


RESEARCH

Open Access



# Using mid-infrared spectroscopy and supervised machine-learning to identify vertebrate blood meals in the malaria vector, *Anopheles arabiensis*

Emmanuel P. Mwanga<sup>1\*</sup> , Salum A. Mapua<sup>1</sup>, Doreen J. Siria<sup>1</sup>, Halfan S. Ngowo<sup>1,4</sup>, Francis Nangacha<sup>1</sup>, Joseph Mgando<sup>1</sup>, Francesco Baldini<sup>4</sup>, Mario González Jiménez<sup>3</sup>, Heather M. Ferguson<sup>4</sup>, Klaas Wynne<sup>3</sup>, Prashanth Selvaraj<sup>2</sup>, Simon A. Babayan<sup>4</sup> and Fredros O. Okumu<sup>1,4,5</sup>

## Abstract

**Background:** The propensity of different *Anopheles* mosquitoes to bite humans instead of other vertebrates influences their capacity to transmit pathogens to humans. Unfortunately, determining proportions of mosquitoes that have fed on humans, i.e. Human Blood Index (HBI), currently requires expensive and time-consuming laboratory procedures involving enzyme-linked immunosorbent assays (ELISA) or polymerase chain reactions (PCR). Here, mid-infrared (MIR) spectroscopy and supervised machine learning are used to accurately distinguish between vertebrate blood meals in guts of malaria mosquitoes, without any molecular techniques.

**Methods:** Laboratory-reared *Anopheles arabiensis* females were fed on humans, chickens, goats or bovines, then held for 6 to 8 h, after which they were killed and preserved in silica. The sample size was 2000 mosquitoes (500 per host species). Five individuals of each host species were enrolled to ensure genotype variability, and 100 mosquitoes fed on each. Dried mosquito abdomens were individually scanned using attenuated total reflection-Fourier transform infrared (ATR-FTIR) spectrometer to obtain high-resolution MIR spectra ( $4000\text{ cm}^{-1}$  to  $400\text{ cm}^{-1}$ ). The spectral data were cleaned to compensate atmospheric water and  $\text{CO}_2$  interference bands using Bruker-OPUS software, then transferred to Python™ for supervised machine-learning to predict host species. Seven classification algorithms were trained using 90% of the spectra through several combinations of 75–25% data splits. The best performing model was used to predict identities of the remaining 10% validation spectra, which had not been used for model training or testing.

**Results:** The logistic regression (LR) model achieved the highest accuracy, correctly predicting true vertebrate blood meal sources with overall accuracy of 98.4%. The model correctly identified 96% goat blood meals, 97% of bovine blood meals, 100% of chicken blood meals and 100% of human blood meals. Three percent of bovine blood meals were misclassified as goat, and 2% of goat blood meals misclassified as human.

**Conclusion:** Mid-infrared spectroscopy coupled with supervised machine learning can accurately identify multiple vertebrate blood meals in malaria vectors, thus potentially enabling rapid assessment of mosquito blood-feeding histories and vectorial capacities. The technique is cost-effective, fast, simple, and requires no reagents other than desiccants. However, scaling it up will require field validation of the findings and boosting relevant technical capacity in affected countries.

\*Correspondence: emwanga@ihi.or.tz

<sup>1</sup> Environmental Health and Ecological Science Thematic Group, Ifakara Health Institute, Morogoro, Tanzania

Full list of author information is available at the end of the article



**Keywords:** Mid-infrared spectroscopy, Supervised machine learning, Malaria, *Anopheles arabiensis*, Mosquito blood meals, Ifakara, Vector surveillance

## Background

The Global Technical Strategy for Malaria Elimination 2016–2030 [1] recommends that countries should integrate effective surveillance as a core intervention in their malaria policies. As such, the World Health Organization (WHO) recently provided guidelines to support measurements of the most important parasitological and entomological indicators [2]. Effective entomological surveillance requires detailed quantitative understanding of key biological attributes which influence overall potential of vector populations to transmit *Plasmodium* to humans [3]. Such attributes may include the likelihood with which specific *Anopheles* populations bite humans as opposed to the other available vertebrate hosts, i.e. the human blood indices (HBI), defined as proportion of all mosquito blood meals obtained from humans [4, 5]. Other attributes include parasite infection rates, i.e. the proportion of females infected with *Plasmodium* [6], survivorship, i.e. whether the mosquitoes can live long enough to allow complete sporogonic development of *Plasmodium* inside them [7], mosquito susceptibility to insecticides commonly used to control them [8], and the location of mosquito biting, i.e. indoors or outdoors, and how it overlaps in space and time with humans [9–12].

Accurate identification of mosquito blood meal sources is important for understanding host–vector interactions, and provides essential information on transmission dynamics of mosquito-borne diseases [13, 14]. Until recently, blood meals in haematophagous insects were typically identified using immunological assays such as the latex agglutination test [15], precipitin test [16] or enzyme-linked immunosorbent assays (ELISA) [13]. Kent et al. published the first polymerase chain-reaction (PCR) based assay, which addressed many limitations of previous methods and enabled accurate detection of blood meals in field-collected mosquitoes up to several hours post-feeding on cows, dogs, human, pigs, and goats [14]. Lately, other techniques, such as matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS) has been applied for mosquito blood meal identification [17–19]. Today, the ELISA [13, 20] and PCR [14] assays are the primary reference methods for measuring HBI in malaria vectors used by most laboratories. Despite generally offering reliable results, these procedures are time-consuming and require repeated supply of reagents making them expensive and unreliable

in poorly-resourced settings. Additionally, the ELISA assays are prone to cross-reactivity if laboratory standards are not regularly updated [21].

