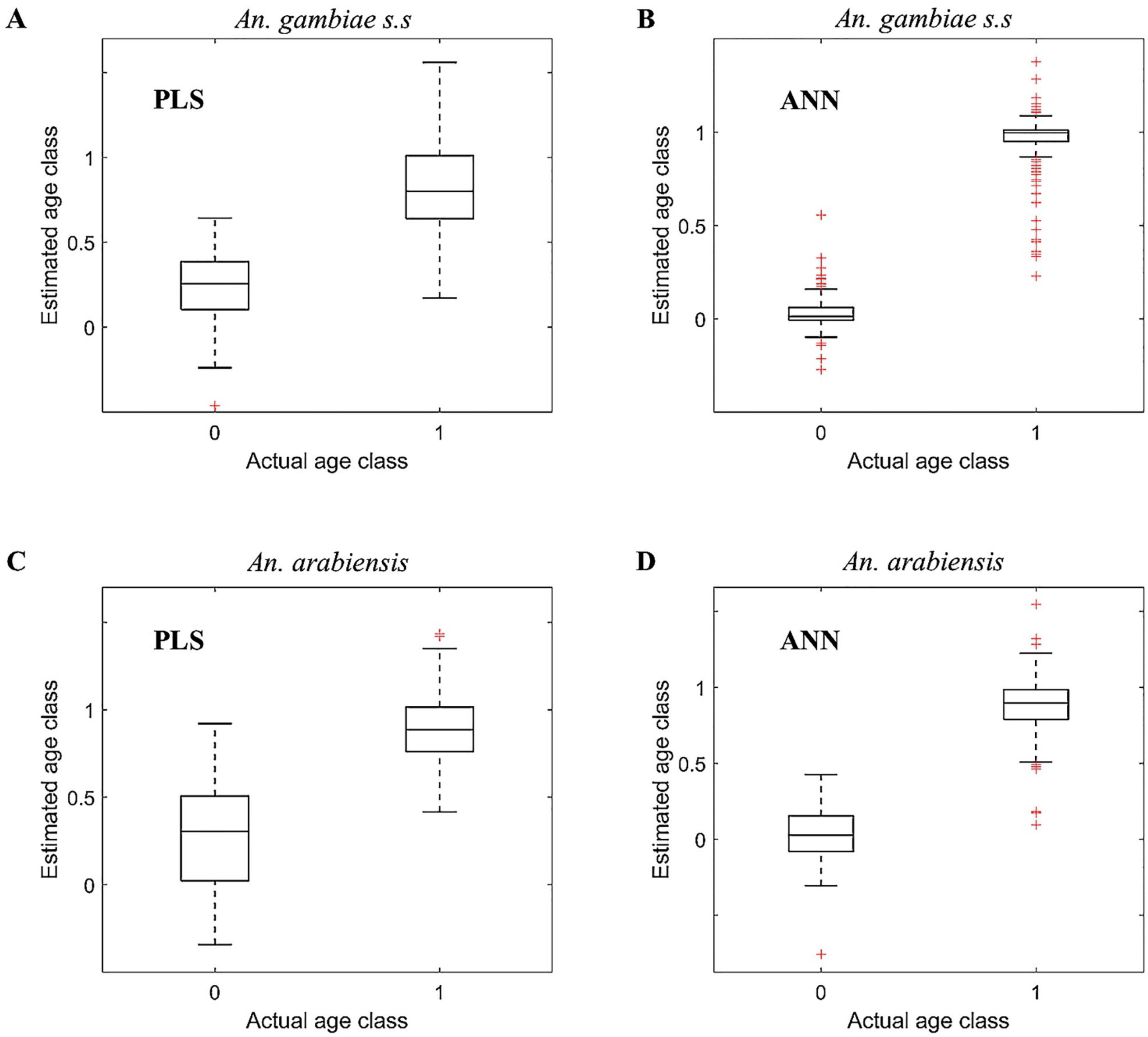

Fig 3. Box plot of directly trained PLS (A and C) and ANN (B and D) binary classifiers for estimating age classes of *An.gambiae* (A and B) *and An. arabiensis* (C and D) in out of sample testing sets.

https://doi.org/10.1371/journal.pone.0209451.g003

Reproducibility of results is one of the key components when testing precision and accuracy of a new measurement or method [50]. We further tested the reproducibility of our analyses on different datasets of *An. gambiae*, *An. arabiensis*, *Aedes aegypti* (males and females infected and not infected with *Wolbachia*) and *Aedes albopictus*, which are already published and freely available for re-use in other studies [20, 24, 32–34]. We found consistency in results between our datasets and different datasets of mosquitoes already published in other studies (S4, S5, S7 and S8 Tables in the supporting information). This consistency strengthens the assertion that ANN models score higher accuracy than PLS models.

