LB-5117

An Autoencoder and Artificial Neural Network-based Method to Estimate Parity Status of Mosquitoes Using NIRS

*Masabho P. Milali[1,2], Samson S. Kiware[1,2], Naveen Bansal[1], Serdar Bozdag[1], Jacques D. Charlwood[3], Floyd E. Dowell[4], George F. Corliss[1] and Richard J. Povinelli[1]
[1]Marquette University USA, [2]Ifakara Health Institute Tanzania, [3]Liverpool School of Tropical Medicine, UK, [4]USDA, Agricultural Research Service, Center for Grain and Animal Health Research

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

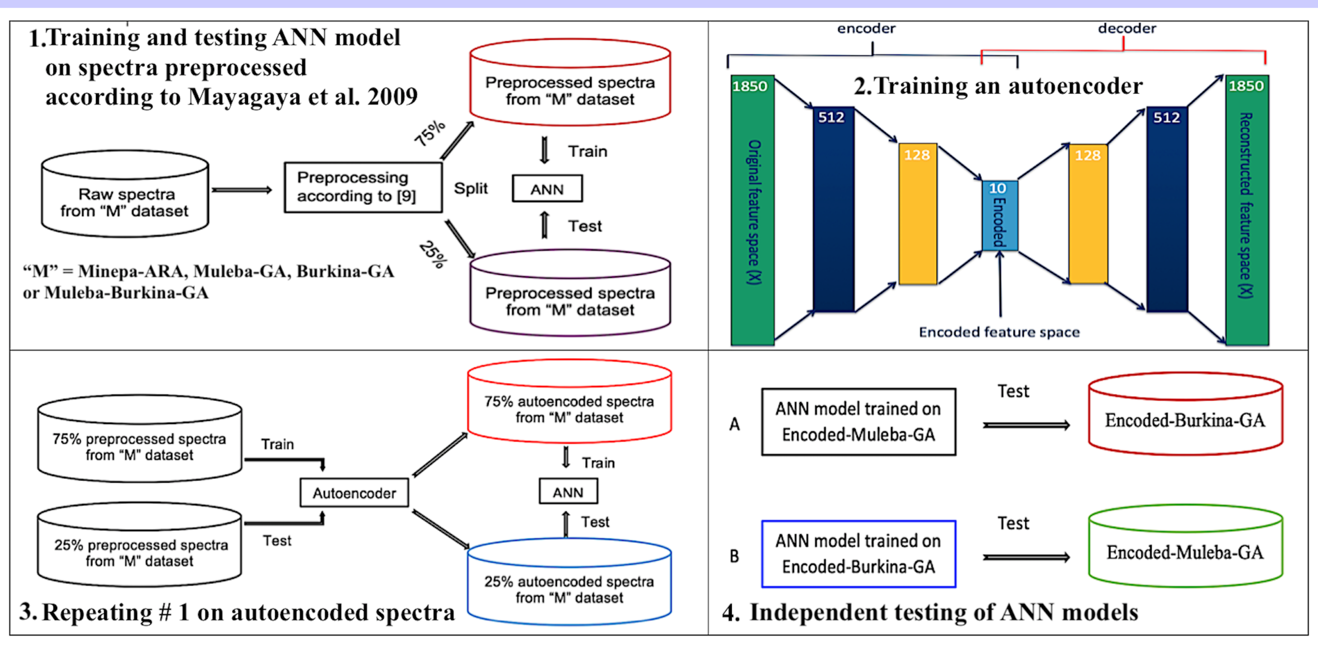
Introduction and Method

Introduction: Knowing the parity status of mosquitoes is useful in control and evaluation of infectious diseases transmitted by mosquitoes, where parous mosquitoes are assumed to be potentially infectious. Ovary dissections, which are currently used to determine parity status of mosquitoes, are very tedious and limited to very few experts. An alternative to ovary dissections is near-infrared spectroscopy (NIRS), which can estimate age in days and infectious state of laboratory and semi-field reared mosquitoes with accuracies between 80 and 99%. No study has tested the accuracy of NIRS for estimating parity status of wild mosquitoes.