Non-molecular techniques such as infrared spectroscopy may offer just as effective but cheaper, quicker, reagent-free and potentially simpler alternatives for assessing key malaria transmission indicators. Indeed, studies have shown that near-infrared (NIR) spectroscopy coupled with chemometrics (mathematical methods of understanding chemical systems) can predict different mosquito ages [22–27], distinguish between mosquito species [22, 28] and even detect presence or absence of pathogens such as *Wolbachia* bacteria, *Plasmodium* and Zika virus in the mosquitoes [29–32]. These successes could be vastly improved by using more effective analytical approaches to process spectral data. Further improvements could potentially be achieved by relying on mid-infrared (MIR) wavelengths ( $4000\text{ cm}^{-1}$  to  $400\text{ cm}^{-1}$ ), which compared to those of NIR ( $12,500\text{ cm}^{-1}$  to  $400\text{ cm}^{-1}$ ), Fig. 1 also allow detection of changes in chemical composition of samples, and can clearly show contributions of different chemical bonds of product constituents in separate peaks [33].

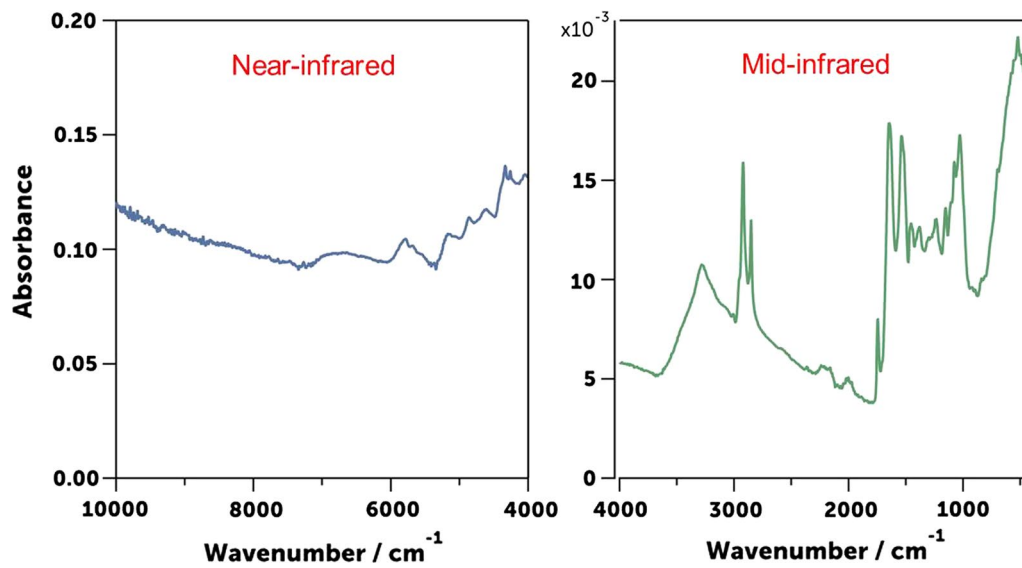
This current study investigated the potential of using supervised machine learning algorithms and MIR spectroscopy to accurately distinguish between blood meals of four different vertebrate species within abdomens of the malaria vector, *Anopheles arabiensis*.

## Methods

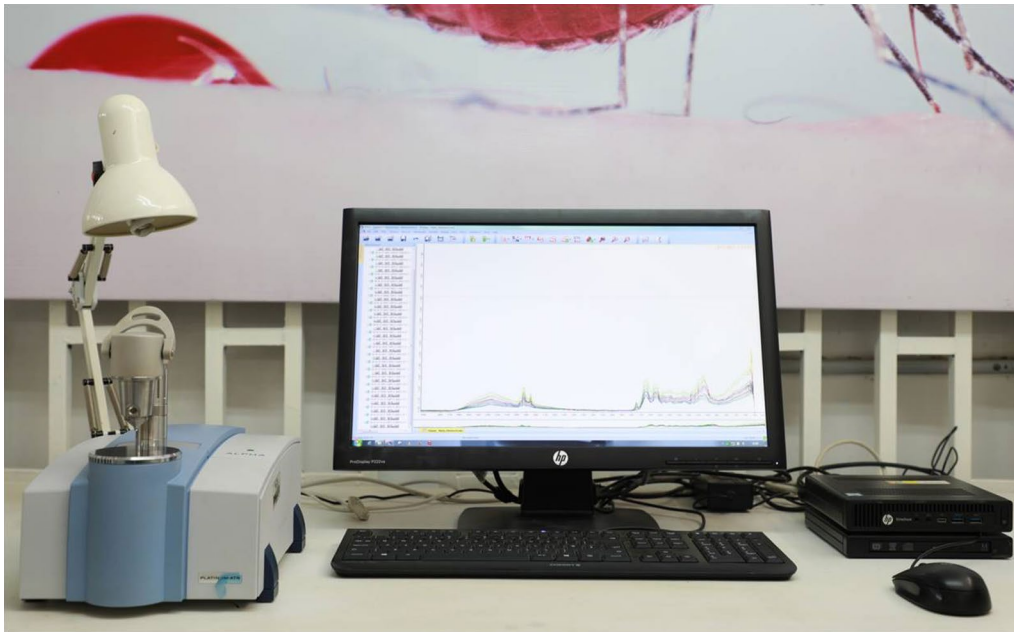
### Mid-infrared spectrometer

A Bruker ALPHA Fourier-transform infrared (FTIR) spectrometer equipped with a Platinum ATR device was used [34]. The spectrometer had a platinum diamond sampling module with a spectral range of  $375\text{--}7500\text{ cm}^{-1}$  and maximum spectral resolution between 2 and  $0.8\text{ cm}^{-1}$ . The infrared optical window was fitted with zinc selenide (ZnSe) to accommodate high humidity conditions.

The unit is small ( $22\text{ cm} \times 30\text{ cm}$ ), highly portable (Fig. 2) and has a permanently aligned interferometer for precise data acquisition [35]. It was installed with an internal validation unit (IVU) with reference standards and programmed to conduct automated instrument tests for operational and performance qualification. Its core is encased in robust metal housing with a lifespan



**Fig. 1** Differences between NIR and MIR spectra obtained from dried mosquito samples collected using ATR-FTIR spectrometer. Compared to near-infrared (NIR), mid-infrared (MIR) allows detection of changes in chemical composition of the samples. Its wavelengths are more sensitive to fundamental vibration of molecular bonds and the different isolated peaks contain information of different chemical components in the mosquito cuticle



**Fig. 2** Mid-infrared ALPHA spectrometer with attenuated total reflectance (ATR), and single reflexion diamond platinum crystal, installed at the VectorSphere, Ifakara Health Institute, Tanzania. A control computer is included for the operator

greater than 10 years, and requires minimal maintenance other than replacement of desiccants depending on humidity inside the mid-IR spectrometer [35].

