2022-08

Effects of sample preservation methods and storage duration on the performance of mid-infrared spectroscopy for predicting the age of malaria vectors

Mgaya, Jacqueline

NM-AIST

https://doi.org/10.58694/20.500.12479/1517

EFFECTS OF SAMPLE PRESERVATION METHODS AND STORAGE DURATION ON THE PERFORMANCE OF MID-INFRARED SPECTROSCOPY FOR PREDICTING THE AGE OF MALARIA VECTORS

Jacqueline N. Mgaya

A Dissertation Submitted in Partial Fulfilment of the Requirements for the Degree of Master of Science in Public Health Research of the Nelson Mandela African Institution of Science and Technology

Arusha, Tanzania

August, 2022

ABSTRACT

The study assessed the effects of different preservation methods and storage durations on the performance of mid-infrared spectroscopy for age-grading malaria transmitting mosquitoes. Laboratory-reared *Anopheles arabiensis* (N=3,681) were collected as 5 or 17-day olds and killed with ethanol then preserved using either silica desiccant at 5°C, freezing at -20°C, or absolute ethanol at room temperatures. For each preservation method, the mosquitoes were divided into three groups and stored for 1, 4 or 8 weeks, then scanned using the mid-infrared spectrometer. Supervised machine learning classifiers were trained with the infrared spectra, and used to predict the mosquito ages. The classification of mosquito ages (as 5 or 17-day old's) was most accurate when the samples used to train the models and samples being tested were preserved the same way or stored for equal durations. However, when the test and training samples were handled differently, the classification accuracies declined significantly. Support vector machine classifiers (SVM) trained using spectra of silica-preserved mosquitoes achieved 95% accuracy when predicting the ages of other silica-preserved mosquitoes, but this declined to 72% and 66% when age-classifying mosquitoes preserved using ethanol and freezing. Similarly, models trained on one-week stored samples had declining accuracies of 97%, 83% and 72% when predicting ages of mosquitoes stored for 1, 4 or 8 weeks respectively. When using mid-infrared spectroscopy and supervised machine learning to age-grade mosquitoes, the highest accuracies are achieved when the training and test samples are preserved the same way and stored for the same durations. Protocols for infrared-based entomological studies should emphasize standardization of sample-handling procedures.

DECLARATION

I, Jacqueline Nicholaus Mgaya, do hereby declare to the Senate of Nelson Mandela African Institution of Science and Technology that this dissertation titled “Effects of sample preservation methods and duration of storage on the performance of mid-infrared spectroscopy and ML for predicting the age of malaria vectors” is my original work and has not been submitted for consideration of a similar degree award in any other University other than NM-AIST.

Jacqueline N. Mgaya

03.08.2022

Date

.

COPYRIGHT

CERTIFICATION

This is to certify this research is submitted as my own as a partial fulfilment of the requirements for the award of Master of Science in Public Health Research of the Nelson Mandela African Institution of Science and Technology.

03.08.2022
Date

Dr.Fredros Okumu
Adjuct Professor, NMAIST

03.08.2022
Date

Dr.Sr John Mary Vianney, PhD
Lecturer, NMAIST

ACKNOWLEDGEMENTS

I express my sincere gratitude to the almighty GOD who made it possible for this work to come to completion. He took care of everything that would have stopped me in my tracks and strengthened me even through my most difficult times. Throughout the writing of this dissertation I have received a great deal of support and assistance.

I respect and thank Prof. Fredros Okumu who has been the ideal supervisor for this project. His wise counsel, insightful criticisms, and patient encouragement assisted the writing of this project in innumerable ways. Your insightful feedback pushed me to sharpen my thinking and brought my work to a higher level. I also thank and appreciate Dr. Sr. John Mary Vianney for her guidance and support throughout the research process. I also want to acknowledge Mr. Emmanuel Mwanga whose expertise was invaluable in formulating the research questions and in data analysis. My colleagues from Ifakara Health Institute (Ifakara), notably the OMC team in Ifakara have my profound thanks for the support they offered throughout data collection, without which the study would not have succeeded

I am indebted to the Ifakara Health Insititute for the financial support they offered in my research project work. I would also like to express my gratitude towards my family, most especially Rosemary Mgaya (My mother, mentor and heroine all rolled into one), my mom Risper, sisters (Lulu, Grace and Naelijwa) and brother Clinton Mgaya whose steadfast support for this project was greatly needed, offered and deeply appreciated.

I am grateful to my best friends Ms.Naomi Urio (Booche), Ms. Jeyla Maulid, Ms. Angela Peter and Ms. Gladness Boniphace for their love, support and being shoulders to lean on, my love for you all can never be quantified.

My special thanks goes to Dr. Shubis Kafuruki and Mrs. Cecilia Francis for their guidance, advice and support. To my classmates Mr. Amos Ngonzi, Mr. Lameck Pashet, Ms. Miriam Kweka, and the rest of the Cohort 3 class of 2020 for their outermost love, care, cooperation and consultation they offered and for being such a blessing in my school life

Finally, I also place in record, my sense of gratitude to one and all, who directly or indirectly participated in this study. May God bless you all.

DEDICATION

This work is dedicated to my Late Mother Rosemary Nicholaus A. Mgaya who encouraged it, supported it but won’t read it. Rosemary, you could not witness my success because death defeated you.

To my Father Nicholaus A. Mgaya who gave me the foundation of education. Ever since then, I have been able to appreciate the value of reading and lifelong learning.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ATR-FTIR	Attenuated Total Reflectance-Fourier transform infrared spectrometer
MIRS	Mid-infrared Spectroscopy
NIRS	Near-infrared Spectroscopy
ML	Machine Learning
KNN	k-nearest neighbors
LR	Logistic regression
SVM	Support vector machine
RF	Random forest
GB	Gradient boosting
ET	Extra-trees classifier
BGC	Bagging classifier
PCR	Polymerase Chain Reaction
IHI	Ifakara Health Institute
IHI-ERB	Ifakara Health Institute Ethical Review Board
NIMR	National Institute of Medical Research

NM-AIST	Nelson Mandela African Institution of Science and Technology
WHO	World Health Organization

CHAPTER ONE

INTRODUCTION

1.1 Background of the problem

Malaria disease remains a public health challenge and it has been resurging despite the efforts made in the fight against it. In 2019 the World Health Organization (WHO), projected 215 million cases with 409000 deaths yearly (WHO, 2020). Most deaths occurring to pregnant women and children under five years of age (WHO, 2019). In 2020, malaria deaths increased by 12% as compared to 2019, reaching a global total of ~ 627000, nearly all of which occurred in sub-Saharan Africa (WHO, 2021). The main methods for controlling the disease currently include improved case management with artemisinin-based combination therapies (ACTs) and vector control with either insecticide-treated mosquito nets (ITNs) or indoor residual spraying (IRS). For over two decades, these tools have been the mainstay of malaria control, contributing significantly to the reductions in cases and deaths (Bhatt *et al.*, 2015).

In addition to case management and vector control, the World Health Organization (WHO), in their 2016-2030 Global Technical Strategy, recommended that improved surveillance should also be included as a core component of malaria control (WHO, 2015). Countries should therefore adopt effective and scalable approaches for surveillance and strategically deploy these across various epidemiological strata. With respect to malaria vectors, improved surveillance may include measures to better understand the dynamics and pathogen transmission activity of *Anopheles* mosquitoes; as well as measures to monitor insecticide resistance and assess the performance of key interventions such as ITNs and IRS (WHO, 2018). Unfortunately, most countries still lack the adequate capacity for vector surveillance and intervention monitoring (Russell *et al.,* 2020). Moreover, current vector control methods are greatly threatened by multiple factors notably insecticide resistance (Kleinschmidt & Rowland, 2021; Strode *et al.,* 2014), human and mosquito behavioral factors (Finda *et al.,* 2019; Monroe *et al.,* 2019; Okumu & Finda, 2021; Sangbakembi-Ngounou *et al.,* 2022), and limited durability of ITNs (Lorenz *et al.,* 2020), among others. Furthermore, African malaria vectors express varied ecological and biological traits, making their detailed surveillance challenging yet critical to optimize control. For example, understanding the mosquito blood–feeding preferences may illuminate the degree to which certain species can carry human pathogens (Kiszewski *et al.,* 2004). Similarly, knowing the age structure of mosquito populations can inform evaluations of the impact of vector control programmes (Silver, 2008); since mosquitoes must attain a certain age to allow maturation of the malaria parasite inside their guts. *Plasmodium falciparum* generally requires more than 10 days incubation period inside their vectors before they become infectious (Guissou *et al.,* 2021; Oakley *et al.,* 2018; Ohm *et al.,* 2018).

Mosquito age-grading previously relied on ovary dissections (Silver, 2008) or, in a few instances, the use of transcriptional profiling (Cook *et al.,* 2006). However, these techniques are laborious, subjective and not optimal for field settings (Hugo *et al.,* 2014). Emerging techniques in spectroscopy have been considered to address these limitations since they can be performed quickly in dry laboratories without expensive reagents or replacement parts (Goh *et al.,* 2021). Both near-infrared spectroscopy (NIRS) and mid-infrared spectroscopy (MIRS) have been demonstrated to be effectively distinguish between mosquito species based on their biochemical components such as proteins, lipids, and carbohydrates (Siria *et al.,* 2022). The techniques can also be used for other entomological assessments such as mosquito age-grading (Goh *et al.,* 2021; Sikulu *et al.,* 2010; Siria *et al.,* 2022), studying blood-feeding histories (Mwanga *et al.,* 2019). The techniques can also be used to detect chronological age grading of house flies (Perez-Mendoza *et al.,* 2002) and detection of pathogens infections in mosquitoes (Fernandes *et al.,* 2018). These approaches are quicker and less expensive compared to alternatives such as polymerase chain reaction (PCR).