Study objective: To apply an autoencoder and artificial neural network (ANN) based method to estimate parity status of wild mosquitoes using near-infrared spectra.

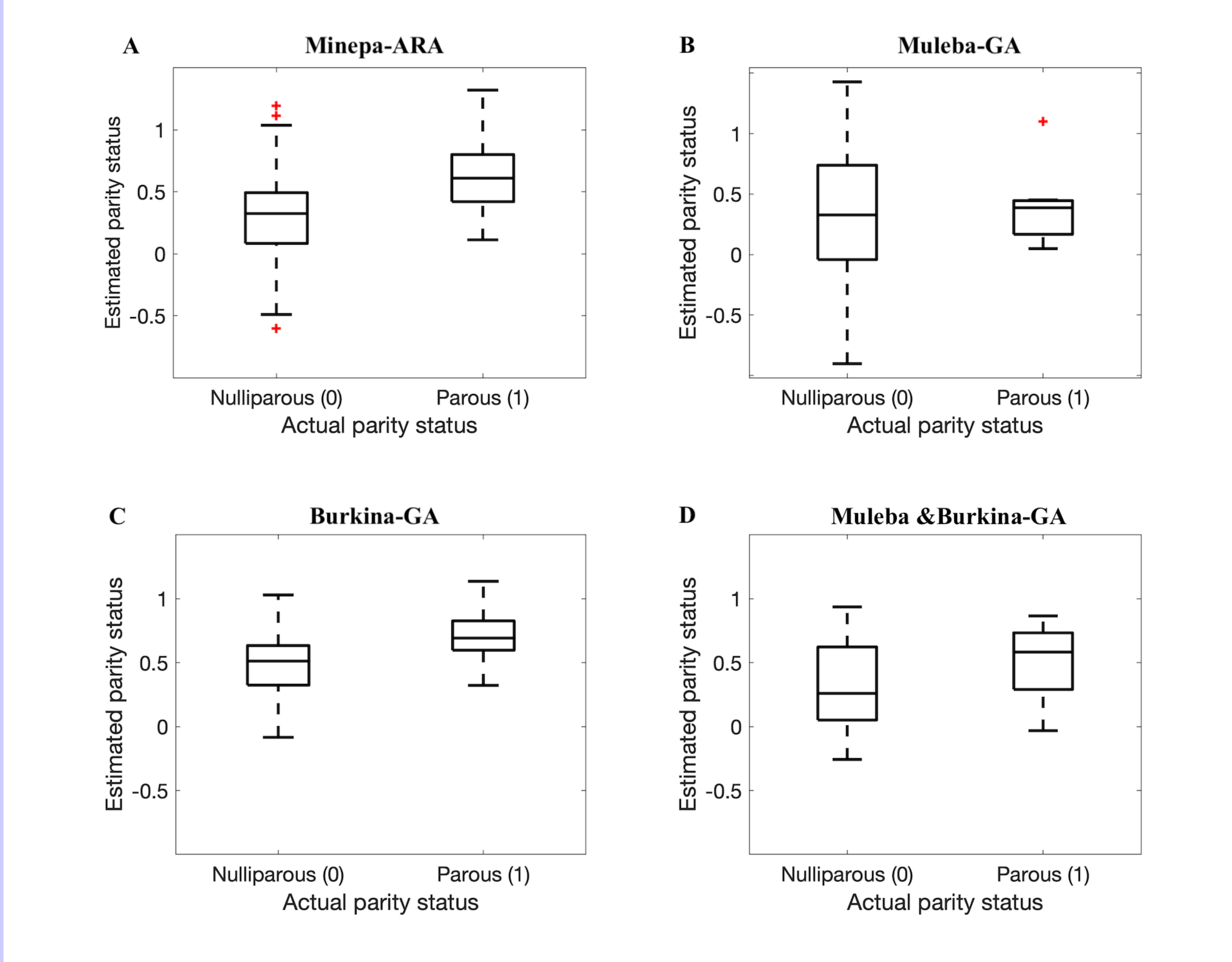
Materials: We use four different datasets; *Anopheles arabiensis* collected from Minepa, Tanzania (Minepa-ARA, N = 927); *Anopheles gambiae s.s* collected from Muleba, Tanzania (Muleba-GA, N = 140); *Anopheles gambiae s.s* collected from Burkina Faso (Burkina-GA, N = 158); and *An.gambiae s.s* from Muleba and Burkina Faso combined (Muleba-Burkina-GA, N = 298). While LabSpec 5000 NIR spectrometer with an integrated light source (ASD Inc., Malvern, UK), was used to collect spectra in Minepa-ARA and Muleba-GA, LabSpec4i spectrometer (ASD Inc., Malvern, UK) was used to collect spectra in Burkina-GA.