Our study is not the first to observe ANN models outperforming PLS models. Despite being reproducible in different datasets, these findings are also supported with other previous studies [27–29, 31] compared the accuracies of ANN and PLS models, where they report ANN

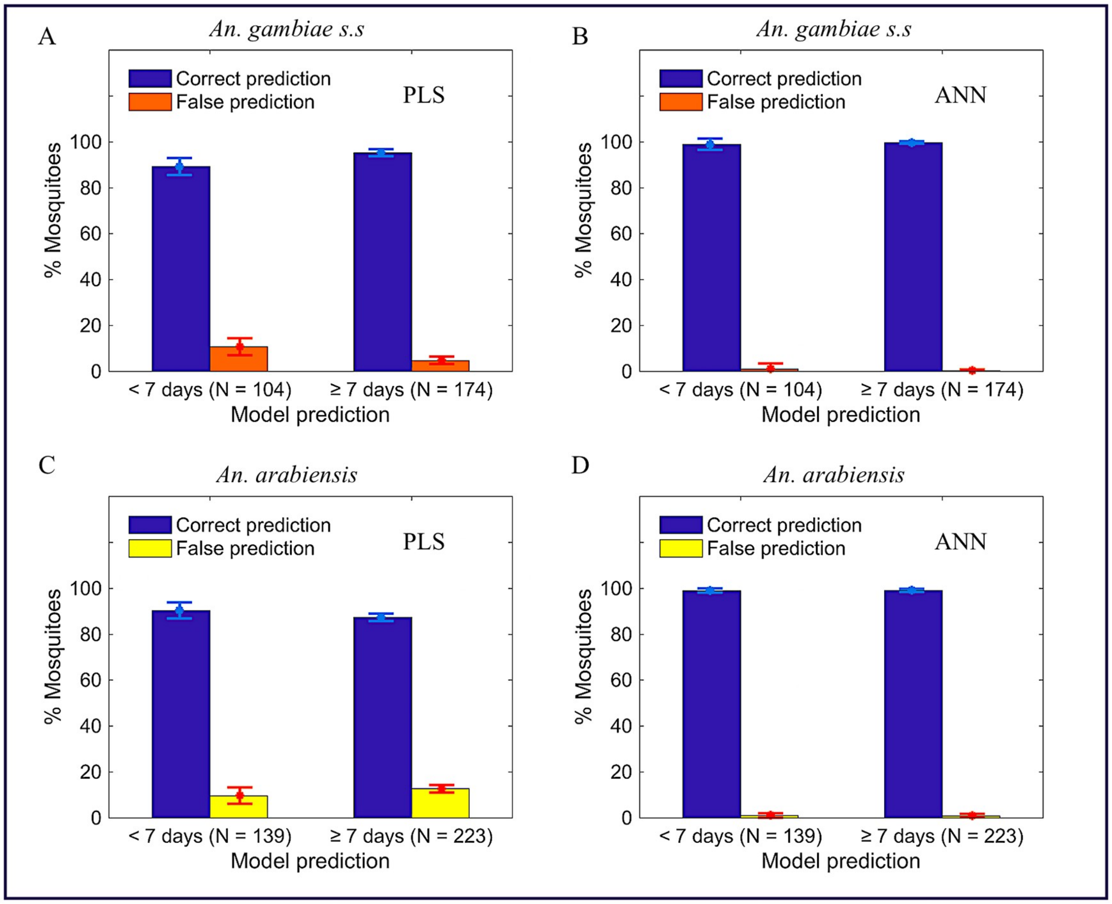

Fig 4. The number of correct and false predictions in each estimated age-class when directly trained PLS (A and C) and ANN (B and D) binary classifiers were applied to classify age of *An. gambiae* (A and B) and *An. arabiensis* (C and D) in testing sets. Results from ten replicates.

https://doi.org/10.1371/journal.pone.0209451.g004

perform better than PLS. The explanation of these results could be that ANN, unlike PLS, considers both linear and unknown non-linear relationships between dependent and independent variables [29–31]; builds independent-dependent relationships that interpolates well even to cases that were not exactly presented by training data; and has a self mechanism of filtering and handling noisy data during training [48, 49]. Hence, ANN models are unbiased estimators in contrast to PLS models (Fig 5 and S7 Fig in the supporting information).