However, since the techniques are still in early-stage development, there are not yet any standardized procedures for handling mosquito samples. For NIRS–specific approaches, past studies have investigated the influence of mosquito physiological states (Ntamatungiro *et al.,* 2013), and also shown that the technique can be used with mosquito samples preserved using different methods (Dowell *et al.,* 2011; Sikulu *et al.,* 2011). Such investigations should be expanded to include mid-infrared spectroscopy (MIRS) applications; and possibly also tested on field-collected mosquitoes, which may have greater variability. In previous applications, mosquitoes scanned on the MIR spectrometer were mostly preserved by drying on silica gel (silicon dioxide) prior to scanning (Jiménez *et al.,* 2019; Mwanga *et al.,* 2019; Mwanga *et al.,* 2019; Siria *et al.,* 2022). Other methods for preserving mosquito samples include RNA*later*® (Ambion, Inc., Austin, TX), ethanol, freezing, DNA-RNA shield, and liquid nitrogen (Gorokhova, 2005; Hugo *et al.,* 2010; Sikulu *et al.,* 2011), but these have not been tested on MIRS–based mosquito applications. Because of this, it is not known whether variations in preservation methods might affect the performance of these techniques or if the chemical components of the mosquito could degrade when stored in different preservation methods for a long period of time.

1.2 Statement of the problem

There is a growing evidence from other studies that the mid-infrared spectroscopy and machine learning approach works in detecting parasites in dried blood spots (DBS), blood meals and age of mosquitoes. The approach is reliable, does not require reagents and is cost-effective for entomological surveillance (Jiménez *et al.,* 2019; Mwanga *et al.,* 2019). However different researchers are using different sample handling techniques, thus there is lack of standardization. The objective of this study was therefore to investigate the effects of different preservation methods on the performance of a previously-implemented mid-infrared based approach for age-grading female malaria vectors. In addition, the study evaluated whether varying the duration of sample storage could influence the performance of age-grading techniques.

1.3 Rationale of the study

Sample handling procedures and quality have great impact on the results when using mid-infrared spectroscopy. Broadly, there is a need for improved approaches for mosquito surveillance that are low cost and can be deployed even at the district level without requiring specialized skills or reagents. This research addresses that gap by demonstrating the most optimal sample handling procedures for MIRS and ML, which have been demonstrated to be potentially effective next-generation surveillance tools. This study also evaluated whether varying the duration of sample storage could influence the performance of age-grading techniques.

1.4 Research Objectives

1.4.1 General Objective

To establish optimal storage and preservation methods for mosquito samples on the performance of Mid-Infrared spectroscopy and Machine learning.

1.4.2 Specific Objectives

(i) Reviewing existing sample handling techniques used in the application of MIR and ML based approaches for malaria surveillance

(ii) Assessing effects of sample storage duration on performance of MIR and ML in estimating the age of malaria vectors

(iii) Assess sample preservation methods on performance of MIR and ML for estimating age of malaria vectors

1.5 Research Questions

(i) What is the effective time period for mosquito sample preservation?

(ii) What are the appropriate storage conditions for the accuracy of MIR spectroscopy and Machine learning?

(iii) What are the required sample preservation conditions for best performance of MIR and Machine learning predictions?

1.6 Significance of the study

This study provides robust information on the effects of different preservation methods and storage on the performance of mid-infrared spectroscopy and machine learning techniques when age grading mosquitoes.

1.7 Delineation of the study

There is a growing evidence from other studies that the mid-infrared spectroscopy and machine learning approach works and is a good entomological surveillance tool. Currently, different researchers are using different sample handling techniques and thus more information and data is needed on sample handling techniques for standardization.

CHAPTER TWO

LITERATURE REVIEW

2.1 Burden of the disease

Malaria is estimated to account for 17% of the global burden of infectious diseases. It is predicted that almost half of the world's population is at risk of malaria which is a primary cause of morbidity and mortality in Sub- Saharan Africa (SSA) (WHO & UNICEF, 2017). Of all cases and deaths accounting to malaria globally, 95% were recorded in sub-Saharan Africa, in which the United Republic of Tanzania was amongst the top 10 countries contributing to cases and death tolls (WHO, 2020). The increase in 2020 was associated with disruption to services during the COVID-19 pandemic.

Many efforts to control malaria have been focused on the development and deployment of vector control strategies. Efforts to control the disease, primarily include improved case management with artemisinin-based combination therapies (ACTs) (Bhatt *et al.*, 2015). Other vector control strategies in reducing malaria transmission by targeting survival rates of mosquitoes have been suggested, with one of the methods being insecticide-treated mosquito nets (ITNs, Indoor residual spraying (IRS), Larval source management (LSM), swarm sprays, spatial repellents (Tizifa *et al.,* 2018). In addition, monitoring and evaluation of malaria programmes as well as the access to diagnostic facilities and improved treatment have also been incorporated in the reduction of malaria strategies (Cibulskis *et al.,* 2016; Farlow *et al.,* 2020; WHO & UNICEF, 2017). Such interventions have effectively reduced the risk of infection in the population, by decreasing biting rates and thus preventing mosquitoes from surviving the parasite development stage and becoming infectious (Koella *et al.*, 2009). This remarkable impact has contributed significantly to the gains since 2020 and has led to the reduction of the global malaria burden (Bhatt *et al.*, 2015).

However, these efforts are greatly threatened by multiple factors notably insecticide resistance (Kleinschmidt & Rowland, 2021; Strode *et al.,* 2014), changes in human and mosquito behavioural factors (Finda *et al.,* 2019; Monroe *et al.,* 2019; Okumu & Finda, 2021), reduced efficacy of LLINS and IRS due to insecticide resistance (Kleinschmidt & Rowland, 2021), changes in ecological structure, user acceptance of ITNs and cost-effectiveness and limited durability of ITNs (Lorenz *et al.*, 2020) among others (Coleman *et al.,* 2017; Hemingway *et al.,* 2016). These setbacks have contributed to mosquito survival and slowed the momentum towards malaria elimination across sub-Saharan Africa. Malaria elimination is difficult without effective mosquito control and it is therefore imperative for key stakeholders to adopt a multifaceted approach in combating insecticide resistance, policies and implementation pathways so that we can prevent the increase of malaria cases and deaths (Hemingway *et al.,* 2016).

2.2 Malaria Surveillance and Elimination

Surveillance entails tracking of the disease and acting based on the data collected. Many Sub-Saharan African countries that have a malaria high malaria burden have weak surveillance systems and are not fully capable of assessing disease trends, thus posing the difficulty to monitor, evaluate programs as well as preventing outbreaks and resurgences (Hammond & Galizi, 2017). Improved surveillance was recommended by WHO in their 2016-2030 Global Technical Strategy, to be included as a core component of malaria programs (WHO, 2015). To support vector control, countries should therefore adopt effective and scalable approaches for surveillance. These may include measures to better understand the vector dynamics and pathogen transmission, monitor insecticide resistance and assess the performance of key interventions such as ITNs and IRS (WHO, 2018). Most countries still lack the adequate capacity for vector surveillance and intervention monitoring (Russell *et al.,* 2020). Furthermore, African malaria vectors have varied ecological and biological traits, making surveillance even more crucial to optimize control. For example, understanding the mosquito blood–feeding preferences may illuminate the degree to which certain species can carry human pathogens (Kiszewski *et al.,* 2004).

2.3 Mosquito age grading

It is important to measure the survival of female mosquitoes as it is an important biological determinant in the essence of understanding malaria transmission (Ferguson*et al.,* 2012). Knowing the age structure of mosquito populations can inform evaluations of the impact of vector control programs; since mosquitoes must attain a certain age to allow maturation of the malaria parasite inside their guts. *Plasmodium falciparum* generally requires more than 10 days incubation period inside their vectors before they become infectious (Guissou *et al.,* 2021; Ohm *et al.,* 2018). Similarly, literature suggests that older female mosquitoes can be potentially infectious, since they have lived to have more than one blood meal and therefore they are potential vectors in transmitting malaria pathogens (Cook *et al.,* 2008; Shaw *et al.,* 2020; Sinkins & O'Neill, 2000).

Mosquito species differ greatly in their ecological and biological behaviours and thus making vector surveillance crucial in the control of malaria transmission. Knowing the age structure of mosquitoes from wild populations is vital in evaluating the impact of vector control programmes (Kiszewski *et al.,* 2004). Therefore, it is essential to estimate the age of mosquitoes in order to assess the impact of the aforementioned vector control interventions. Mosquito age-grading previously relied on ovary dissections (Silver, 2008) or, in a few instances, the use of transcriptional profiling (Cook *et al.,* 2006). However, these techniques are laborious, subjective and not optimal for field settings (Hugo *et al.,* 2014). Emerging techniques in spectroscopy have been considered to address these limitations since they can be performed quickly in dry laboratories without expensive reagents or replacement parts (Goh *et al.,* 2021) . For MIRS, the process is based on the MIRS measurement of the amount of light absorbed by the mosquito cuticle. Considering cuticular composition varies with age, MIR spectra will be used to predict these features. These approaches are quicker, can be utilized on large set of mosquito samples and less expensive compared to alternatives such as polymerase chain reaction (PCR) (Bass *et al.,* 2007).

2.4 Differences between NIRS and MIRS

As one of the entomological surveillance and disease monitoring tools and vector control strategies, near- infrared spectroscopy (NIRS) have been considered to address these limitations due to their rapid spectra acquirement with minimum sample preparation methods (Johnson *et al.,* 2020). NIRS has been used in mosquito species differentiation based on their biochemical components such as proteins, lipids, water and carbohydrates (Dowell *et al.,* 2011), mosquito age grading (Sikulu *et al.,* 2010) and preservation of mosquito species (Sikulu *et al.,* 2011).