The spectrometer together with the operating computer was installed in the vector biology laboratory, the VectorSphere, at Ifakara Health Institute (IHI), Ifakara, Tanzania (Fig. 2). Proprietary OPUS software version 7.5

[36], licensed to IHI, was also installed to record and process the MIR spectra.

### Mosquitoes

The malaria vector *An. arabiensis* was used in this study because of the natural plasticity of its blood-feeding preferences, and its readiness to feed on non-human hosts when humans are not available [4, 5]. Laboratory-reared females were used. Larvae were reared in plastic basins and fed on Tetramin<sup>®</sup> fish food (Tetra GmbH, Melle, Germany), while adults were maintained on 10% sugar meals and human arm-feeding for colony maintenance. Adult mosquitoes were maintained at temperature of  $27 \pm 2$  °C, relative humidity (RH) of  $80 \pm 5\%$  as previously described [37]. Females aged 4–6 days old were used for experiments, all starved for 6 h prior to direct blood feeding as described below.

### Mosquito blood-feeding on different vertebrate hosts

We identified four vertebrate host species widely available and commonly fed upon by *Anopheles* mosquitoes [5, 38] in rural Tanzania. These included: bovine, chicken, goat, and human. For each host species, five individuals were recruited, and 100 female *An. arabiensis* fed upon each one. This way, every host species had 500 blood-fed mosquitoes (100 per individual host/replicate). All humans were all males and recruited from the research team, and other animal hosts (i.e. bovine, chicken, and goat) were bought and were part of the research project. All hosts were restrained and mosquitoes were fed until they were fully engorged. The blood-feeding took place over several days so that the groups consisted of individual mosquitoes from different reproductive batches in the mosquito colony. After blood-feeding, mosquitoes were held for 6 h for digestion to begin and to minimize potential differences associated with extent of blood digestion in the gut [39]. After the holding period, the mosquitoes were killed using chloroform [33] and preserved in micro-centrifuge tubes with silica gel to keep them dry before scanning. Each sample was labelled by date, vertebrate host type, mosquito species, sample ID and age.

### Scanning the preserved mosquitoes

Abdomens were first separated from the heads and thoraxes of the dried mosquitoes. Since the study was primarily focused on gut content, the heads and thoraxes were discarded and instead only the abdomens were scanned. The mosquito specimen (abdomen) was placed at the center of the MIR crystal plate, and supported by the spectrometer arm (the anvil). MIR spectra were captured (spectral range  $4000\text{--}400$   $\text{cm}^{-1}$ ), with spectral resolution set at  $5$   $\text{cm}^{-1}$ . Each individual specimen was scanned 32 times in 30 s, and resulting spectra averaged

to obtain a single representative spectrum as previously described [33]. The spectra were recorded in absorbance units and stored using Bruker OPUS software [36].