Table 3. 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	< 0.001	< 0.001
	Sensitivity (%)	94.4 ± 1.6	99.3 ± 1.4	< 0.001	< 0.001
	Specificity (%)	92.4 ± 1.9	99.5 ± 0.7	< 0.001	< 0.001
An. arabiensis	Accuracy (%)	88.7 ± 1.1	99.0 ± 0.6	< 0.001	< 0.001
	Sensitivity (%)	95.4 ± 1.4	99.5 ± 0.5	< 0.001	< 0.001
	Specificity (%)	75.2 ± 3.4	98.3 ± 1.3	< 0.001	< 0.001

https://doi.org/10.1371/journal.pone.0209451.t003

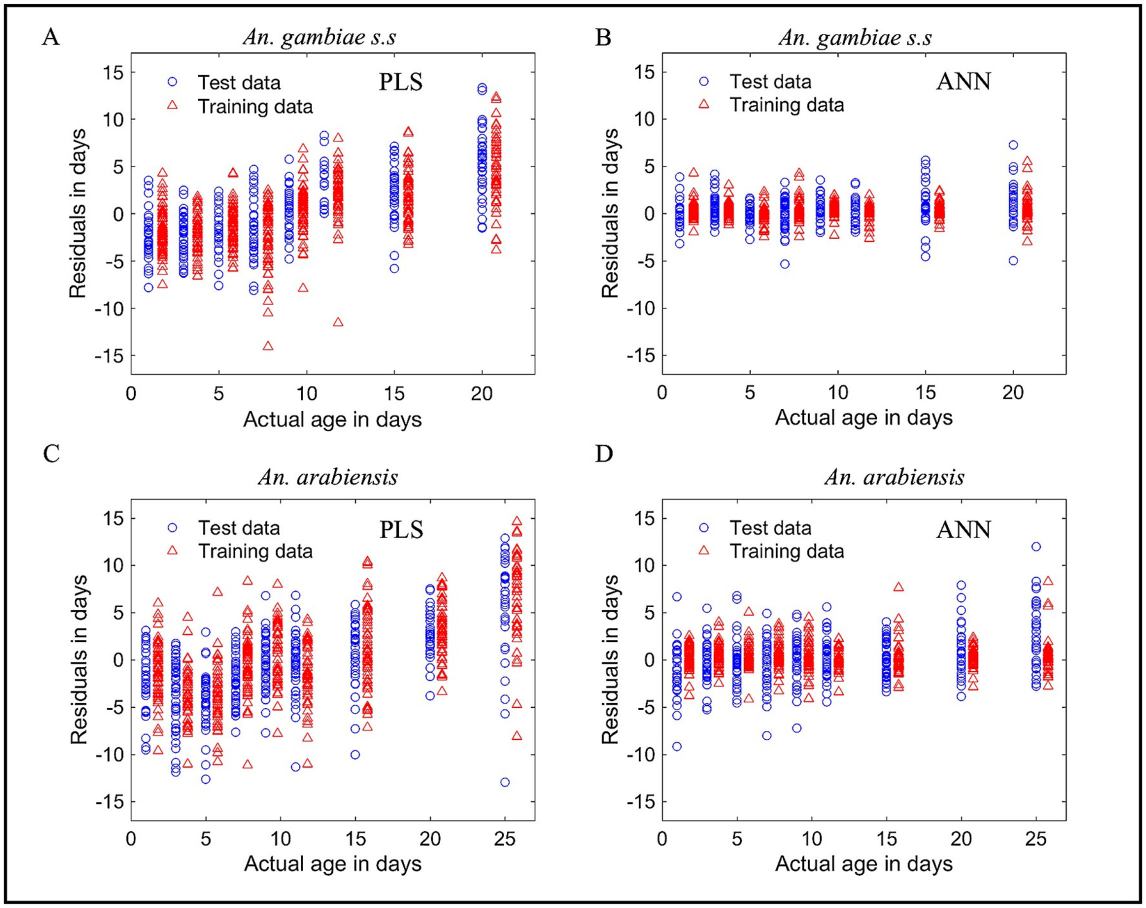

Fig 5. Error distribution per actual age of *An. gambiae* and *An. arabiensis* when ANN and PLS regressors applied to estimate the actual ages of mosquitoes in training and test data sets, showing a uniform distribution of errors (un-biased estimating) across actual ages of mosquitoes for the ANN regressor and an un-uniform distribution of errors (biased estimating) for the PLS regressor.

https://doi.org/10.1371/journal.pone.0209451.g005

We also found that ANN model extrapolates better than PLS model when tested on datasets whose samples have different characteristics than the samples used to train them (S9 Table in supporting information). These results strengthen the assertion that ANNs can filter and handle noisy data better than PLS models. Furthermore, these results suggest that training neural networks on samples with varying characteristics such as different killing methods, scanning instruments, and geographical regions, might yield a model with better performance than the one presented in S9 Table in supporting information. The only caveat with this is a need for large dataset to train the model.

Conclusion

We conclude that training both regression and binary classification age artificial neural network models yield higher accuracies than partial least squares models. Also, training a binary classifier scores higher accuracy than training a regression model and interpreting it as a binary classifier. Hence, 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.

Supporting information

S1 Fig. The percentage of variance explained in the dependent variable against the number of PLS components: A) *An. gambiae* B) *An. arabiensis*.
(TIF)

S2 Fig. Illustration on how we reproduced our analysis on different datasets.
(TIFF)

S3 Fig. Illustration on how ANN and PLS models trained on IFA-ARA and IFA-GA datasets were tested on independent datasets.
(TIFF)