Subsequently, NIRS has also been used to differentiate *Plasmodium falciparum* infected mosquitoes from uninfected in the *Anopheles gambiae* mosquito vector species with an accuracy of 95% (Maia *et al.,* 2019). This approach however is composed of a few weak signals dominated by a combination of water which relies more on the physiological and environmental aspects of the mosquito vector other than characteristics such as age and species. In addition, NIRS approach has been limited in predicting age of wild mosquitoes' due to their ecological variability (Lambert *et al.,* 2018; Sikulu *et al.,* 2010). These drawbacks favoured the use of mid-infrared spectroscopy coupled with machine learning technique which was shown to have a better accuracy and speed in detecting the age and species of mosquitoes (Jiménez *et al.,* 2019; Siria*et al.,* 2022). Mid-infrared spectroscopy has proven to be better in performance than near-infrared spectroscopy in many ways. Firstly, it measures chemical compositions in biological samples in the mid-infrared region at 2500 - 25000nm. MIRS has also been reported to accurately estimate the age classes of mosquitoes responsible for spreading malaria up to two weeks of age, whilst NIRS- can define age up to seven days old with an estimated accuracy of about 80% (Mayagaya *et al.,* 2009). The MIRS bands are intense and are clearly defined since the fundamental bonds vibrate upon absorbance of specific wavelengths of the mid-infrared spectrum. Thus, making it easier to independently quantify the biochemical constituents (Jiménez *et al.,* 2019; Mwanga *et al.,* 2019).

Additionally, it has further been proved to be an alternative fast and cheap method in diagnosing *Wolbachia* infection for both laboratory and field *Aedes aegypti* mosquitoes with an accuracy of up to 97% (Khoshmanesh *et al.,* 2017). It has also been proved to be reliable in screening of malaria parasites in dried blood spots with 92% specificity in Tanzania, age grading and species identification of mosquitoes (Jiménez *et al.,* 2019) and detecting blood meal sources (Mwanga*et al.,* 2019). Subsequently, it has also been proven to be reliable in identifying species and age class of genetically *An. gambiae, An. arabiensis*, and *An. coluzzii* female mosquitoes (Siria*et al.,*2022). The approach was proved to be quicker and inexpensive than using polymerase chain reaction (PCR) which is currently considered as the gold standard (Mwanga *et al.,* 2019; Mwang a*et al.,* 2019). In general, little technical expertise is required to scan samples and high scanning throughputs are achieved with minimal reagent costs. Furthermore, it has been illustrated that, Mid-Infrared spectroscopy and machine learning approach can detect malaria parasites in the laboratory, having similar sensitivity to the PCR (Heraud *et al.,* 2019).

However, since the techniques are still in early-stage development, there are not yet any standardized procedures for handling mosquito samples. For NIRS – specific approaches, studies have investigated the influence of mosquito physiological states (Ntamatungiro *et al.,* 2013) and showed that the technique can be used with mosquito samples preserved using different methods (Dowell *et al.,* 2011; Sikulu *et al.,* 2011). Such investigations should be expanded to include mid-infrared spectroscopy applications, and possibly also get tested on field-collected mosquitoes, which often have greater variability.

2.5 Mosquito preservation methods

In previous applications, mosquitoes scanned on the MIR spectrometer were mostly preserved by drying on silica gel (silicon dioxide) prior to scanning as previously described (Jiménez *et al.,* 2019; Mwanga *et al.,* 2019; Mwanga *et al.*, 2019; Siria *et al.,* 2022). Apart from desiccation, other methods for preserving mosquito samples includeRNA*later*® (Ambion, Inc., Austin, TX), ethanol, freezing, DNA-RNA shield, and liquid nitrogen. These techniques have been widely used to store mosquito samples for dissections, DNA extraction, as well as genetic sequencing (Gorokhova, 2005; Hugo *et al.,* 2010; Sikulu *et al.,* 2011). Unfortunately, most of these preservation methods have not been tested on MIRS–based mosquito applications. Because of this, it is not yet known whether variations in preservation methods and duration of storage would alter the chemical compositions of the mosquitoes, which would therefore affect the performance of MIRS by predicting the age of mosquitoes with low prediction accuracy.

Additionally, for a broader range of applications, it is also necessary to test different mosquito sample preservation methods and storage durations, particularly for large-scale studies where samples are obtained from widely dispersed sampling sites and can be stored for various amounts of time before analysis.

CHAPTER THREE

MATERIALS AND METHODS

3.1 Study Area Description

This study was conducted in the VectorSphere Laboratory in Ifakara, South-eastern Tanzania.

3.2 Mosquitoes

Laboratory-reared *Anopheles arabiensis* females were used in this study. The mosquitoes were maintained in standard insectary conditions (27 ± 1°C, 70% humidity and a 12hr: 12hr light-dark cycle) at the Ifakara Health Institute's vector biology laboratory, the VectorSphere. They were fed on 10% glucose solution but not blood and were sampled at the ages of 5 and 17 days old post-emergence, to constitute two distinct age classes of young and old mosquitoes. A total of 3681 mosquitoes were used, including 1840 that were 5-day olds and 1841 that were 17-day olds.

3.3 Preservation and storage

Upon collection, mosquitoes were collected in disposable cups, anesthetized and killed using absolute ethanol for 30 minutes (Fig. 1). The mosquitoes of each age category were immediately packed in pools of ten in 2 ml micro-centrifuge tubes, and then preserved separately using three different techniques, namely: a.) silica gel desiccation at 5°C temperature (n = 1231), b.) Freezing at -20°C(n=1226) or c.) absolute ethanol at room temperature (n=1224). Desiccation over silica, being the method, most commonly used by the research team in previous entomological studies, was considered as the reference.

For each preservation method, mosquitoes were stored for 1, 4, or 8 weeks separately, before being scanned (Table 1). Prior to sample scanning, ethanol preserved samples were placed on paper towels to allow evaporation of the liquid. The procedure was also performed on mosquito samples that had been frozen to allow the moisture to evaporate (Fig. 1).

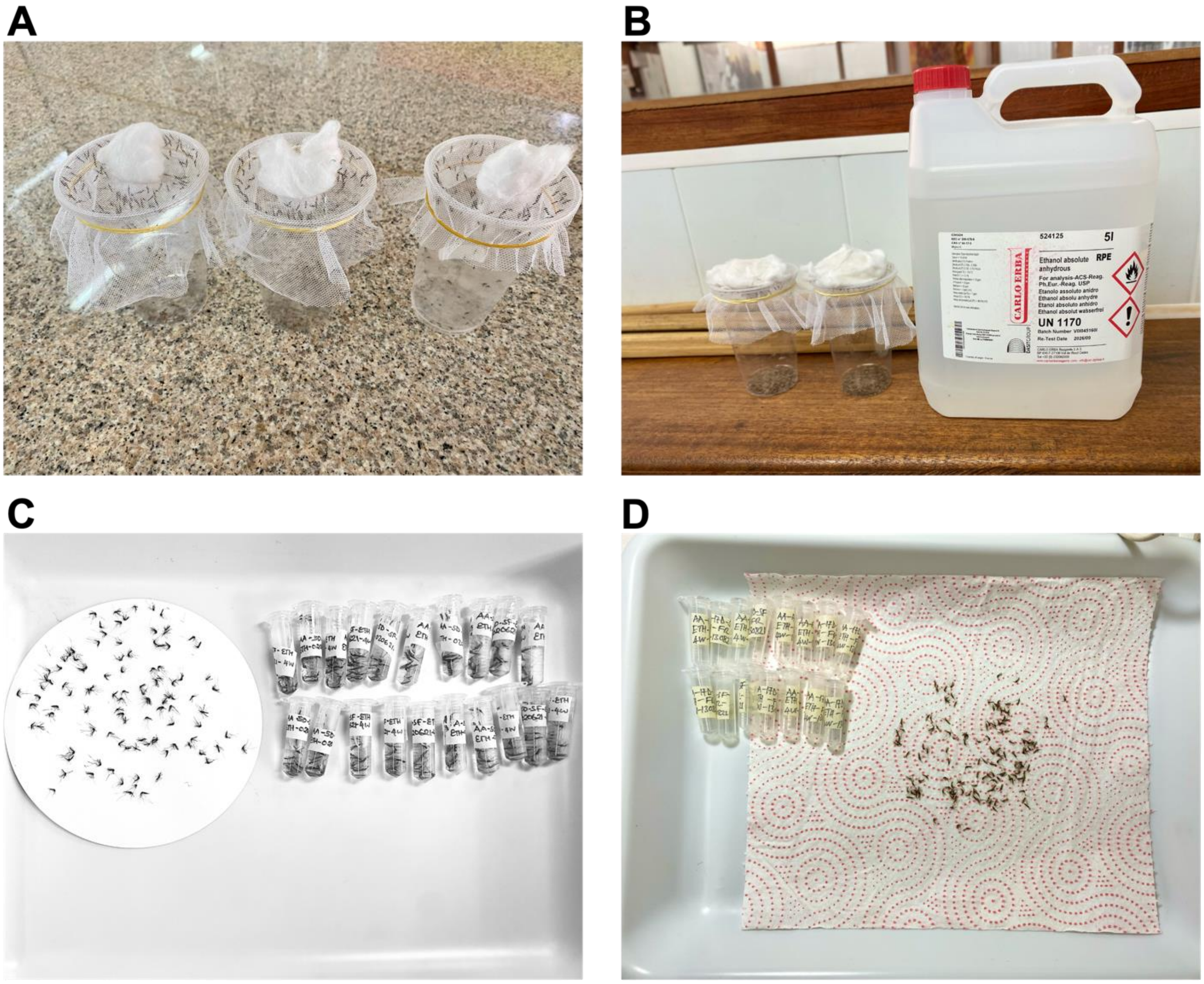

Figure 1: Mosquitoes collection, killing, packing and drying process before scanning

3.4 Mosquito Scanning

The heads and thoraces of the individual mosquitoes were scanned using attenuated total reflection-Fourier transform infrared (ATR-FTIR) ALPHA II spectrometer, as previously described (Jiménez *et al.,* 2019; Mwanga *et al.,* 2019; Mwanga *et al.,* 2019).