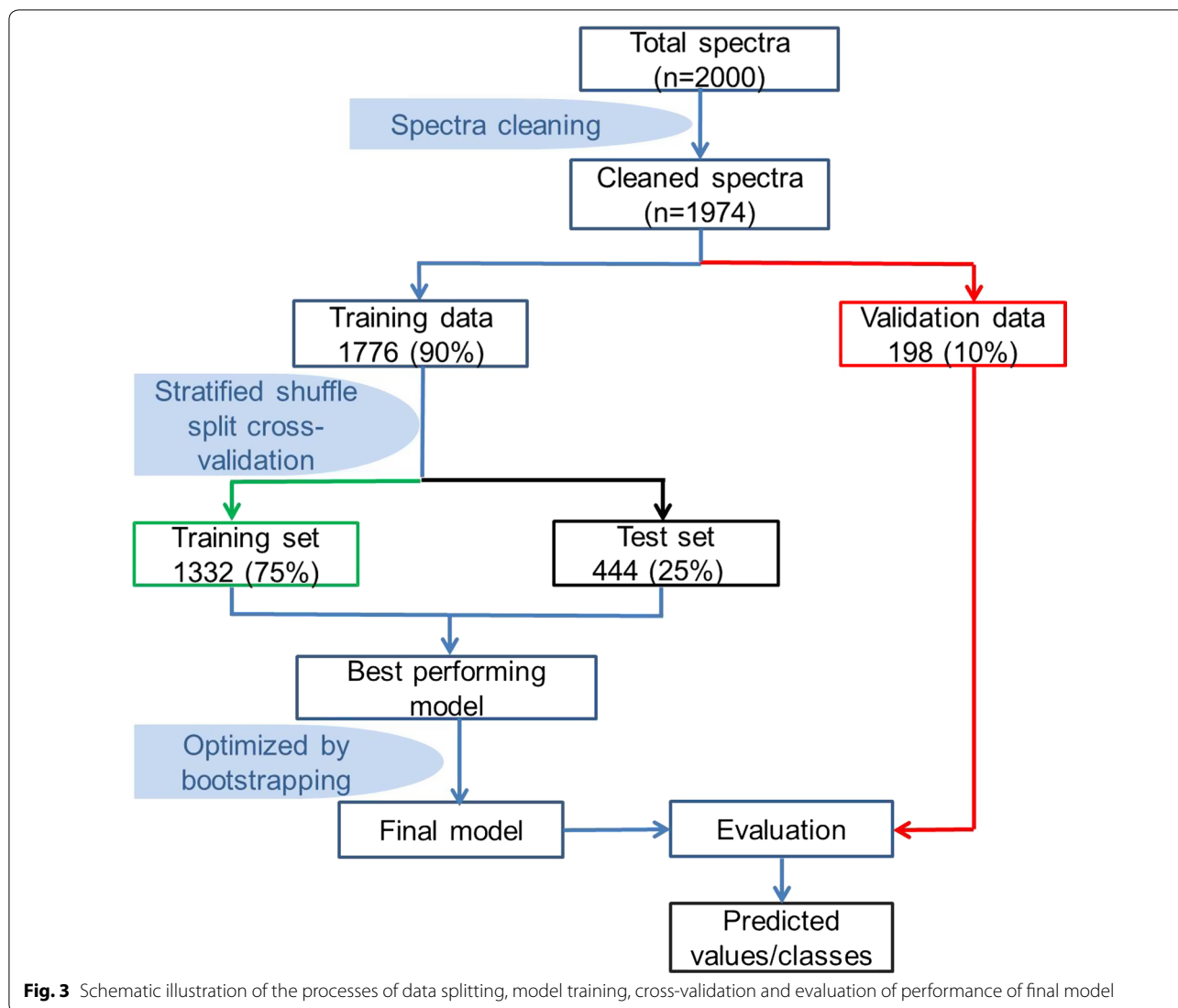
### Pre-processing of the MIR spectra

The OPUS software [36] was used to clean and compensate spectra with water vapor absorption bands (intense bands centered around  $2340$   $\text{cm}^{-1}$ ,  $3600$  and  $550$   $\text{cm}^{-1}$  wavenumbers) and carbon dioxide ( $\text{CO}_2$ ) interference bands ( $4000\text{--}3400$   $\text{cm}^{-1}$  and  $2200\text{--}1300$   $\text{cm}^{-1}$ ) as previously described [33]. The cleaned spectral data were converted from the Bruker OPUS format to text files in Python<sup>™</sup>. Further spectral cleaning was done by discarding spectra with low intensities (below 0.11 absorbance units; mostly between  $400$  and  $500$   $\text{cm}^{-1}$ ) and spectra with no features (flat spectra) [33]. The final cleaned spectra matrix was saved in comma separated values (CSV) format for further analysis.

### Data training, validation, and prediction of blood meal sources

The MIR spectra were analyzed in the Python<sup>™</sup> programming language with the *scikit-learn* library [40]. Supervised machine learning techniques were used to train models on known and predicted classes for the validation set, based on a multi-class classification strategy [40]. Models were trained and cross-validated using the strategy as illustrated in Fig. 3. First, the whole dataset was partitioned into training (90% of the data) and validation datasets (10%), ensuring proportional representation of the different host types and individuals (Fig. 3). The validation set was split from the whole dataset before training process such that the 10% subset was neither used for training nor for selection of the trained models. Instead, it was preserved for evaluating accuracies of the final model. Once the training dataset was separated, it was itself subjected to multiple rounds of randomly-stratified splits into training sets (75%) and test sets (25%) as illustrated in Fig. 3, to achieve rigorous classification of the different blood meal types, this time involving only the training dataset. The spectra data were used to classify blood meals into one of four host species classes.

To achieve this, classification algorithms were used to learn patterns from matrices of features which represent different blood meal sources. Seven different algorithms were tested using default settings on the training set. The candidate algorithms included: k-nearest neighbours classifier (KNN), logistic regression (LR), support vector machine classifier (SVM), naïve Bayes (NB), random forest classifier (RF), XGBoost classifier (XGB), Multilayer perceptron (MLP) [40]. Prediction scores were presented in terms of percentage accuracy.



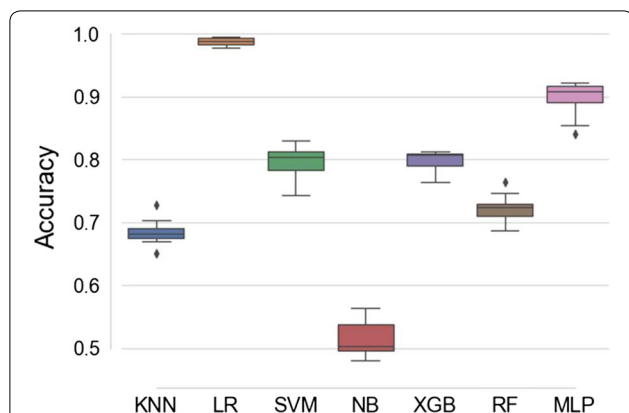
Finally, the best performing model of the seven above, i.e. model with highest accuracy (percentage of times a blood meal was correctly classified to the right host species) and precision (variability between actual estimates), was selected. The model was optimized by fitting 100 bootstrapping regressions, and bagged to increase prediction performance. Performance of this final model was evaluated using the validation set.

**Results**

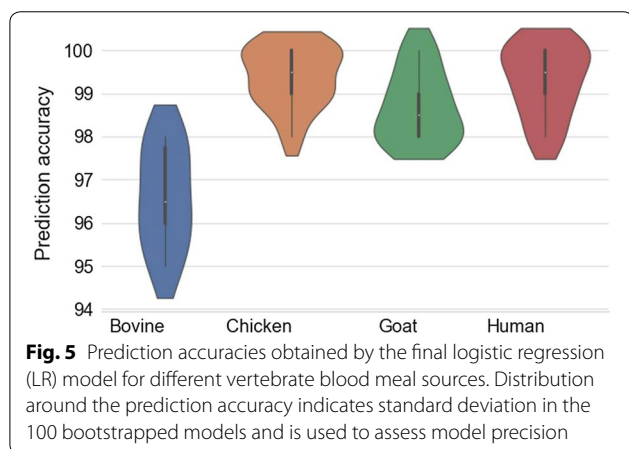
Of the 2000 individual spectra collected, 26 randomly distributed were discarded during data cleaning because of sub-standard quality. Of the remaining 1974 spectra, 1776 were used in training and testing of supervised machine learning models and 198 were used for final validation of the final model.