S4 Fig. PLS (A and C) and ANN (B and D) regression models, estimating actual age of training and testing samples of *An. gambiae* (A and B) and *An. arabiensis* (C and D), respectively.
(TIF)

S5 Fig. Regression coefficients weights against wavelengths: A) *An. gambiae* B) *An. arabiensis*.
(TIF)

S6 Fig. The consistency in accuracies of directly trained PLS (A and C) and ANN (B and D) binary classifiers for estimating age classes of *An.gambiae* (A and B) and *An. arabiensis* (C and D) in both training and testing sets.
(TIF)

S7 Fig. Error distribution per actual age class of *An. gambiae* and *An. arabiensis* when directly trained ANN and PLS binary classifiers applied to estimate age classes of mosquitoes in training and test data sets, showing uniform distribution of errors (un-biased estimating) across actual age classes of mosquitoes for ANN binary classifiers and un-uniform (biased estimating) distribution for PLS classifiers.
(TIF)

S1 Table. List and summary of mosquito datasets used to test reproducibility of our study. Numbers in brackets are references of the studies where dataset is originally published.
(DOCX)

S2 Table. Number of mosquitoes per age group in each dataset used to test reproducibility of our study.
(DOCX)

S3 Table. Percentage of mosquitoes in each age group correctly classified when ANN and PLS regression models are interpreted as binary classifiers.
(DOCX)

S4 Table. Reproducibility analysis of PLS and ANN regression models on estimating age of *An. gambiae* and *An. arabiensis* in different datasets already used in other publications. Results from ten-fold Monte Carlo cross-validation.
(DOCX)

S5 Table. Performance analysis of PLS and ANN regression models on estimating age of *Aedes albopictus*, *Wolbachia* free and *Wolbachia* infected male and female *Aedes aegypti*. Results from ten-fold Monte Carlo cross-validation.
(DOCX)

S6 Table. Comparison of accuracies between directly trained binary classifiers and regressers interpreted as binary classifiers. Results from ten-fold Monte Carlo cross-validation.
(DOCX)

S7 Table. Comparison of the accuracy of directly trained ANN and PLS classification models on *An. gambiae* and *An. arabiensis* in datasets from other published studies.
(DOCX)

S8 Table. Comparison of the accuracies of directly trained ANN and PLS classification models on *Aedes aegypti* and *Aedes albopictus* in datasets from other published studies.
(DOCX)