The mid-infrared spectra were recorded at 4000cm^{-1} to 400cm^{-1} frequencies and averaged from 32 co-added scans, with a spectral resolution of 2 cm^{-1}. Background scans were performed before starting the recordings of the spectra from mosquito samples, and thereafter repeated after every 20 individual sample scans. The scanning was done at Vector Biology laboratory, the VectorSphere at Ifakara Health Institute where the ALPHA II spectrometer is installed.

The proprietary Bruker–OPUS software version 7.5 was used to record and process the MIR spectra. At least 203 mosquitoes were used for each age (5 and 17 day-olds), each storage duration (1, 4 and 8 weeks), and each preservation method (silica, freezing and ethanol (Table 1).

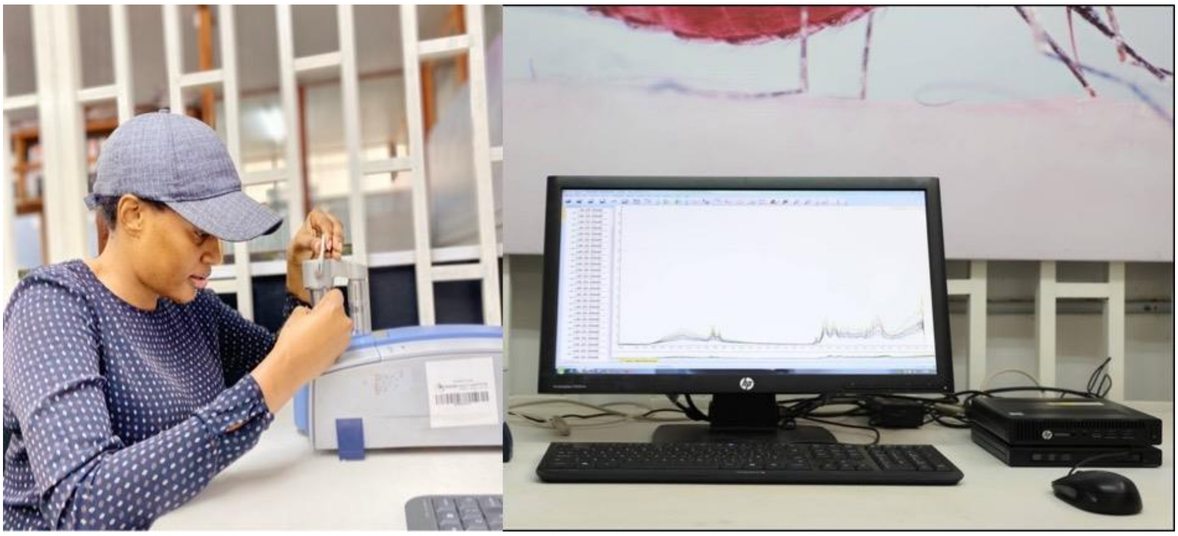

Figure 2: The scanning process of mosquito samples using attenuated total reflection Fourier transform infrared (ATR-FTIR) ALPHA II spectrometer

Table 1: The number of mosquitoes scanned for each age, preservation method and storage duration

Preservation method	No. mosquitoes scanned 5 days old	No. mosquitoes scanned 17 days old	Storage duration	Storage Temperature
Silica gel	203	204	1 Week	
	208	208	4 Weeks	5°C
	204	204	8 Weeks	
Ethanol (100%)	204	204	1 Week	
	203	203	4 Weeks	26°C
	205	205	8 Weeks	
Freezing	204	204	1 Week	
	204	204	4 Weeks	-20°C
	205	205	8 Weeks	

3.4 Data analysis

The spectral data were first cleaned to eliminate the bands associated with atmospheric water and CO_2 interference, then transferred to Python for supervised machine learning to predict mosquito ages for each preservation method and storage duration. During this pre-processing, up to 21 individual spectra, which either had significant atmospheric inferences from water (H_20) and carbon dioxide (CO_2), or abnormal spectral background noise were discarded from the main dataset as previously described (Jiménez *et al.*, 2019; Mwanga *et al.*, 2019).

The remaining 3660 spectra (1823 from 5 day-old mosquitoes and 1837 from 17 day-old mosquitoes) were further analysed in python version 3.8 using *Scikit-learn* version 0.23.2. The corresponding plots and visualizations were done using Seaborn version 0.11 and Matplotlib Version 3.3.2. Supervised machine learning approaches were used to train and predict the age of *An. arabiensis* preserved on different preservation techniques, and stored for different durations. The intensities of the entire MIR wavelengths were passed as a matrix of features and the mosquito ages (5 and 17 days) were used as labels. Features in the spectra dataset were rescaled to have a mean of 0 and a standard deviation of 1, bringing them into a similar scale without distorting the variations in the range of values. Seven machine learning models were evaluated, and the ones most suited for mosquito age classification were identified by comparing the baseline accuracies.

The evaluated classifiers included: (a) K-Nearest Neighbors (KNN), (b) Logistic regression (LR), (c) Support vector machine (SVM), (d) random forest (RF), (e) Gradient Boosting (GB), (f) extra-trees classifier (ET) and (g) Bagging classifier (BGC). Grid-search cross validation was used to further optimize the best-performing algorithm by turning its hyper-parameters. To evaluate and estimate the performance of the models on unseen data and avoid the risk of over-fitting, the training and test sets were evaluated using *K-fold* cross validation/rotational estimation.

For each analysis, the data was a sub-set, so that 80% was used to train the models, and the other 20% used as unseen data to evaluate the performance of the models. Data was combined from all storage durations (1, 4 and 8 weeks) for each preservation method when evaluating the influence of different preservation techniques on MIR-based age-classifications. Similarly, data from all preservation methods (silica gel, freezing, and ethanol) for each storage time were combined when assessing the effects of storage duration on the age-classifications.

Models trained with samples preserved in one way were evaluated for predicting the age classes of samples preserved the same way as well as for predicting the age of samples preserved with other methods. Similarly, models trained on samples stored for one week were tested for predicting age classes in samples kept for the same duration or longer periods (4 or 8 weeks).

3.5 Ethics and Consents

This study was approved by the Ifakara Health Institute's institutional review board (Ref: IHI/IRB/No: 24-2021) and the National Institute for Medical Research (NIMR/HQ/R.8a/Vol. IX/3557). The National Institute of Medical Research (NIMR) also granted permission to publish this work, with the reference number NIMR/HQ/P.12VOLXXXIV/77.

CHAPTER FOUR

RESULTS AND DISCUSSION

4.1 Results

4.1.1 Effect of preservation methods

The best performing model in predicting the ages of mosquitoes preserved by different methods was the support vector machine (SVM) (Fig. 3). The figure shows, the evaluation of different machine learning classifiers for predicting age for mosquito samples preserved in silica gel. The other three panels show confusion matrices with mosquito age predictions from an SVM classifier trained with silica-preserved mosquitoes and used to evaluate samples preserved in (B) silica gel, (C) ethanol and (D) freezing.

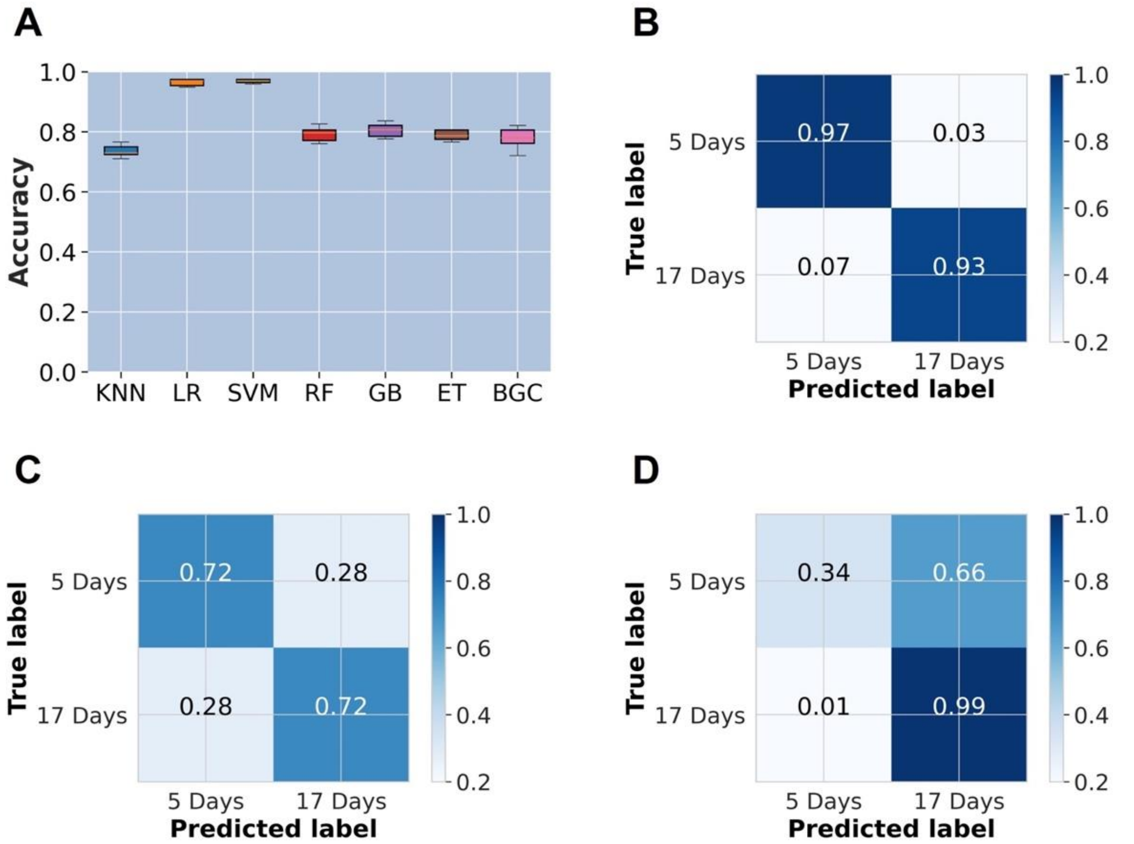

Figure 3: Evaluation of different machine learning classifiers for predicting age for mosquito samples preserved in silica gel, and the confusion matrices show mosquito age predictions from an SVM classifier

