Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Mid-infrared spectroscopy and supervised machine learning can determine age class of arboviral vector Culex pipiens at high age resolution

Ryan McCaffrey Carmichael
rcarmichael01@qub.ac.uk

Queen's University Belfast
Mauro Pazmiño-Betancourth
University of Glasgow
Jean-Philippe Parvy
MRC University of Glasgow Centre for Virus Research
Sai Kulkarni
MRC University of Glasgow Centre for Virus Research
Jaimie T A Dick
Queen's University Belfast
Katarzyna Modrzynska
University of Glasgow
Simon Babayan
University of Glasgow
Ross N Cuthbert
Queen's University Belfast
Heather M Ferguson
University of Glasgow
Francesco Baldini
University of Glasgow
Emilie Pondeville
MRC University of Glasgow Centre for Virus Research

Research Article

Keywords: Age determination, mosquito, mid-infrared spectroscopy, arbovirus transmission, vector control, machine learning

Posted Date: July 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10183333/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Culex pipiens is a widespread temperate mosquito species and the principal vector of several medically important arboviruses, most notably West Nile virus, in temperate regions. Recent years have seen increased transmission of *Culex*-borne arboviruses, underscoring the need for targeted vector surveillance, transmission risk prediction, and control. Since vector age is a key determinant of the efficiency with which mosquitoes can transmit a pathogen, determining the age structure of field mosquito populations is critical for accurately assessing the risk of pathogen transmission. However, existing methods for determining the age of *Cx. pipiens* are currently unsuitable, limiting our ability to assess the transmission potential of natural populations.

Methods

Here, we developed a successful model of age determination in *Cx. pipiens* using mid-infrared spectroscopy (MIRS). MIRS captures the chemical signature of the mosquito cuticle, which changes over the mosquito lifespan, thereby enabling mosquitoes to be classified into age groups with epidemiological relevance and resolved with sufficient precision to allow fine-scale assessments of vector age and pathogen transmission risk. In this study, over 1500 individual female and 500 male mosquito spectra from a recently established laboratory colony of *Cx. pipiens* were obtained, with female age determined at “low” (two age classes), “medium” (four age classes) and “high” resolution (eight age classes) and male at three age classes.

Results

Using a multi-layer perceptron classifier, a machine learning algorithm was trained on different biochemical signatures of age. MIRS could predict age class of *Cx. pipiens* with relatively good accuracy, with variable performance depending on the resolution. Using an unseen spectra dataset, the trained model successfully predicted the age class of *Cx. pipiens*, though accuracy declined at medium and high resolutions. Notably, MIRS detected newly emerged mosquitoes with high accuracy, even at a high resolution.

Conclusions

This demonstrates the utility of MIRS as a tool for age determination in *Cx. pipiens*, which could have applications in determining the risk of transmission by natural *Culex* populations as well as evaluating vector control strategies. More work is needed to validate this technique in field populations and expand to related species.

Introduction

Mosquito-borne diseases pose an increasing global public and animal health threat, and understanding the transmission potential of natural mosquito populations is essential for accurately assessing disease risk and informing mitigation strategies. The efficiency of mosquito-borne pathogen transmission by natural populations of mosquitoes – *i.e.*, vectorial capacity - which is the number of potentially infectious bites that would arise per day from all mosquitoes biting a single infectious host [1] can be estimated using a mathematical model taking into account different factors. These include the abundance of vector mosquitoes, the biting rate, vector survival, the duration of the extrinsic incubation period (EIP) after which the vector can transmit the pathogen to a new host, the vector survival after the EIP period is completed, and the vector competence, which is the mosquito's ability to become infected following an infectious blood meal and to subsequently transmit the virus [2]. Among these factors, mosquito survival has the greatest influence on transmission potential. This is because survival affects overall mosquito abundance (number of adult mosquitoes and progeny), and importantly, only females that survive long enough to complete the EIP can transmit the pathogen [3]. Consequently, determining the vector population age structure is key to understanding risk of disease transmission, as longer-lived populations are consequently more likely to have increased biting rates, and reproductive output, as well as completed the EIP and become infectious before they die. Since the duration of the EIP can vary substantially among mosquito population and pathogen combinations [4–6], being able to determine age of mosquitoes at high resolution is beneficial for obtaining the most accurate risk assessment. Mosquito age determination is also useful to evaluate efficacy of vector control strategies, such as the ones aiming at reducing population abundance. Monitoring the population age structure pre- and post-management intervention (*e.g.*, targeted adulticide as in [7]) should indeed show a reduction in adult mosquito longevity and increase in proportion of young adult mosquitoes.