S9 Table. Results when both regression and directly trained binary classifiers trained on IFA-GA and IFA-ARA datasets were tested on independent test sets.
(DOCX)

S1 Appendix. Excel file with IFA-GA data. Column header, wavelengths in 'nm'.
(XLSX)

S2 Appendix. Excel file with IFA-ARA. Column header, wavelengths in 'nm'.
(XLSX)

S3 Appendix. Matlab code used to run the analysis.
(M)

S4 Appendix. Matlab code used to pre-process spectra.
(M)

S5 Appendix. Zip folder with data used to test reproducibility of our study.
(ZIP)

S6 Appendix. Zip folder with boxplots generated after performing reproducibility analysis of PLS and ANN regression models on estimating age of *An. gambiae* and *An. arabiensis* in different datasets already used in other publications.
(ZIP)

S7 Appendix. Zip folder with boxplots generated after analysis of PLS and ANN regression models on estimating age of *Aedes albopictus*, *Wolbachia* free and *Wolbachia* infected male and female *Aedes aegypti*.
(ZIP)

S8 Appendix. Zip folder with boxplots generated when we directly trained ANN and PLS classification models on *An. gambiae* and *An. arabiensis* in datasets from other published studies.
(ZIP)

S9 Appendix. Zip folder with boxplots generated when we directly trained ANN and PLS classification models on *Aedes aegypti* and *Aedes albopictus* in datasets from other published studies.
(ZIP)

S10 Appendix. Zip folder with boxplots after both regression and directly trained binary classifiers trained on IFA-GA and IFA-ARA datasets were tested on independent test sets.
(ZIP)

S11 Appendix. Zip folder with boxplots and a table with accuracies generated when our models trained on datasets DS1—DS6 and IFA-GA were applied on independent test sets (nulliparous vs sporozoite positive field samples) as presented by Krajacich et al. 2017. (ZIP)

Acknowledgments

We thank Andrew Kafwenji and Paulina Kasanga for help maintaining the mosquito colony, Marta F. Maia, Fredros O. Okumu, and Sheila Ogoma for participating in grant writing and managing of the project that produced the data used in this manuscript, and the USDA, Agricultural Research Service, Center for Grain and Animal Health Research, USA for loaning us the near-infrared spectrometer used to scan the mosquitoes. We also thank Michael Henry and Nikita Lysenko for helping with mosquito scanning to collect spectra in Tanzania; Alex Ntamatungiro and Benjamin Krajacich for allowing us to use their datasets to test reproducibility of our study; and Gustav Mkandawile who worked tirelessly to make sure we obtained mosquitoes and Ben Durette who participated in the initial stages of data analysis.

Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity provider and employer.

Author Contributions

Conceptualization: Masabho P. Milali, George F. Corliss, Richard J. Povinelli.

Data curation: Masabho P. Milali.

Formal analysis: Masabho P. Milali.

Funding acquisition: Masabho P. Milali, Maggy T. Sikulu-Lord.

Investigation: Masabho P. Milali.

Methodology: Masabho P. Milali, George F. Corliss, Richard J. Povinelli.

Project administration: Masabho P. Milali, Maggy T. Sikulu-Lord.

Resources: Floyd E. Dowell.

Software: Masabho P. Milali.

Supervision: Maggy T. Sikulu-Lord, Samson S. Kiware, George F. Corliss, Richard J. Povinelli.

Validation: Maggy T. Sikulu-Lord, Samson S. Kiware, Richard J. Povinelli.

Visualization: Richard J. Povinelli.

Writing – original draft: Masabho P. Milali.

Writing – review & editing: Masabho P. Milali, Maggy T. Sikulu-Lord, Samson S. Kiware, Floyd E. Dowell, George F. Corliss, Richard J. Povinelli.