The SVM model was trained using data of mosquito samples preserved in silica gel and then used to predict the age of unseen mosquito samples stored in silica, ethanol, and freezing (Fig. 4). The model performed best when predicting mosquito ages of samples stored with silica gel, achieving a maximum classification accuracy of 95% on the unseen data (Fig. 4). When the same model was used to classify samples preserved in either ethanol or freezing by age, the prediction accuracy decreased to 72% and 66%, respectively (Fig.4). Declining accuracies were observed when the training set was changed from silica to either ethanol or freezing (Fig. 4). Here, the SVM models are trained with mid-infrared spectra of mosquitoes preserved using either silica (A), ethanol (B) or freezing (C) then used to predict age classes of samples preserved by either of the three methods. The figure also shows the results of the SVM models trained with mid-infrared spectra of mosquitoes stored for 1 week (E), 4 weeks (F) or 8 weeks (G) then used to predict ages of samples stored for either of the three durations. Reference samples are marked with stars. In all cases, the classification accuracy was highest when the training and test samples were handles the same way.

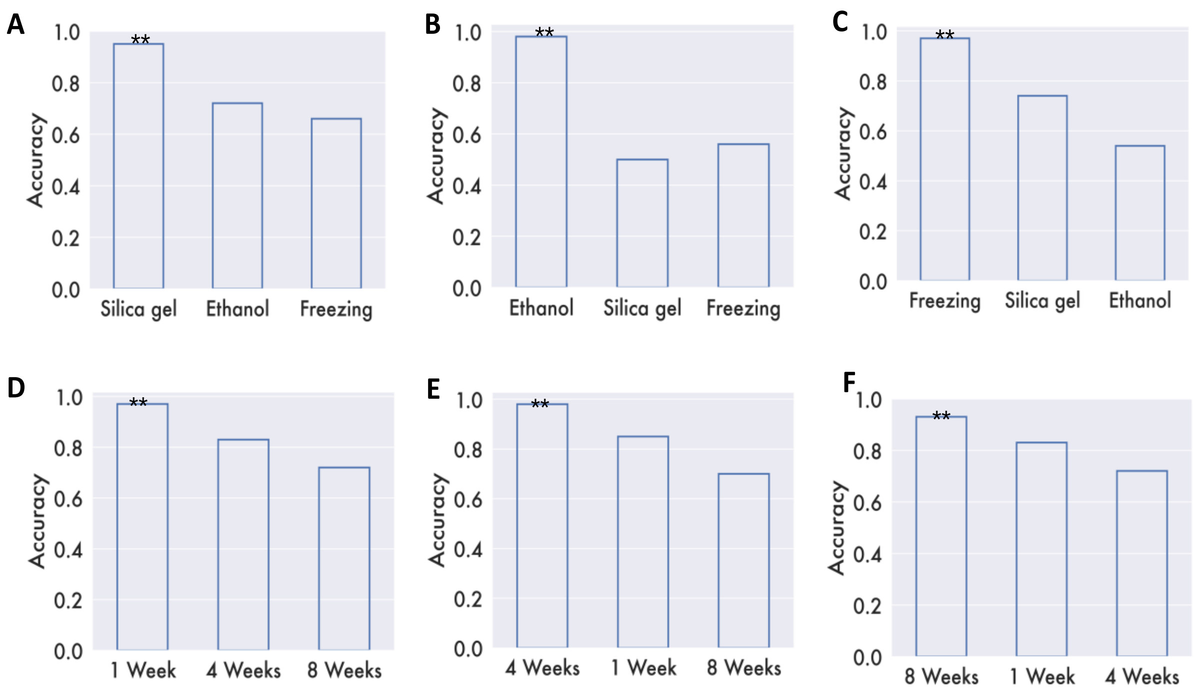

Figure 4: Bar plots showing the declines in classification accuracies when test and training datasets are handled similarly or differently for different preservation methods and storage durations

4.1.2 Effect of storage durations

With a maximum accuracy of 99%, the support vector machine (SVM) was again the best performing of the seven machine learning classifiers tested for determining the age of mosquitoes stored for varied time periods (Fig. 5). Evaluation of different machine learning classifiers for predicting the age of mosquito samples stored for one week (A). The other three panels show confusion matrices with prediction of mosquito ages from an SVM classifier trained with one-week samples and used to evaluate samples stored for 1 week (B), 4 weeks (C) and 8 weeks (D).

When trained with data from mosquitoes stored for one week, and used to predict the age of mosquitoes stored for the same period, the accuracy was 97%. However, the performance deteriorated with an increase in storage time of the test samples, such that the accuracies were 83% for samples stored for 4 weeks and 72% for 8 weeks stored samples (Fig. 4 & 5). The same trend of declining accuracies was observed when the training set was changed from 1 to either 4 weeks or 8 weeks stored samples (Fig. 4).

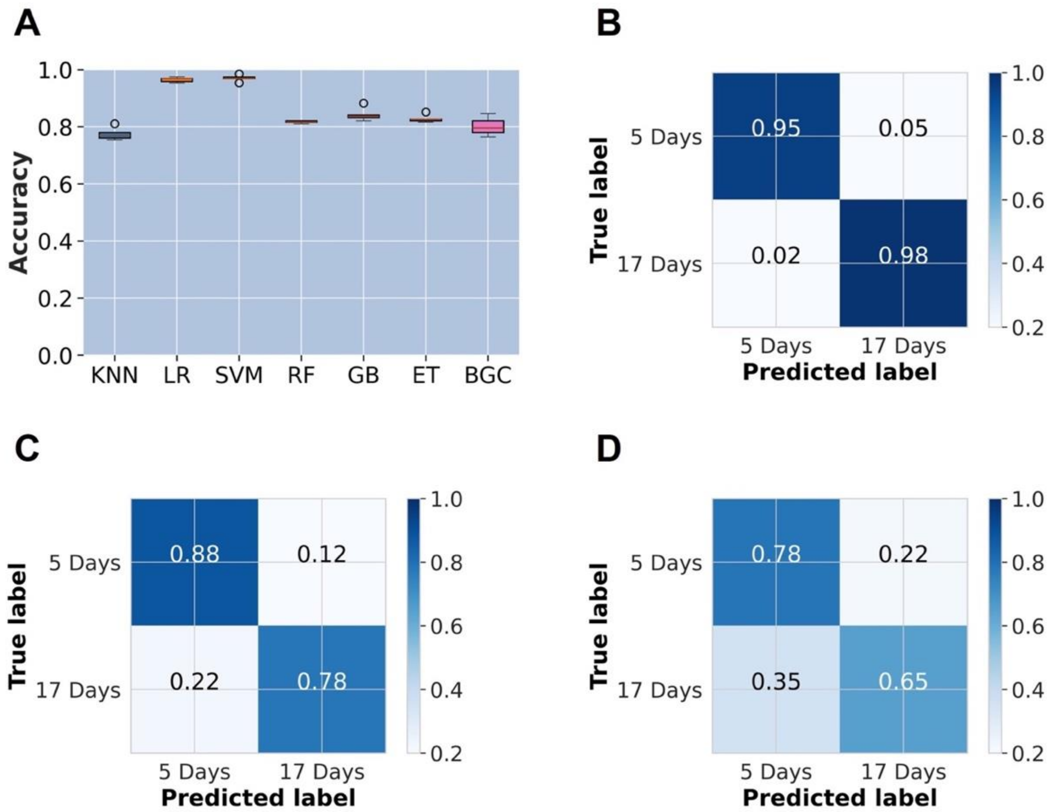

Figure 5: Evaluation of different machine learning classifiers for predicting age of mosquito samples stored for one week and the confusion matrices show mosquito age predictions from an SVM classifier

4.1.3 Using mosquitoes preserved in silica and stored for one week as a reference

Here, mosquitoes preserved using silica and stored for one week were used to train an SVM classifier to predict the ages of other mosquitoes preserved by different methods (silica gel, ethanol or freezing) and stored for different durations (1 4 and 8 weeks). Table 2 summarized the data from all the nine tests completed, which are the classification accuracies of a support vector machine model. The model was trained using mid-infrared spectra from mosquitoes preserved on silica desiccant and stored for one week, then used to age-grade other mosquitoes handled in the same or alternative ways. The resulting classification accuracies varied, greatly and were highest for mosquitoes that had been handled the same way (i.e. one-week storage on the silica desiccant (Fig. 6). Overall, mosquitoes stored in silica gel generally had the highest classification accuracy up to 4 weeks, whereas mosquitoes stored by freezing had the lowest classification accuracies (Table 2).

A decline in prediction accuracies was observed for samples stored in ethanol from 1 week to 8 weeks. The age-classification accuracies for 5 days old and 17 days old mosquitoes preserved using different methods and stored for different durations are summarized in Fig. 6. Confusion matrices showing prediction accuracies of mosquito ages from a standard SVM classifier trained with samples preserved in silica gel and stored for one week. Thereafter, it was used to predict age-classes of test samples handled the same way or differently. Silica-preserved samples are shown in panels A, B, C; ethanol-preserved samples on panels D, E, F and frozen samples on panels D, H and I respectively.