Of the seven classification algorithms tested, logistic regression (LR) was identified as the best approach since it outperformed the other six classifiers in identifying mosquito blood meal sources of laboratory-reared *An. arabiensis* (Fig. 4). After additional optimization by bootstrapping, LR successfully predicted mosquito blood meal sources by correctly identifying the source of host blood meals more than 90% of the time. A total of 100 bootstrapped models were fitted, which when aggregated predicted mosquito blood meals with an overall accuracy of 98.6%, Fig. 5). Average accuracies by class were 98% for bovine and human blood, 99% for goat blood, and 100% for chicken blood (Fig. 6).

In the final validation on held-out 198 previously unseen spectra, the optimized model predicted correct identities of blood meals in this new dataset by 98.4% overall accuracy (96% for goat blood, to 97% for bovine



**Fig. 4** Prediction accuracies for different classification algorithms. Models tested include k-nearest neighbours (KNN), logistic regression (LR), support vector machines (SVM), naïve Bayes (NB), XGBoost (XGB), random forest (RF), Multilayer perceptron (MLP). Based on prediction accuracy and precision achieved, the best performing model was LR



**Fig. 5** Prediction accuracies obtained by the final logistic regression (LR) model for different vertebrate blood meal sources. Distribution around the prediction accuracy indicates standard deviation in the 100 bootstrapped models and is used to assess model precision

blood and 100% for chicken and human blood) (Fig. 7). Three-percent of bovine blood samples were misclassified as goat blood and 2% of goat blood misclassified as human blood (Fig. 7).

## Discussion

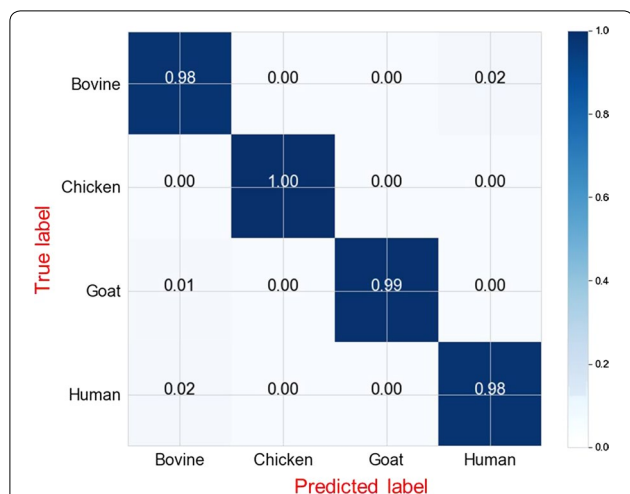
This work has demonstrated that mid-infrared spectroscopy (MIR) coupled with supervised machine learning can accurately distinguish between mosquito blood meals originating from common vertebrate hosts, i.e. bovine, chicken, goat and human, without requiring molecular techniques. It also represents the first evaluation of an infrared-spectroscopy approach for blood meal analysis, and application of machine learning to analyze mosquito blood feeding histories spectral data. The experiments were done using laboratory-reared *An. arabiensis* mosquitoes but with known blood hosts replicated across five individuals for each host type. *An. arabiensis* was selected

for these experiments because of its natural plasticity of host preferences [4, 5], but similar analyses would apply for all other mosquito species.

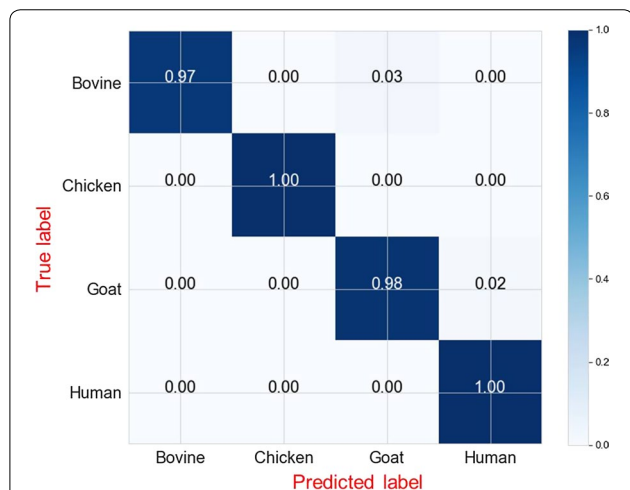
This work builds upon a series of studies using NIR spectroscopy, most of which have already demonstrated the potential of these technologies for analysis of several important mosquito traits. Examples include prediction of different mosquito ages [22–27], use of NIR spectroscopy and chemometrics to distinguish between two malaria vector species, *Anopheles gambiae* sensu stricto and *An. arabiensis* [22, 28], and detection of pathogens such as Zika virus [30], malaria parasites [31, 32] and the symbiont, *Wolbachia* [29, 30].

Until now a key challenge of this approach has been the lack of high-capacity statistical tools to handle the massive quantities of spectral data generated from the procedures, and also lack of field validation of many of these approaches. Both these challenges are now on the verge of being addressed by multiple research groups. As recently described by González-Jiménez et al. [33], this current study also applied MIR as opposed to the NIR wavelengths previously used. MIR extends from 4000 to 400  $\text{cm}^{-1}$ , and is between the far-infrared (FIR) region (400  $\text{cm}^{-1}$  to 10  $\text{cm}^{-1}$ ) and NIR region (12,500  $\text{cm}^{-1}$  to 400  $\text{cm}^{-1}$ ) [41]. As shown in Fig. 1, it records spectra with greater information content compared to NIR. Moreover, key features of MIR spectra such as number of infrared absorption bands and their intensities combined with the advantage of robust instrumentation available for MIR such as used here [34]. The mosquito abdomen was squashed during the scanning so blood meal was present at the surface of the specimen. Different isolated peaks in MIR contain information of different chemical components in the mosquito cuticle, at it appears proteins and lipids may be responsible for the difference in spectra for different vertebrate as also observed in MALDI-TOF MS approach [17–19]. The MIR-based approach has great potential for accuracy and scalability. In this study, only mosquito abdomen was used but also other parts (head and thorax) could be used to distinguish between species and age.