References

1. Dye C. The Analysis of Parasite Transmission by Bloodsucking Insects. Annu Rev Entomol. 1992; 37 (1):1–19.
2. WHO. World Malaria Report. Geneva, World Health Organization. 2017.
3. Beier JC. Malaria Parasite Development in Mosquitoes. Annu Rev Entomol. 1998; 43(1):519–43.
4. Robert V, Carnevale P. Influence of Deltamethrin Treatment of Bed Nets on Malaria Transmission in the Kou Valley, Burkina Faso. Bull World Health Organ. 1991; 69(6):735. PMID: 1786622

5. Magesa SM, Wilkes TJ, Mnzava A, Njunwa KJ, Myamba J, Kivuyo MDP, et al. Trial of Pyrethroid Impregnated Bednets in an Area of Tanzania Holoendemic for Malaria Part 2. Effects on the Malaria Vector Population. Acta Trop. 1991; 49(2):97–108. PMID: 1680284
6. Detinova TS. Age Grouping Methods in Diptera of Medical Importance with Special Reference to Some Vectors of Malaria. Monogr Ser World Health Organization. 1962; 47:13–108.
7. Hugo LE, Kay BH, Eaglesham GK, Holling N, Ryan PA. Investigation of Cuticular Hydrocarbons for Determining the Age and Survivorship of Australasian Mosquitoes. American Journal of Tropical Medicine and Hygiene. 2006; 74(3):462–74. PMID: 16525108
8. Caputo B, Dani FR, Horne GL, Petrarca V, Turillazzi S, Coluzzi M, et al. Identification and Composition of Cuticular Hydrocarbons of the Major Afrotropical Malaria Vector *Anopheles gambiae ss* (Diptera: *Culicidae*): Analysis of Sexual Dimorphism and Age-related Changes. Journal of Mass Spectrometry. 2005; 40(12):1595–604. https://doi.org/10.1002/jms.961 PMID: 16320293
9. Wang M, Marinotti O, James AA, Walker E, Githure J, Yan G. Genome-wide Patterns of Gene Expression During Aging in the African Malaria Vector *Anopheles gambiae*. PLoS One. 2010; 5(10):e13359. https://doi.org/10.1371/journal.pone.0013359 PMID: 20967211
10. Cook PE, Sinkins SP. Transcriptional Profiling of *Anopheles gambiae* Mosquitoes for Adult Age Estimation. Insect Mol Biol. 2010; 19(6):745–51. https://doi.org/10.1111/j.1365-2583.2010.01034.x PMID: 20695922
11. Sikulu MT, Monkman J, Dave KA, Hastie ML, Dale PE, Kitching RL, et al. Proteomic Changes Occurring in the Malaria Mosquitoes *Anopheles gambiae* and *Anopheles stephensi* During Aging. Journal of Proteomics. 2015; 126:234–44. https://doi.org/10.1016/j.jprot.2015.06.008 PMID: 26100052
12. Sikulu MT, Monkman J, Dave KA, Hastie ML, Dale PE, Kitching RL, et al. Mass Spectrometry Identification of Age-associated Proteins from the Malaria Mosquitoes *Anopheles gambiae ss* and *Anopheles stephensi*. Data in brief. 2015; 4:461–7. https://doi.org/10.1016/j.dib.2015.07.007 PMID: 26306320