Table 2: Classification accuracies of an SVM model trained using mosquitoes preserved on silica desiccant and stored for one week

Preservation method	Storage duration	Classification accuracy
Silica gel	1 Week	100 %
	4 Weeks	88 %
	8 Weeks	61 %
Ethanol (100%)	1 Week	76 %
	4 Weeks	71 %
	8 Weeks	70 %
Freezing -20°C	1 Week	52 %
	4 Weeks	54 %
	8 Weeks	51 %

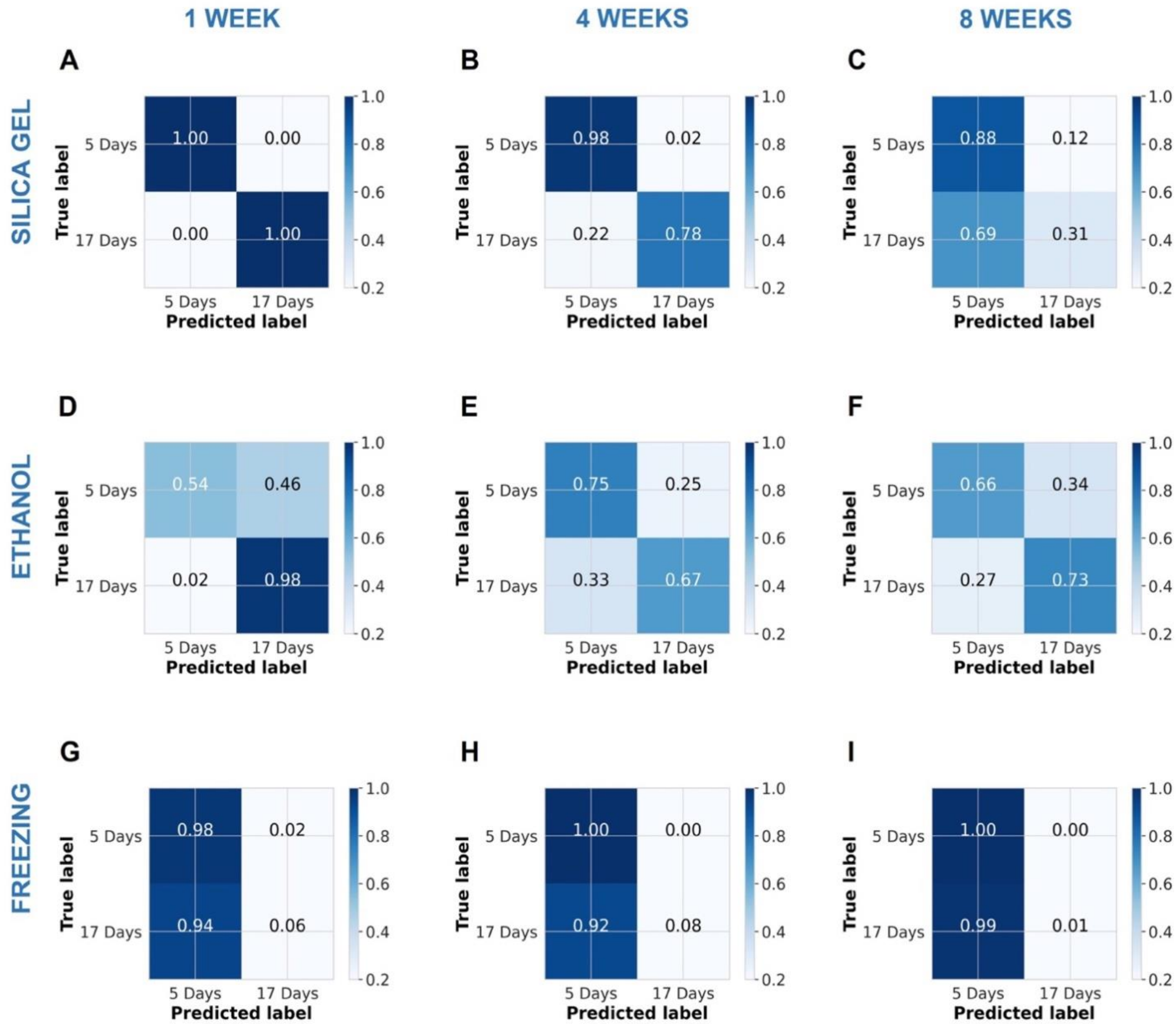

Figure 6: Confusion matrices showing prediction accuracies of mosquito ages from a standard SVM classifier trained with samples preserved in silica gel and stored for one week as a reference

4.2 Discussion

Infrared-based techniques are increasingly being used for entomological studies such as age-grading and species identification of malaria vectors vectors (Goh *et al.,* 2021; Hugo *et al.,* 2010; Lambert *et al.,* 2018; Mwanga *et al.,* 2019; Siria *et al.* ,2022). As these techniques are still operational at a small scale, researchers have mostly depended on specific sample handling approaches, with limited considerations for either standardization or alternatives. However, for NIR-specific uses, it has been demonstrated that the methods can work with mosquitoes preserved using multiple techniques. Dowell *et al.,* for example, demonstrated that the chronological ages of mosquitoes could be predicted from NIR-spectra to within 1.4 days when using either desiccants, ethanol, Carnoy, RNA *later*®, or refrigeration (Dowell *et al.,* 2011). Separately, Sikulu *et al,* showed that preserving mosquitoes in RNA *later*® reduced the likelihood of misclassifying the age of *Anopheles* mosquitoes, further emphasizing the potential of this preservative (Sikulu *et al.,* 2011). These studies demonstrated the expanded potential for infrared-based applications and provided a basis for additional investigations.

Since the infrared spectroscopy applications now increasingly also include the mid-infrared spectral range, studies on different sample handling techniques should be expanded to the MIR application as well. This study investigated the effects of different preservation methods and tested whether different sample storage durations could influence the performance of the age-grading models using laboratory-reared mosquitoes that were either young (5-days old) or old (17-days old), based on mid-infrared spectra. Desiccation over silica gel is the most commonly used preservation method by entomologists and has also been widely used in previous infrared-based studies, thus it was considered in this study as the primary reference, against which other methods were compared. Similarly, the storage durations of 1, 4 and 8 weeks were selected to represent a practical range over which samples would normally be stored before analysis even in cases where there are limited analytical resources. One week was considered as the baseline, against which other durations could be evaluated.

Of the seven classification models tested, support vector machine (SVM) was the best performing at predicting the ages of mosquitoes preserved in different preservation methods, achieving about 96% accuracy. Broadly, the MIRS-based approach could accurately classify the age groups of mosquitoes even after 8 weeks of storage, even though the performance was best for samples stored in 1 week. When data was put together for all storage durations, and the SVM model trained using just samples preserved in silica, the highest accuracy was obtained when the unseen data being predicted was also from silica-preserved mosquitoes. However, when this model was used to predict samples preserved in other methods, the accuracy declined significantly, suggesting the need to standardize the preservation method. The same observation was made when the model was trained using samples preserved by silica for one week and then used to perform the predictions for other preservations and storage durations. Here too, the highest predictions accuracies were obtained from the preservation method and storage duration that was used as a reference (Table 2).

While these data do not necessarily offer a comprehensive analysis of all possible preservation methods, they clearly demonstrate the need to either standardize the preservation methods or at least deploy an additional layer of statistical procedures such as transfer learning (Weiss *et al.*, 2016), where a small amount of different data is introduced into the training set, to neutralize the differences introduced by using different preservatives. Such statistical approaches have been applied to improve predictions on samples collected in different countries or laboratories for mosquito age grading (Siria *et al.*, 2022). In the study conducted by Siria *et al*, the transfer-learning approach was demonstrably effective at extending the utility of the deep learning models to predict the ages of field-collected mosquitoes in both Tanzania and Burkina Faso (Siria *et al.*, 2022). Perhaps the most practical option would simply be to require standardized treatments of samples in both laboratory and field studies.

In particular, silica gel has been used by researchers in preserving many samples for both species identification, age grading and blood meal experiments at a relatively low cost (Lambert *et al.*, 2018; Mwanga *et al.*, 2019; Siria *et al.*, 2022). This makes it ideal for storing large numbers of mosquito samples in field settings over long durations. This current study has also shown that silica gel desiccation may be ideal for storing mosquito samples for a short period at ~ 5°C, as this were the samples for which age classification was most accurate. On the contrary, frozen samples achieved far lower age predictions across the different storage durations, when the SVM model was trained on silica gel as a reference (Table 2). These samples were not as dry as those preserved in silica gel for the same duration; therefore, excess water content may have limited the full potential of the machine learning predictions of the spectral data even after the data was cleaned. Because of the moisture content even after the drying period before scanning, the frozen samples were easily crushed by the anvil and provided no resistance when pressed against the ATR crystal of the spectrometer. This may add additional complexity to data analysis, requiring that certain data points are discarded due to excessive water content as previously suggested by González *et al* .(2019). Lastly, ethanol has also been used widely for mosquito preservation, especially where the nucleic acid component is needed for further analyses (Torres *et al.,* 2019). In this study, the prediction accuracies of models trained with ethanol-preserved samples dropped to 50% when age-classifying silica-preserved mosquitoes and to 56% when age-classifying frozen samples. Further analysis, beyond the scope of this current study, may be needed to evaluate these comparisons, and possibly include additional statistical approaches.

This study also allowed direct assessment of whether variations in storage duration can impact the accuracy of mid-infrared-based approaches for mosquito age-grading. Here, the samples stored for one-week were initially used as the standard reference and used to train the basic machine learning models, which were then used to predict age classes of mosquitoes stored for different durations. It is particularly important to evaluate these differences since entomological surveys are typically time-consuming and can generate very large numbers of samples that cannot be analysed on the same day immediately. As a result, some form of extended storage for weeks or months is often necessary, especially where the equipment for sample analysis does not exist on-site. Overall, these results suggest that mid-infrared spectroscopy coupled with machine learning can predict the age of mosquito species stored in different preservation methods for different periods of up to 8 weeks. As previously demonstrated, the approach has the advantages of being quick to perform, cost-effective, and reagent-free, this being reliable even in low-resource settings (Jiménez *et al.,* 2019; Siria *et al.,* 2022).