Model training: We train ANN models on datasets with spectra only pre-processed according to previous protocols. We then use autoencoders to reduce the spectra feature dimensions from 1851 to 10 and re-train ANN models. On each dataset, using ten Monte-Carlo cross validations and Levenberg-Marquardt optimization, a one hidden layer, ten-neuron feed-forward ANN model with logistic regression as a transfer function was trained and tested in Matlab



Acknowledgment: We extend our gratitude to Fredros O. Okumu, Sheila Ogoma and Marta F. Maia for supporting grant writing and managing of the project that produced the Minepa-ARA dataset; Maggy Sikulu-Lord for helping to scan the mosquitoes in the Muleba-GA, and PI of the original project that enabled collection of mosquitoes in the Minepa-ARA dataset; Benjamin Krajacich for allowing us to use Burkina-GA dataset; Gustav Mkandawile who made sure we obtained mosquitoes in Minepa-ARA and Muleba-GA datasets; 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.

Results - Before autoencoder



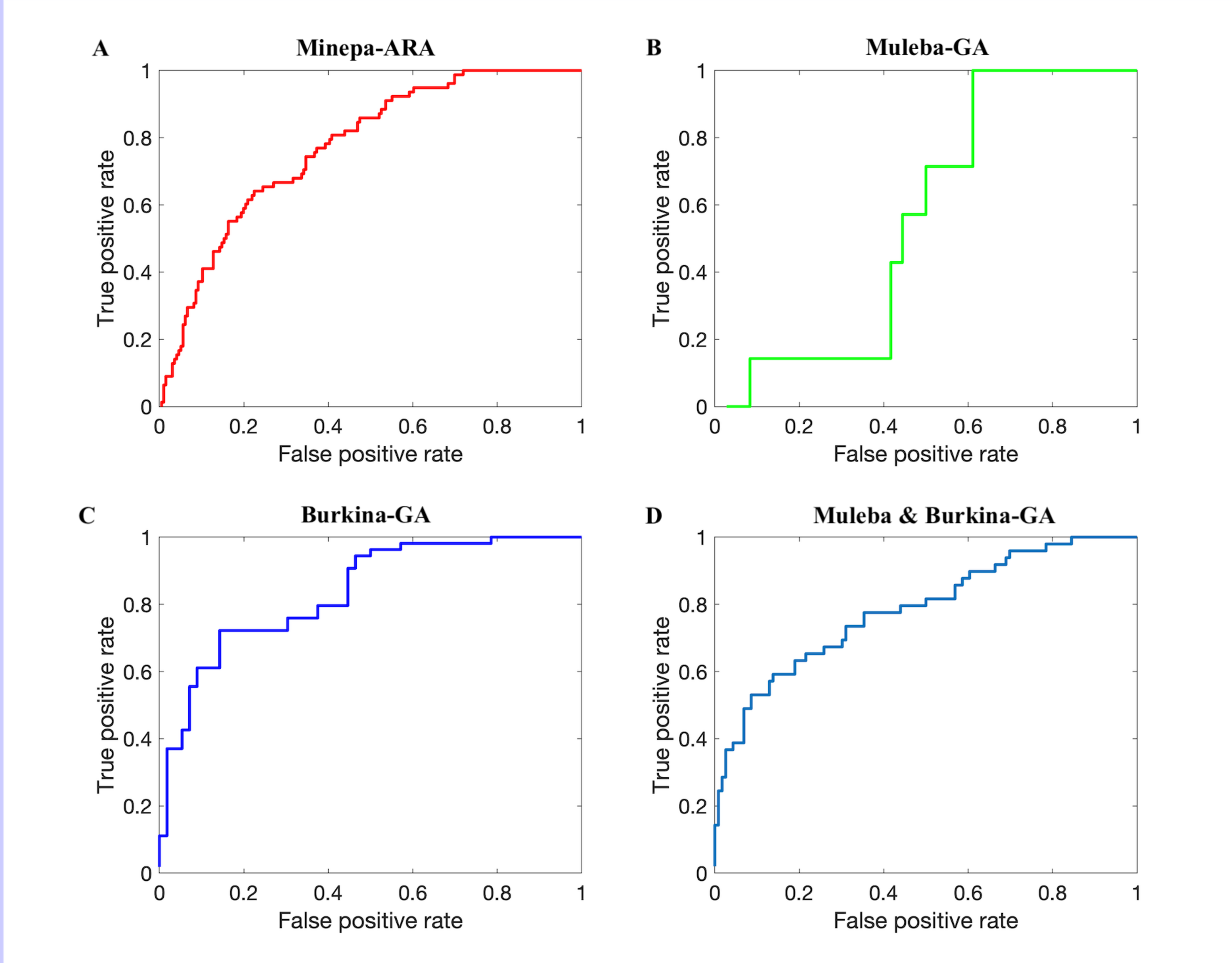


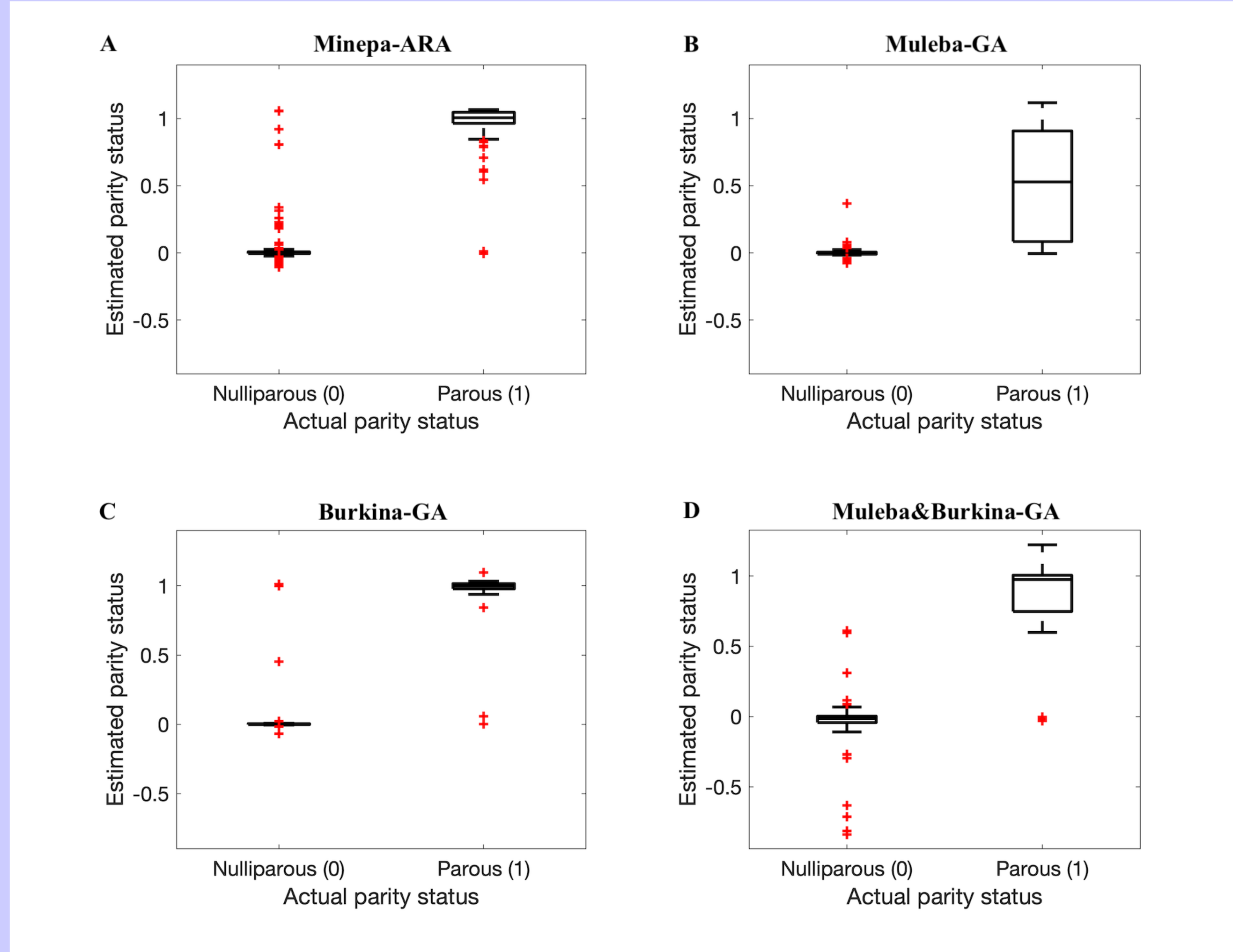
Table 1: Performance of an ANN model trained on 75% of mosquito spectra with 1851 features (before autoencoder) and tested on the remaining 25% spectra (out of the sample testing). Minepa-ARA (Nulliparous = 656, Parous = 271), Muleba-GA (Nulliparous = 119, Parous = 21) Burkina-GA (Nulliparous = 80, Parous = 78)