To process the MIR data, this study deployed multiple supervised machine learning algorithms before selecting the most accurate and precise candidate for final analysis. Machine learning is increasingly applied in medical and public health industry [33, 42–48], and will likely become dominant in disease predictions and surveillance [42, 43]. For example, it has been used to solve problems in genomic medicine [44, 45] and to predict responses to antiretroviral treatment, in one case with ~78% accuracy [46]. Nearly 10 years ago, the approaches were already being used to predict diabetes and pre-diabetes outcomes with greater than 80% accuracy [47]. Recently, Chen



**Fig. 6** Normalized confusion matrix for the trained model (training set = 1332 spectra; test set = 444 spectra; total spectra = 1776). Each row represents instances in actual class (true label), while each column represents instances in predicted class (predicted label). From the top left to bottom right, the blue line highlights final prediction accuracies in each class



**Fig. 7** Normalized confusion matrix for final model evaluation (training data = 1776 spectra; validation data = 198 spectra; total spectra = 1974). Each row represents instances in actual class (true label), while each column represents instances in predicted class (predicted label). From the top left to bottom right, the blue line highlights final prediction accuracies in each class

meal sources. Optimization of the LR model with 100 bootstrapped realizations of the dataset led to very high prediction accuracy of 98.8%, achieving perfect score in a few instances (Fig. 4). The selected model remained very highly accurate, exceeding 98% even when challenged with a new dataset not previously seen by the model.

The technique has shown to be highly effective, achieving accuracies previously achieved by ELISA [13] and PCR. However, the samples here were prepared just 6 to 8 h after blood-feeding on known animal bloods, so future studies should consider using different digestion stages as this may influence accuracy [14, 39, 49]. The technique will need to be evaluated whether it can detect and distinguish blood feeding histories of malaria mosquitoes for more than 6 h as well as mixed blood meals. In this study, there was no uniformity between the individuals of the same host species in terms of attributes such as sex, age, weight and health status. Mosquitoes selected in this study were fully engorged; it is still unknown whether the technique will also detect mosquitoes with partial blood meals. Additional advantages of this technique over direct ELISA include the fact that it is time-saving in both sample preparation and analysis, and has reduced cross-reactivity. Despite the fact that PCR is even more sensitive than ELISA and has low risk of cross reactivity, they still consume time and require skilled expertise in deoxyribonucleic acid (DNA) extraction [14]. Lastly, all mosquito samples used in this study were blood-fed on known hosts in a controlled environment. Future work should therefore validate these findings using field collections, in which case PCR assays would be the best standard reference [14].

The costs are also significantly lower, and particularly since no reagents are required other than desiccants. ELISA systems currently cost approximately 13,044 USD, while the cost for MIR spectrometer we used (Bruker ALPHA Fourier-transform infrared (FTIR) spectrometer equipped with a platinum ATR) was approximately 29,000 USD, including shipment and installation, and the costs were incurred only at the initial purchase of the machine. Unlike the MIR spectrometry, ELISA systems require additional reagents, such that the cost per sample can be between 1 and 1.5 USD. MIR is more cost-effective as it does not require repeated reagents for sample processing. This means that in an active laboratory, it would take just 1 year for the overall financial investment in ELISA systems to exceed the costs of MIR systems, and thereafter the costs per sample would continue to reduce. The turnaround time between sample preparation and results for ELISA is approximately 2 days for every 100 samples (including sample preparation, processing and results reading). On the other hand, once the MIR system has been calibrated and established, users

et al. proposed a faster neural network approach based on multimodal disease risk prediction using data from health care facilities [48]. This approach reached 94% accuracy in evaluating risk of cerebral infarction disease.

In this current work, logistic regression (LR) models were found to be the best performing for quantifying variations of MIR spectral information on mosquito blood

can scan 300 or more samples in a day, each sample taking approximately 1 min to scan. The scale of use also suggests that this tool could greatly improve district-wide or nation-wide vector surveillance efforts.

## Conclusion

In conclusion, mid-infrared spectroscopy coupled with supervised machine learning can accurately identify multiple vertebrate blood meals in malaria vectors, thus enabling rapid assessment of mosquito blood-feeding histories and vectorial capacities. The technique is cost-effective, fast, simple and requires no reagents other than desiccants. All the analyses were done in open source software, except for data extraction done using the proprietary Bruker OPUS software, but which can also be done with available open source scripts. Nonetheless, scaling up this approach will require field validation of the findings and specific training to improve technical capacity in affected countries. Once validated, this approach could potentially replace current molecular techniques for blood meal analysis (i.e. PCR and ELISA).

## Abbreviations

HBI: Human Blood Index; ELISA: enzyme-linked immunosorbent assay; PCR: polymerase chain reaction; MIR: mid-infrared; ATR-FTIR: attenuated total reflectance-Fourier transform infrared spectrometer; MALDI-TOFMS: matrix-assisted laser desorption ionization-time of flight mass spectrometry; NIR: near-infrared; ATR: attenuated total reflectance; FTIR: Fourier transform infrared spectrometer; KNN: k-nearest neighbours; LR: logistic regression; NB: naïve Bayes; RF: random forest; XGB: gradient boosting; MLP: Multilayer perceptron.

## Acknowledgements

We express our gratitude to for the research and administrative support team at Ifakara Health Institute. The authors thank Khamis Kifungo and Rukiyah M. Njalambaha for their assistance in logistics management. The authors thank Marcelina Finda, and Johnson K. Swai, Arnold S. Mmbando for assessing initial drafts of this manuscript and providing important comments.

## Authors' contributions

EPM, SAM, DJS, HSN, and FOO designed the study. EPM performed semi-field experiments. EPM and FOO wrote and revised the manuscript. FN, JM helped in the semi-field experiments in blood feeding mosquitoes to different vertebrates. FB, PS, MG, HMF, SAB, KW supported in spectrometry and data analysis, taught EPM to apply supervised machine learning techniques for big data analysis and reviewed the manuscript. All authors read and approved the final manuscript.