One limitation of this study was that the mosquitoes used were not blood-fed, and therefore did not fully simulate the natural mosquito life cycle processes. To further reduce experimental variations, the experiments used only two mosquito ages (5 and 17 days old). These factors could be addressed in future studies by expanding the range of ages and physiological states of mosquitoes so as to be more representative of the natural world. It may also be necessary to evaluate silica-preserved mosquitoes without any refrigeration as done in this study, since it will not be operationally feasible for large scale field studies and areas with limited or no access to electricity. Moreover, the range of preservatives and durations of storage were limited to just three each, to ensure feasibility. Further investigation may reveal that such models may respond differently when using an expanded range of preservatives or storage durations.

Nonetheless, the data shows that silica-based preservation is a satisfactory starting point for samples destined for spectroscopy; and can be used for several weeks of storage.

CHAPTER FIVE

CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

This study has demonstrated that both the preservation methods and storage durations are important determinants of the classification accuracy used to predict mosquito ages using mid-infrared spectra data. Furthermore, we observed that the highest accuracies are achieved when the training samples are preserved the same way and stored for the same duration as the test samples. Additionally, among all the preservation methods used, drying over silica gel was the best method and could be used for up to several weeks performing slightly better than ethanol. Alternatively, additional machine learning techniques such as transfer learning and deep learning approaches may be incorporated to improve prediction accuracy between distinct groups. Protocols for entomological studies should therefore specify the need to standardize sample handling procedures for infrared-based approaches. The development of such a protocol will allow for the immediate integration of this technology into large-scale vector surveillance programs, providing key insights to aid in the control of malaria vectors.

5.2 Recommendations

The ability of MIRS coupled with ML technique to detect the ages of the malaria vector with an accuracy of over 80% is ecologically relevant and it can be used to effectively strategize and monitor vector control programs and for malaria surveillance. Thus, from the findings of this study we can recommend the following:

(i) Further research incorporating additional preservation methods and storage durations.

(ii) Extended analysis (such as transfer learning) or training and validating of models using samples stored for the same duration; which may enable users to analyze their samples for longer periods.

(iii) More research to investigate possible variations in mosquito species transmitting malaria, in terms of machine learning techniques.

(iv) Standardized protocols for sample handling should be developed for infrared-based approaches.

REFERENCES

Bass, C., Williamson, M. S., Wilding, C. S., Donnelly, M. J., & Field, L. M. (2007). Identification of the main malaria vectors in the *Anopheles gambiae* species complex using a TaqMan real-time PCR assay. *Malaria Journal*, *6*(1), 1–9.

Bhatt, S., Weiss, D. J., Cameron, E., Bisanzio, D., Mappin, B., Dalrymple, U., Battle, K. E., Moyes, C. L., Henry, A., & Eckhoff, P. A. (2015). The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. *Nature*, *526*(7572), 207–211.

Cibulskis, R. E., Alonso, P., Aponte, J., Aregawi, M., Barrette, A., Bergeron, L., Fergus, C. A., Knox, T., Lynch, M., & Patouillard, E. (2016). Malaria: Global progress 2000–2015 and future challenges. *Infectious Diseases of Poverty*, *5*(1), 1–8.

Coleman, M., Hemingway, J., Gleave, K. A., Wiebe, A., Gething, P. W., & Moyes, C. L. (2017). Developing global maps of insecticide resistance risk to improve vector control. *Malaria Journal*, *16*(1), 1–9.

Cook, P. E., Hugo, L. E., Iturbe-Ormaetxe, I., Williams, C. R., Chenoweth, S. F., Ritchie, S. A., Ryan, P. A., Kay, B. H., Blows, M. W., & O'Neill, S. L. (2006). The use of transcriptional profiles to predict adult mosquito age under field conditions. *Proceedings of the National Academy of Sciences*, *103*(48), 18060–18065.

Cook, P. E., McMeniman, C. J., & O'Neill, S. L. (2008). Modifying insect population age structure to control vector-borne disease. *Transgenesis and the Management of Vector-Borne Disease*, 126–140.

Dowell, F. E., Noutcha, A. E. M., & Michel, K. (2011). The effect of preservation methods on predicting mosquito age by near infrared spectroscopy. *The American Journal of Tropical Medicine and Hygiene*, *85*(6), 1093.

Farlow, R., Russell, T. L., & Burkot, T. R. (2020). Nextgen Vector Surveillance Tools: sensitive, specific, cost-effective and epidemiologically relevant. *Malaria Journal*, *19*(1), 1–13.

Ferguson, H. M., Maire, N., Takken, W., Lyimo, I. N., Briët, O., Lindsay, S. W., & Smith, T. A. (2012). Selection of mosquito life-histories: A hidden weapon against malaria? *Malaria Journal*, *11*(1), 1–5.

Fernandes, J. N., dos Santos, L. M. B., Chouin-Carneiro, T., Pavan, M. G., Garcia, G. A., David, M. R., Beier, J. C., Dowell, F. E., Maciel-de-Freitas, R., & Sikulu-Lord, M. T. (2018). Rapid, non-invasive detection of Zika virus in *Aedes aegypti* mosquitoes by near-infrared spectroscopy. *Science Advances*, *4*(5), eaat0496.

Finda, M. F., Moshi, I. R., Monroe, A., Limwagu, A. J., Nyoni, A. P., Swai, J. K., Ngowo, H. S., Minja, E. G., Toe, L. P., & Kaindoa, E. W. (2019). Linking human behaviours and malaria vector biting risk in south-eastern Tanzania. *PloS One*, *14*(6), e0217414.

Goh, B., Ching, K., Soares Magalhães, R. J., Ciocchetta, S., Edstein, M. D., Maciel-de-Freitas, R., & Sikulu-Lord, M. T. (2021). The application of spectroscopy techniques for diagnosis of malaria parasites and arboviruses and surveillance of mosquito vectors: A systematic review and critical appraisal of evidence. *PLoS Neglected Tropical Diseases*, *15*(4), e0009218.

Gorokhova, E. (2005). Effects of preservation and storage of microcrustaceans in RNAlater on RNA and DNA degradation. *Limnology and Oceanography: Methods*, *3*(2), 143–148.

Guissou, E., Waite, J. L., Jones, M., Bell, A. S., Suh, E., Yameogo, K. B., Djègbè, N., Da, D. F., Hien, D. F. D. S., & Yerbanga, R. S. (2021). A non-destructive sugar-feeding assay for parasite detection and estimating the extrinsic incubation period of *Plasmodium falciparum* in individual mosquito vectors. *Scientific Reports*, *11*(1), 1–14.

Hammond, A. M., & Galizi, R. (2017). Gene drives to fight malaria: current state and future directions. *Pathogens and Global Health*, *111*(8), 412–423.

Hemingway, J., Ranson, H., Magill, A., Kolaczinski, J., Fornadel, C., Gimnig, J., Coetzee, M., Simard, F., Roch, D. K., & Hinzoumbe, C. K. (2016). Averting a malaria disaster: Will insecticide resistance derail malaria control? *The Lancet*, *387*(10029), 1785–1788.

Heraud, P., Chatchawal, P., Wongwattanakul, M., Tippayawat, P., Doerig, C., Jearanaikoon, P., Perez-Guaita, D., & Wood, B. R. (2019). Infrared spectroscopy coupled to cloud-based data management as a tool to diagnose malaria: A pilot study in a malaria-endemic country. *Malaria Journal*, *18*(1), 1–11.

Hugo, L. E., Cook, P. E., Johnson, P. H., Rapley, L. P., Kay, B. H., Ryan, P. A., Ritchie, S. A., & O'Neill, S. L. (2010). Field validation of a transcriptional assay for the prediction of age of uncaged *Aedes aegypti* mosquitoes in northern Australia. *PLoS Neglected Tropical Diseases*, *4*(2), e608.

Hugo, L. E., Quick-Miles, S., Kay, B. H., & Ryan, P. A. (2014). Evaluations of mosquito age grading techniques based on morphological changes. *Journal of Medical Entomology*, *45*(3), 353–369.

Jiménez, M. G., Babayan, S. A., Khazaeli, P., Doyle, M., Walton, F., Reedy, E., Glew, T., Viana, M., Ranford-Cartwright, L., & Niang, A. (2019). Prediction of mosquito species and population age structure using mid-infrared spectroscopy and supervised machine learning. *Wellcome Open Research, 4*.

Johnson, B. J., Hugo, L. E., Churcher, T. S., Ong, O. T. W., & Devine, G. J. (2020). Mosquito age grading and vector-control programmes. *Trends in Parasitology*, *36*(1), 39–51.

Khoshmanesh, A., Christensen, D., Perez-Guaita, D., Iturbe-Ormaetxe, I., O'Neill, S. L., McNaughton, D., & Wood, B. R. (2017). Screening of Wolbachia endosymbiont infection in *Aedes aegypti* mosquitoes using attenuated total reflection mid-infrared spectroscopy. *Analytical Chemistry*, *89*(10), 5285–5293.

Kiszewski, A., Mellinger, A., Spielman, A., Malaney, P., Sachs, S. E., & Sachs, J. (2004). A global index representing the stability of malaria transmission. *The American Journal of Tropical Medicine and Hygiene*, *70*(5), 486–498.

Kleinschmidt, I., & Rowland, M. (2021). Insecticides and malaria. *Ecology and Control of Vector-borne Diseases*, 266. Wageningen Academic Publishers.

Koella, J. C., Lynch, P. A., Thomas, M. B., & Read, A. F. (2009). Towards evolution-proof malaria control with insecticides. *Evolutionary Applications*, *2*(4), 469–480.