	Minepa-ARA (N = 927)	Muleba-GA (N = 140)	Burkina-GA (N = 158)	Muleba-Burkina-GA (N = 298)
Accuracy (%)	81.9 ± 2.8	68.7 ± 4.8	80.3 ± 2.0	75.7 ± 2.5
Sensitivity (%)	79.7 ± 3.2	37.8 ± 6.6	76.5 ± 2.1	70.2 ± 3.1
Specificity (%)	86.0 ± 1.6	80.1 ± 2.7	88.3 ± 2.3	77.6 ± 2.9
Precision (%)	74.3 ± 3.4	31.3 ± 5.2	77.8 ± 1.8	68.8 ± 3.2
AUC (%)	77.2	55.9	83.6	76.4

Conclusion

These results show that a combination of an autoencoder and ANNs on NIR spectra yields models that can be used as an alternative tool to estimate parity status of wild mosquitoes, especially since NIRS is a high-throughput, reagent-free, and simple-to-use technique compared to ovary dissections.

Results - After autoencoder



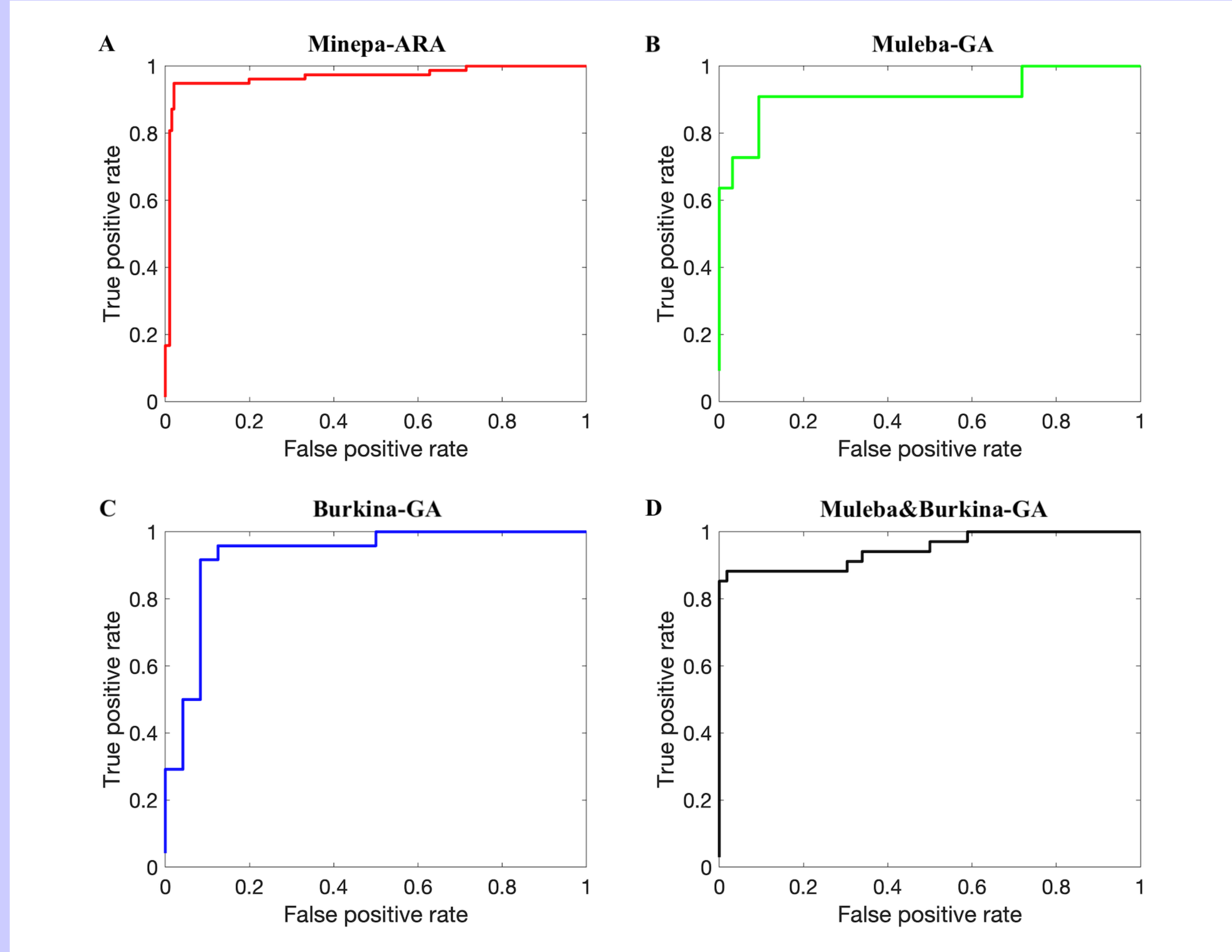


Table 2: Performance of an ANN model trained on 75% of the encoded mosquito spectra (10 features) and tested on the remaining 25% of the encoded mosquito spectra. Minepa-ARA (Nulliparous = 656, Parous = 271), Muleba-GA (Nulliparous = 119, Parous = 21), Burkina-GA (Nulliparous = 80, Parous = 78).