## Funding

This research was supported by the Wellcome trust Intermediate Fellowship in Public Health & Tropical Medicine awarded to FOO (Grant No. WT102350/Z/13/Z) and a Howard Hughes Medical Institute (HHMI)-Gates International Research Scholarship awarded to FOO (Grant No. OPP1099295). EPM, FB and DJS were also supported by MRC (Grant No. MR/P025501/1). SAM was also supported by Wellcome Trust Masters Fellowship in Tropical Medicine & Hygiene (Grant No. 212633/Z/18/Z).

## Availability of data and materials

All data for this study will be available upon request.

## Ethics approval and consent to participate

Ethical approval for the study was obtained from Ifakara Health Institute Institutional Review Board (Ref. IHI/IRB/EXT/No: 005-2018), and from the Medical

Research Coordinating Committee (MRCC) at the National Institutes of Medical Research (NIMR), Ref: NIMR/HQ/R.8c/Vol.II/880.

## Consent for publication

Permission to publish the work was also obtained from NIMR (NIMR/HQ/P.12 VOL XXVI/77).

## Competing interests

The authors declare that they have no competing interests.

## Author details

<sup>1</sup> Environmental Health and Ecological Science Thematic Group, Ifakara Health Institute, Morogoro, Tanzania. <sup>2</sup> Institute for Disease Modeling, Bellevue, WA 98005, USA. <sup>3</sup> School of Chemistry, University of Glasgow, Glasgow G12 8QQ, UK. <sup>4</sup> Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, Glasgow G12 8QQ, UK. <sup>5</sup> School of Public Health, University of Witwatersrand, Johannesburg, South Africa.

Received: 27 March 2019 Accepted: 25 May 2019

Published online: 30 May 2019

## References

- WHO. Global technical strategy for malaria 2016–2030. Geneva: World Health Organization; 2015. [http://apps.who.int/iris/bitstream/10665/1176712/1/9789241564991\\_eng.pdf?ua=1](http://apps.who.int/iris/bitstream/10665/1176712/1/9789241564991_eng.pdf?ua=1). Accessed 25 Mar 2019.
- WHO. Malaria surveillance, monitoring & evaluation: a reference manual. Geneva: World Health Organization; 2018. <http://www.who.int/malaria/publications/atoz/9789241565578/en/>. Accessed 25 Mar 2019.
- MacDonald G. Epidemiological basis of malaria control. *Bull World Health Organ.* 1956;15:613–26.
- Tirados I, Costantini C, Gibson G, Torr SJ. Blood feeding behaviour of the malarial mosquito *Anopheles arabiensis*: implications for vector control. *Med Vet Entomol.* 2006;20:425–37.
- Takken W, Verhulst NO. Host preferences of blood-feeding mosquitoes. *Annu Rev Entomol.* 2013;58:433–53.
- Kaindoa EW, Matowo NS, Ngowo HS, Mkandawile G, Mmbando A, Finda M, et al. Interventions that effectively target *Anopheles funestus* mosquitoes could significantly improve control of persistent malaria transmission in south-eastern Tanzania. *PLoS One.* 2017;12:e0177807.
- Day JF, Edman JD, Scott TW. Reproductive fitness and survivorship of *Aedes aegypti* (Diptera: Culicidae) maintained on blood, with field observations from Thailand. *J Med Entomol.* 1994;31:611–7.
- Ranson H, Guesson RN, Lines J, Moiroux N, Nkuni Z, Corbel V. Pyrethroid resistance in African anopheline mosquitoes: what are the implications for malaria control? *Trends Parasitol.* 2011;27:91–8.
- Mahande A, Masha F, Mahande J, Kweka E. Feeding and resting behaviour of malaria vector, *Anopheles arabiensis* with reference to zoophyloxaxis. *Malar J.* 2007;6:100.
- Russell TL, Govella NJ, Azizi S, Drakeley CJ, Kachur SP, Killeen GF. Increased proportions of outdoor feeding among residual malaria vector populations following increased use of insecticide-treated nets in rural Tanzania. *Malar J.* 2011;10:80.
- Ngowo HS, Kaindoa EW, Matthiopoulos J, Ferguson HM, Okumu FO. Variations in household microclimate affect outdoor-biting behaviour of malaria vectors. *Wellcome Open Res.* 2017;2:102.
- Monroe A, Moore S, Koenker H, Lynch M, Ricotta E. Measuring and characterizing night time human behaviour as it relates to residual malaria transmission in sub-Saharan Africa: a review of the published literature. *Malar J.* 2019;18:6.
- Beier JC, Perkins PV, Wirtz RA, Koros J, Diggs D, Gargan TP, et al. Bloodmeal identification by direct enzyme-linked immunosorbent assay (ELISA), tested on *Anopheles* (Diptera: Culicidae) in Kenya. *J Med Entomol.* 1988;25:9–16.
- Kent RJ, Norris DE. Identification of mammalian blood meals in mosquitoes by a multiplexed polymerase chain reaction targeting cytochrome B. *Am J Trop Med Hyg.* 2005;73:336–42.
- Boorman J, Mellor PS, Boreham PFL, Hewett RS. A latex agglutination test for the identification of blood meals of Culicoides (Diptera: Ceratopogonidae). *Bull Entomol Res.* 1977;67:305–11.