Lambert, B., Sikulu-Lord, M. T., Mayagaya, V. S., Devine, G., Dowell, F., & Churcher, T. S. (2018). Monitoring the age of mosquito populations using near-infrared spectroscopy. *Scientific Reports*, *8*(1), 1–9.

Lorenz, L. M., Bradley, J., Yukich, J., Massue, D. J., Mageni Mboma, Z., Pigeon, O., Moore, J., Kilian, A., Lines, J., & Kisinza, W. (2020). Comparative functional survival and equivalent annual cost of 3 long-lasting insecticidal net (LLIN) products in Tanzania: A randomised trial with 3-year follow up. *PLoS Medicine*, *17*(9), e1003248.

Maia, M. F., Kapulu, M., Muthui, M., Wagah, M. G., Ferguson, H. M., Dowell, F. E., Baldini, F., & Ranford-Cartwright, L. (2019). Detection of *Plasmodium falciparum* infected *Anopheles gambiae* using near-infrared spectroscopy. *Malaria Journal*, *18*(1), 1–11.

Monroe, A., Mihayo, K., Okumu, F., Finda, M., Moore, S., Koenker, H., Lynch, M., Haji, K., Abbas, F., & Ali, A. (2019). Human behaviour and residual malaria transmission in Zanzibar: Findings from in-depth interviews and direct observation of community events. *Malaria Journal*, *18*(1), 1–13.

Mwanga, E. P., Mapua, S. A., Siria, D. J., Ngowo, H. S., Nangacha, F., Mgando, J., Baldini, F., González Jiménez, M., Ferguson, H. M., & Wynne, K. (2019). Using mid-infrared spectroscopy and supervised machine-learning to identify vertebrate blood meals in the malaria vector, *Anopheles arabiensis*. *Malaria Journal*, *18*(1), 1–9.

Mwanga, E. P., Minja, E. G., Mrimi, E., Jiménez, M. G., Swai, J. K., Abbasi, S., Ngowo, H. S., Siria, D. J., Mapua, S., & Stica, C. (2019). Detection of malaria parasites in dried human blood spots using mid-infrared spectroscopy and logistic regression analysis. *Malaria Journal*, *18*(1), 1–13.

Mayagaya, V. S., Michel, K., Benedict, M. Q., Killeen, G. F., Wirtz, R. A., Ferguson, H. M., & Dowell, F. E. (2009). Non-destructive determination of age and species of *Anopheles gambiae* sl using near-infrared spectroscopy. *The American journal of tropical medicine and hygiene,81*(4), 622-630.

Ntamatungiro, A. J., Mayagaya, V. S., Rieben, S., Moore, S. J., Dowell, F. E., & Maia, M. F. (2013). The influence of physiological status on age prediction of *Anopheles arabiensis* using near infra-red spectroscopy. *Parasites & Vectors*, *6*(1), 1–6.

Oakley, M. S., Verma, N., Myers, T. G., Zheng, H., Locke, E., Morin, M. J., Tripathi, A. K., Mlambo, G., & Kumar, S. (2018). Transcriptome analysis based detection of *Plasmodium falciparum* development in *Anopheles stephensi* mosquitoes. *Scientific Reports*, *8*(1), 1–12.

Ohm, J. R., Baldini, F., Barreaux, P., Lefevre, T., Lynch, P. A., Suh, E., Whitehead, S. A., & Thomas, M. B. (2018). Rethinking the extrinsic incubation period of malaria parasites. *Parasites & Vectors*, *11*(1), 1–9.

Okumu, F., & Finda, M. (2021). Key Characteristics of Residual Malaria Transmission in Two Districts in South-Eastern Tanzania - Implications for Improved Control. *International Journal of Infectious Diseases*, 223,S143–S154).

Perez-Mendoza, J., Dowell, F. E., Broce, A. B., Throne, J. E., Wirtz, R. A., Xie, F., Fabrick, J. A., & Baker, J. E. (2002). Chronological age-grading of house flies by using near-infrared spectroscopy. *Journal of Medical Entomology*, *39*(3), 499–508.

Russell, T. L., Farlow, R., Min, M., Espino, E., Mnzava, A., & Burkot, T. R. (2020). Capacity of National Malaria Control Programmes to implement vector surveillance: A global analysis. *Malaria Journal*, *19*(1), 1-9.

Sangbakembi-Ngounou, C., Costantini, C., Longo-Pendy, N. M., Ngoagouni, C., Akone-Ella, O., Rahola, N., Cornelie, S., Kengne, P., Nakouné, E. R., & Komas, N. P. (2022). Diurnal biting of malaria mosquitoes in the Central African Republic indicates residual transmission may be "out of control."*Proceedings of the National Academy of Sciences*, *119*(21), e2104282119.

Shaw, W. R., Holmdahl, I. E., Itoe, M. A., Werling, K., Marquette, M., Paton, D. G., Singh, N., Buckee, C. O., Childs, L. M., & Catteruccia, F. (2020). Multiple blood feeding in mosquitoes shortens the *Plasmodium falciparum* incubation period and increases malaria transmission potential. *PLoS Pathogens*, *16*(12), e1009131.

Sikulu, M., Dowell, K. M., Hugo, L. E., Wirtz, R. A., Michel, K., Peiris, K. H. S., Moore, S., Killeen, G. F., & Dowell, F. E. (2011). Evaluating RNA Later® as a preservative for using near-infrared spectroscopy to predict *Anopheles gambiae* age and species. *Malaria Journal*, *10*(1), 1–8.

Sikulu, M., Killeen, G. F., Hugo, L. E., Ryan, P. A., Dowell, K. M., Wirtz, R. A., Moore, S. J., & Dowell, F. E. (2010). Near-infrared spectroscopy as a complementary age grading and species identification tool for African malaria vectors. *Parasites & Vectors*, *3*(1), 1–7.

Silver, J. B. (2008). Methods of age-grading adults and estimation of adult survival rates. *Mosquito Ecology: Field Sampling Methods*, 1161–1271.

Sinkins, S. P., & O'Neill, S. L. (2000). Wolbachia as a vehicle to modify insect populations.

Siria, D. J., Sanou, R., Mitton, J., Mwanga, E. P., Niang, A., Sare, I., Johnson, P. C. D., Foster, G. M., Belem, A. M. G., & Wynne, K. (2022a). Rapid age-grading and species identification of natural mosquitoes for malaria surveillance. *Nature Communications*, *13*(1), 1–9.

Strode, C., Donegan, S., Garner, P., Enayati, A. A., & Hemingway, J. (2014). The impact of pyrethroid resistance on the efficacy of insecticide-treated bed nets against African *anopheline* mosquitoes:Systematic review and meta-analysis. *PLoS Medicine*, *11*(3), e1001619.

Tizifa, T. A., Kabaghe, A. N., McCann, R. S., van den Berg, H., van Vugt, M., & Phiri, K. S. (2018). Prevention efforts for malaria. *Current Tropical Medicine Reports, 5*(1), 41–50.

Torres, M. G., Weakley, A. M., Hibbert, J. D., Kirstein, O. D., Lanzaro, G. C., & Lee, Y. (2019). Ethanol as a potential mosquito sample storage medium for RNA preservation. *F1000Research*, *8*.

Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. *Journal of Big Data*, *3*(1), 1–40.

World Health Organization. (2015). *Global technical strategy for malaria 2016-2030*. World Health Organization.

World Health Organization. (2018). *Malaria surveillance, monitoring and evaluation:* A reference manual.

World Health Organization. (2019). *Global malaria report 2019:* World Health Organization, Geneva.

World Health Organization. (2020). *World malaria report 2020:* 20 years of global progress and challenges.

World Health Organization. (2021). *World Malaria Report 2021*.

World Health Organization, & UNICEF. (2017). *Global vector control response 2017-2030*.

RESEARCH OUPUTS

Output 1: Paper published by *Parasites & Vectors* journal

Mgaya, J. N., Siria, D. J., Makala, F. E., Mgando, J. P., Vianney, J. M., Mwanga, E. P., & Okumu, F. O. (2022). Effects of sample preservation methods and duration of storage on the performance of mid-infrared spectroscopy for predicting the age of malaria vectors.

Parasites & Vectors <vinothini.mani@springernature.com> Sat, Jul 9, 3:58 PM
to me

Ref: Submission ID 83487bd0-5bb5-47ad-b608-432599dfb1a8

Dear Dr Mgaya,

Re: "Effects of sample preservation methods and duration of storage on the performance of mid-infrared spectroscopy for predicting the age of malaria vectors"

We're delighted to let you know that your manuscript has been accepted for publication in Parasites & Vectors.

Prior to publication, our production team will check the format of your manuscript to ensure that it conforms to the journal's requirements. They will be in touch shortly to request any necessary changes, or to confirm that none are needed.

Checking the proofs

Once we've prepared your paper for publication, you will receive a proof. At this stage, please check that the author list and affiliations are correct. For the main text, only errors that have been introduced during the production process, or those that directly compromise the scientific integrity of the paper, may be corrected.

As the corresponding (or nominated) author, you are responsible for the accuracy of all content, including spelling of names and current affiliations.

To ensure prompt publication, your proofs should be returned within two working days.

Publication policies

Acceptance of your manuscript is conditional on all authors agreeing to our publication policies at: https://www.springernature.com/gp/policies/editorial-policies

Article Processing Charge

You will shortly receive an email asking you to confirm your institutional affiliation and arrange payment of your article-processing charge (APC), if applicable. To find out more about APCs, visit our support portal: https://support.springernature.com/en/support/solutions/6000138386

Once again, thank you for choosing Parasites & Vectors, and we look forward to publishing your article.

Kind regards,

Filipe Dantas-Torres
Editor
Parasites & Vectors

Output 2: Poster