13. Mayagaya VS, Michel K, Benedict MQ, Killeen GF, Wirtz RA, Ferguson HM, et al. Non-destructive Determination of Age and Species of *Anopheles gambiae sl* Using Near-infrared Spectroscopy. American Journal of Tropical Medicine and Hygiene. 2009; 81(4):622–30. https://doi.org/10.4269/ajtmh.2009.09-0192 PMID: 19815877
14. Sikulu M, Killeen GF, Hugo LE, Ryan PA, Dowell KM, Wirtz RA, et al. Near-infrared Spectroscopy as a Complementary Age Grading and Species Identification Tool for African Malaria Vectors. Parasites & Vectors. 2010; 3(1):1.
15. Dowell FE, Throne JE, Wang D, Baker JE. Identifying Stored-grain Insects Using Near-infrared Spectroscopy. Journal of Economic Entomology. 1999; 92(1):165–9.
16. Perez-Mendoza J, Dowell FE, Broce AB, Throne JE, Wirtz RA, Xie F, et al. Chronological Age-grading of House Flies by Using Near-infrared Spectroscopy. Journal of Medical Entomology. 2002; 39(3):499–508. https://doi.org/10.1603/0022-2585-39.3.499 PMID: 12061447
17. Perez-Mendoza J, Throne JE, Dowell FE, Baker JE. Chronological Age-grading of Three Species of Stored-product Beetles by Using Near-infrared Spectroscopy. Journal of Economic Entomology. 2004; 97(3):1159–67. https://doi.org/10.1603/0022-0493(2004)097[1159:caotso]2.0.co;2 PMID: 15279305
18. Reeves WK, Peiris K, Scholte E, Wirtz RA, Dowell FE. Age-grading the Biting Midge *Culicoides sonorensis* Using Near-infrared Spectroscopy. Med Vet Entomol. 2010; 24(1):32–7. https://doi.org/10.1111/j.1365-2915.2009.00843.x PMID: 20377729
19. Aldrich BT, Maghirang EB, Dowell FE, Kambhampati S. Identification of Termite Species and Subspecies of the Genus *Zootermopsis* Using Near-infrared Reflectance Spectroscopy. J Insect Sci. 2007; 7(1):18.
20. Sikulu MT, Majambere S, Khatib BO, Ali AS, Hugo LE, Dowell FE. Using a Near-infrared Spectrometer to Estimate the Age of *Anopheles* Mosquitoes Exposed to Pyrethroids. PloS one. 2014; 9(3):e90657. https://doi.org/10.1371/journal.pone.0090657 PMID: 24594705
21. Sikulu M, Dowell KM, Hugo LE, Wirtz RA, Michel K, Peiris KH, et al. Evaluating RNA Later® as a Preservative for Using Near-infrared Spectroscopy to Predict *Anopheles gambiae* Age and Species. Malaria Journal. 2011; 10(1):186.
22. Dowell FE, Noutcha AE, Michel K. The Effect of Preservation Methods on Predicting Mosquito Age by Near-infrared Spectroscopy. American Journal of Tropical Medicine and Hygiene. 2011; 85(6):1093–6.
23. Mayagaya VS, Ntamatungiro AJ, Moore SJ, Wirtz RA, Dowell FE, Maia MF. Evaluating Preservation Methods for Identifying *Anopheles gambiae ss* and *Anopheles arabiensis* Complex Mosquitoes Species Using Near-infrared Spectroscopy. Parasites & Vectors. 2015; 8(1):60.
24. Sikulu-Lord MT, Milali MP, Henry M, Wirtz RA, Hugo LE, Dowell FE, et al. Near-infrared Spectroscopy, a Rapid Method for Predicting the Age of Male and Female Wild-Type and Wolbachia Infected *Aedes*