16. Gomes LAM, Duarte R, Lima DC, Diniz BS, Serrão ML, Labarthe N. Comparison between precipitin and ELISA tests in the bloodmeal detection of *Aedes aegypti* (Linnaeus) and *Aedes fluviatilis* (Lutz) mosquitoes experimentally fed on feline, canine and human hosts. *Mem Inst Oswaldo Cruz*. 2001;96:693–5.
17. Niare S, Berenger JM, Dieme C, Doumbo O, Raoult D, Parola P, et al. Identification of blood meal sources in the main African malaria mosquito vector by MALDI-TOF MS. *Malar J*. 2016;15:87.
18. Niare S, Almeras L, Tandina F, Yssouf A, Bacar A, Toilibou A, et al. MALDI-TOF MS identification of *Anopheles gambiae* Giles blood meal crushed on Whatman filter papers. *PLoS One*. 2017;12:e0183238.
19. Tandina F, Laroche M, Davoust B, K Doumbo O, Parola P. Blood meal identification in the cryptic species *Anopheles gambiae* and *Anopheles coluzzii* using MALDI-TOF MS. *Parasite*. 2018;25:40.
20. Beier JC, Perkins PV, Koros JK, Onyango FK, Gargan TP, Wirtz RA, et al. Malaria sporozoite detection by dissection and ELISA to assess infectivity of afrotropical *Anopheles* (Diptera: Culicidae). *J Med Entomol*. 1990;27:377–84.
21. Chow E, Wirtz RA, Scott TW. Identification of blood meals in *Aedes aegypti* by antibody sandwich enzyme-linked immunosorbent assay. *J Am Mosq Control Assoc*. 1993;9:196–205.
22. 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. *Am J Trop Med Hyg*. 2009;81:622–30.
23. Lambert B, Sikulu-Lord MT, Mayagaya VS, Devine G, Dowell F, Churcher TS. Monitoring the age of mosquito populations using near-infrared spectroscopy. *Sci Rep*. 2018;8:5274.
24. 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. *Sci Rep*. 2018;8:9590.
25. Krajacich BJ, Meyers JI, Alout H, Dabiré RK, Dowell FE, Foy BD. Analysis of near infrared spectra for age-grading of wild populations of *Anopheles gambiae*. *Parasit Vectors*. 2017;10:552.
26. 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 infra-red spectroscopy. *Parasit Vectors*. 2013;6:298.
27. Milali MP, Sikulu-Lord MT, Kiware SS, Dowell FE, Corliss GF, Povinelli RJ. Age grading *An. gambiae* and *An. arabiensis* using near infrared spectra and artificial neural networks. *bioRxiv* 490326. 2018.
28. 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. *Parasit Vectors*. 2010;3:49.
29. 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:e0004759.
30. Fernandes JN, dos Santos LMB, Chouin-Carneiro T, Pavan MG, Garcia GA, David MR, et al. Rapid, noninvasive detection of Zika virus in *Aedes aegypti* mosquitoes by near-infrared spectroscopy. *Sci Adv*. 2018;4:eaat0496.
31. Esperança PM, Blagborough AM, Da DF, Dowell FE, Churcher TS. Detection of *Plasmodium berghei* infected *Anopheles stephensi* using near-infrared spectroscopy. *Parasit Vectors*. 2018;11:377.
32. Ferreira Maia M, Kapulu M, Muthui M, Wagah M, Ferguson H, Dowell F, et al. Detection of malaria in insectary-reared *Anopheles gambiae* using near-infrared spectroscopy. *Malar J*. 2019;18:85.
33. Gonzalez-Jimenez M, Babayan SA, Khazaeli P, Doyle M, Walton F, Reedy E, et al. Prediction of malaria mosquito species and population age structure using mid-infrared spectroscopy and supervised machine learning. *Wellcome Open Res*. 2019;4:76.
34. Bruker Optics. ALPHA II—the Compact FTIR spectrometer for any industry. 2019. <https://www.bruker.com/products/infrared-near-infrared-and-raman-spectroscopy/ft-ir-routine-spectrometers/alpha/overview.html>. Accessed 26 Mar 2019.
35. Bruker Optics. ALPHA: the very compact and smart FTIR spectrometer. 2017.
36. Bruker Optics. OPUS spectroscopy software. 2019. <https://www.bruker.com/products/infrared-near-infrared-and-raman-spectroscopy/opus-spectroscopy-software.html>. Accessed 26 Mar 2019.
37. Siria DJ, Batista EPA, Opiyo MA, Melo EF, Sumaye RD, Ngowo HS, et al. Evaluation of a simple polytetrafluoroethylene (PTFE)-based membrane for blood-feeding of malaria and dengue fever vectors in the laboratory. *Parasit Vectors*. 2018;11:236.
38. Mayagaya VS, Nkwengulila G, Lyimo IN, Kihonda J, Mtambala H, Nkonyani H, et al. The impact of livestock on the abundance, resting behaviour and sporozoite rate of malaria vectors in southern Tanzania. *Malar J*. 2015;14:17.
39. Mukabana RW, Takken W, Seda P, Killeen GF, Hawley WA, Knols BGJ. Extent of digestion affects the success of amplifying human DNA isolated from blood meals of *Anopheles gambiae* (Diptera: Culicidae). *Bull Entomol Res*. 2002;92:233–9.
40. Pedregosa F, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12:2825–30.
41. Gauglitz G, Moore DS. *Handbook of Spectroscopy*. 2nd Ed. 2014.
42. Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. *Cancer Inform*. 2006;2:59–77.
43. Okser S, Pahikkala T, Airola A, Salakoski T, Ripatti S, Aittokallio T. Regularized machine learning in the genetic prediction of complex traits. *PLoS Genet*. 2014;10:e1004754.
44. Leung MKK, Delong A, Alipanahi B, Frey BJ. Machine learning in genomic medicine: a review of computational problems and data sets. *Proc IEEE*. 2016;104:176–97.
45. Babayan SA, Orton RJ, Streicker DG. Predicting reservoir hosts and arthropod vectors from evolutionary signatures in RNA virus genomes. *Science*. 2018;362:577–80.
46. Prosperi MC, Di Giambenedetto S, Fanti I, Meini G, Bruzzone B, Callegaro A, et al. A prognostic model for estimating the time to virologic failure in HIV-1 infected patients undergoing a new combination antiretroviral therapy regimen. *BMC Med Inform Decis Mak*. 2011;11:40.
47. Yu W, Liu T, Valdez R, Gwinn M, Khoury MJ. Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes. *BMC Med Inform Decis Mak*. 2010;10:16.
48. Chen M, Hao Y, Hwang K, Wang L, Wang L. Disease prediction by machine learning over big data from healthcare communities. *IEEE Access*. 2017;5:8869–79.
49. Ngo KA, Kramer LD. Identification of mosquito bloodmeals using polymerase chain reaction (PCR) with order-specific primers. *J Med Entomol*. 2003;40:215–22.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.