Age-grading has been widely applied to understand the transmission potential and response to interventions in tropical mosquito vectors (*e.g.*, the African malaria *Anopheles* or *Aedes* vectors [8, 9]). Historically, mosquito age grading has primarily relied on ovarian dissection methods, including examination of tracheolar coiling or follicular dilatations to estimate parity and number of previous ovipositions [10, 11]. Although these approaches have been validated in different mosquito species (*e.g.*, [12]), they provide only a crude estimate of chronological age as not considering number of ovipositions, or length and frequency of gonotrophic cycles, which can result in low age resolution and underestimation of age. In addition, they are labour intensive, difficult to standardise at scale [13]. Mid-infrared spectroscopy (MIRS) has recently emerged as a viable tool to age-grade mosquitoes [14]. MIRS specifically quantifies the light absorbed by an 'object' at different wavelengths in the mid-infrared region (4000 – 400 cm^{-1}), corresponding to the fundamental vibrations of biochemical bonds, and using supervised machine learning can disentangle the obtained spectra to provide information on its chemical composition and molecular structure [15]. Since the biochemical profile of the mosquito cuticle varies over the mosquito lifespan [16], MIRS can be used to differentiate mosquito age classes based on these variations. MIRS is advantageous over other techniques because it is reagent-free, relatively easy

to carry out, and can allow for high-throughput processing. While MIRS has been applied to age determination in the major vectors of malaria in Africa, the *Anopheles gambiae* complex [15], including validation in semi-field conditions [17], and also demonstrated to be effective in age determination in other species like *Aedes albopictus* and *An. funestus* [18, 19] as well as other arthropod disease vectors [20], it has not yet been tested in *Culex* mosquitoes. This represents an important gap given the increasing relevance of temperate *Culex* populations under climate change, where their role in disease transmission may expand [21]. In Europe, *Culex pipiens* sensu lato (s.l.) is widely distributed and frequently reported as the most abundant species captured in European mosquito surveillance programmes [22–24] and is considered to be the primary vector of both Usutu (USUV) and West Nile virus (WNV) [25]. Both USUV and WNV are flaviviruses which are maintained horizontally in an enzootic cycle primarily between avian hosts and *Culex pipiens* s.l. mosquito vectors, particularly *Cx. pipiens pipiens*, while the anthropophilic form *Cx. pipiens molestus* is considered an important bridge vector capable of transmitting these viruses to dead-end human hosts [25, 26]. *Cx. pipiens pipiens* has been reported to be able to feed on humans [27, 28], suggesting that it may also contribute to bridge transmission.[29–33] *Culex*-transmitted arboviruses are also transmitted by vertical transmission from an infected female to its offspring [29–31]. Diapausing *Cx. pipiens* females can also act as overwintering reservoirs and contribute to seasonal viral re-emergence (*e.g.*, USUV; [32]). In addition, *Culex* mosquitoes are often long-lived under cool environmental conditions [33], particularly during overwintering diapause, with some individuals surviving for several weeks to months. Notably, *Cx. pipiens* from Scotland (UK) have recently been demonstrated to live up to six months under laboratory conditions at 18°C, the longest ever recorded survival for female mosquitoes in the absence of diapause [34]. Consequently, it remains unclear whether MIRS can provide sufficiently fine-scale age resolution across these extended lifespans to generate epidemiologically meaningful estimates of transmission risk in *Culex* populations. Additionally, no method has been validated for age-grading male mosquitoes in *Culex* despite the importance of male longevity to vector control interventions like the sterile insect technique (*e.g.*, [35]), though testis morphology can be used to distinguish between age classes of male *Anopheles gambiae* [36]. Thus, the development of novel methodologies that are more reproducible, reliable and scalable is important for successful age determination in *Culex* mosquitoes.