	Minepa-ARA (N = 927)	Muleba-GA (N = 140)	Burkina-GA (N = 158)	Muleba-Burkina-GA (N = 298)
Accuracy (%)	97.1 ± 2.2	89.8 ± 1.7	93.3± 1.2	92.7 ± 1.8
Sensitivity (%)	94.9 ± 1.6	70.1 ± 2.3	91.7 ± 1.9	88.2 ± 2.9
Specificity (%)	98.6 ± 1.3	96.9 ± 1.2	96.4 ± 1.6	94.7 ± 2.1
Precision (%)	93.7 ± 2.4	62.5 ± 3.2	91.3 ± 1.4	93.1 ± 2.5
AUC (%)	96.7	91.5	93.1	94.9

Table 3: Independent testing of ANN models trained on Muleba-GA and Burkina-GA encoded datasets

	ANN model trained on Encoded-Muleba-GA, tested on Encoded-Burkina-GA	ANN model trained on Encoded-Burkina-GA, tested on Encoded-Muleba-GA
Accuracy (%)	68.6	88.3
Sensitivity (%)	26.5	86.1
Specificity (%)	94.4	92.2

Financial support: We thank Marquette University GasDay and the Ifakara Health Institute for funding the study.

References: 9. Mayagaya VS, et al. 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