aegypti. PLoS Negl Trop Dis. 2016; 10(10):e0005040. https://doi.org/10.1371/journal.pntd.0005040 PMID: 27768689

25. Sikulu-Lord MT, Maia MF, Milali MP, Henry M, Mkandawile G, Kho EA, et al. Rapid and Non-destructive Detection and Identification of two Strains of Wolbachia in *Aedes aegypti* by Near-infrared Spectroscopy. PLoS Negl Trop Dis. 2016; 10(6):e0004759. https://doi.org/10.1371/journal.pntd.0004759 PMID: 27362709
26. Mouazen AM, Kuang B, De Baerdemaeker J, Ramon H. Comparison Among Principal Component, Partial Least Squares and Back Propagation Neural Network Analyses for Accuracy of Measurement of Selected Soil Properties with Visible and Near-infrared Spectroscopy. Geoderma. 2010; 158(1):23–31.
27. 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. https://doi.org/10.1007/s00421-011-2118-6 PMID: 21861111
28. Zheng H, Jiang L, Lou H, Hu Y, Kong X, Lu H. Application of Artificial Neural Network (ANN) and Partial Least-squares Regression (PLSR) to Predict the Changes of Anthocyanins, Ascorbic Acid, Total Phenols, Flavonoids, and Antioxidant Activity During Storage of Red Bayberry Juice Based on Fractal Analysis and Red, Green, and Blue (RGB) Intensity Values. Journal of Agricultural and Food Chemistry. 2011 Jan 26,; 59(2):592. https://doi.org/10.1021/jf1032476 PMID: 21190362
29. Bhandare P, Mendelson Y, Peura RA, Janatsch G, Kruse-Jarres JD, Marbach R, et al. Multivariate Determination of Glucose in Whole Blood Using Partial Least-squares and Artificial Neural Networks Based on Mid-infrared Spectroscopy. Appl Spectrosc. 1993; 47(8):1214–21.
30. Khotanzad A, Elragal H, Lu T. Combination of Artificial Neural-network Forecasters for Prediction of Natural Gas Consumption. IEEE Trans Neural Networks. 2000; 11(2):464–73. https://doi.org/10.1109/72.839015 PMID: 18249775
31. Hadjiiski L, Geladi P, Hopke P. A Comparison of Modeling Nonlinear Systems with Artificial Neural Networks and Partial Least Squares. Chemometrics Intellig Lab Syst. 1999; 49(1):91–103.
32. Ntamatungiro AJ, Mayagaya VS, Rieben S, Moore SJ, Dowell FE, Maia MF. The Influence of Physiological Status on Age Prediction of *Anopheles arabiensis* Using Near-infrared Spectroscopy. Parasites & vectors. 2013; 6(1):1.
33. Krajacich BJ, Meyers JI, Alout H, Dabire RK, Dowell FE, Foy BD. Analysis of Near-infrared Spectra for Age-grading of Wild Populations of *Anopheles gambiae*. Parasites & Vectors. 2017 Jan 1,; 10(1):1–13.
34. Sikulu-Lord MT, Devine GJ, Hugo LE, Dowell FE. First Report on the Application of Near-infrared Spectroscopy to Predict the Age of *Aedes albopictus* Skuse. Scientific Reports. 2018; 8(1):9590. https://doi.org/10.1038/s41598-018-27998-7 PMID: 29941924
35. Ståhle L, Wold S. Partial Least Squares Analysis with Cross-validation for the Two-class Problem: A Monte Carlo study. J Chemometrics. 1987; 1(3):185–96.
36. Ballabio D, Consonni V, Todeschini R. The Kohonen and CP-ANN Toolbox: A Collection of MATLAB Modules for Self Organizing Maps and Counterpropagation Artificial Neural Networks. Chemometrics Intellig Lab Syst. 2009; 98(2):115–22.
37. Cawley GC, Talbot NL. On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. Journal of Machine Learning Research. 2010; 11(Jul):2079–107.
38. Chai T, Draxler RR. Root Mean Square Error (RMSE) or Mean Absolute Error (MAE)?–Arguments Against Avoiding RMSE in the Literature. Geoscientific Model Development. 2014; 7(3):1247–50.
39. Willmott CJ, Matsuura K. Advantages of the Mean Absolute Error (MAE) Over the Root Mean Square Error (RMSE) in Assessing Average Model Performance. Climate Research. 2005; 30(1):79–82.
40. Hyndman RJ, Koehler AB. Another Look at Measures of Forecast Accuracy. Int J Forecast. 2006; 22 (4):679–88.
41. Altman DG, Bland JM. Statistics Notes: Diagnostic Tests 1: Sensitivity and Specificity. BMJ. 1994 Jun 11,; 308(6943):1552.
42. Smith C. Diagnostic Tests (1)–Sensitivity and Specificity. Phlebology. 2012 Aug; 27(5): 250–1. https://doi.org/10.1258/phleb.2012.012J05 PMID: 22956651
43. Lalkhen AG, McCluskey A. Clinical Tests: Sensitivity and Specificity. Continuing Education in Anaesthesia, Critical Care & Pain. 2008 Dec; 8(6): 221–3.
44. Moré JJ. The Levenberg-Marquardt Algorithm: Implementation and Theory. In Numerical Analysis. Springer, Berlin, Heidelberg. 1978 (pp. 105–116).
45. Xu Q, Liang Y. Monte Carlo Cross Validation. Chemometrics Intellig Lab Syst. 2001; 56(1):1–11.
46. Xu Q, Liang Y, Du Y. Monte Carlo Cross-validation for Selecting a Model and Estimating the Prediction Error in Multivariate Calibration. A Journal of the Chemometrics Society. 2004; 18(2):112–20.

47. Dubitzky W, Granzow M, Berrar DP. Fundamentals of Data Mining in Genomics and Proteomics. Springer Science & Business Media; 2007.

48. Rosenblatt F. Principles of Neurodynamics. Perceptrons and the Theory of Brain Mechanisms. 1961.

49. McCulloch WS, Pitts W. A Logical Calculus of the Ideas Immanent in Nervous Activity. Bull Math Biophys. 1943; 5(4): 115–33.

50. ASTM E. Standard Practice for Use of the Terms Precision and Bias in ASTM Test Methods. ASTM International. 2008.