In this study, we developed and evaluated a model for age determination in both female and male *Cx. pipiens pipiens* mosquitoes using MIRS spectra and supervised machine learning. Using spectral data obtained from the Scottish *Cx. pipiens* colony [34], mosquitoes of different ages and physiological states were analysed, including unfed, blood-fed, and gravid females. Female mosquitoes up to 79 days old, the longest lifespan assessed to date using MIRS, were included in the analyses. The resulting models successfully classified female mosquito age across low-, medium-, and high-resolution age categories, with high accuracy for the extreme age classes, and reduced for intermediate age classes, especially at high resolution. In addition, a separate model was developed to distinguish newly emerged male mosquitoes from older individuals. Collectively, these findings demonstrate the potential of MIRS as a robust tool for age determination in *Culex* mosquitoes. Incorporation of such models into existing One Health surveillance programmes could provide valuable insights into the age structure and longevity of

monitored *Culex* populations, thereby improving understanding of vector population dynamics and the risk of pathogen transmission.

Methods

Mosquito rearing

A colony of *Culex pipiens pipiens* mosquitoes originating from Bearsden, Scotland, was previously established in August 2023 and reared as described [34]. Sampling took place from November 2024 to June 2025. Approximately 200 larvae were reared per pan containing 2L of dechlorinated tap water, fed with one pellet of commercial cat food (Friskies) changed every 2-3 days to prevent bacterial contamination. Emerging mosquitoes were collected using a handheld aspirator, placed in a Bugdorm cage (Watkins & Doncaster, ref. E6099) and provided with 10% sucrose. Mosquitoes were maintained at 24°C, 80% relative humidity and an 18:6 light:dark photoperiod, including 1h dusk and 1h dawn. Mosquitoes were fed with chicken blood-meals (Envigo, UK) through a hog intestine membrane (Tongmaster, Amazon) using a Hemotek (Hemotek Ltd, Blackburn, UK). Prior to blood-feeding, mosquitoes were starved for at least 24 hours.

Sampling design

This study aimed to classify age classes of *Cx. pipiens* using MIRS spectra. Compared to other mosquito vectors investigated previously with MIRS, *Cx. pipiens* are long-lived, with females living up to 93 days under lab conditions at constant 24°C [34]. Over three independent replicates, female mosquitoes were sampled through their full lifespan, starting from 1-3 days old, then sampled at regular 11-day intervals thereafter up until and including 77-79 days old, resulting in eight age classes (AC1 to AC8, Table 1). Each replicate was an independent cohort sampled at constant intervals, *i.e.*, replicate 1 was sampled at 1 day, 11 days then every 11 days thereafter, replicate 2 was sampled at 2 days, 12 days then every 11 days thereafter, so on and so forth. To evaluate the accuracy of our models and their resolution (*e.g.*, how precise age determination is), data were grouped by age classes together to allow low (lifespan divided in two age classes), medium (four age classes) and high resolution (eight age classes) estimation of age class (Table 2). Male *Cx. pipiens* mosquitoes have a significantly shorter lifespan relative to females [34], so were sampled at 1-3 days old, 11-13 days old and 22-24 days old.

Table 1: Summary of sampling regime for female and male *Cx. pipiens* mosquitoes in this study. Approximately 50-70 mosquitoes were sampled for each age class (indicated in days post emergence) per replicate. AC = age class, R = replicate, n = sample size of each age class (three replicates). AC1 = 1-3d, AC2 = 11-13d, AC3 = 22-24d, AC4 = 33-35d, AC5 = 44-46d, AC6 = 55-57d, AC7 = 66-68d, AC8 = 77-79d.

Cx. pipiens females								
	AC1	**AC2**	**AC3**	**AC4**	**AC5**	**AC6**	**AC7**	**AC8**
R1	1	11	22	33	44	55	66	77
R2	2	12	23	34	45	56	67	78
R3	3	13	24	35	46	57	68	79
N	170	207	203	202	204	188	178	170
Cx. pipiens **males**								
	AC1	**AC2**	**AC3**	**AC4**	**AC5**	**AC6**	**AC7**	**AC8**
R1	1	11	22	-	-	-	-	-
R2	2	12	23	-	-	-	-	-
R3	3	13	24	-	-	-	-	-
N	208	208	152	-	-	-	-	-

Table 2: Summary of low, medium and high resolution organisation of *Cx. pipiens* female mosquito age classes.

Resolution	Number of age classes	Age classes (chronological days)
Low	2	1-35, 44-79
Medium	4	1-13, 22-35, 44-57, 66-79
High	8	1-3, 11-13, 22-24, 33-35, 44-46, 55-57, 66-68, 77-79

To generate variability in physiological state and improve applicability of the eventual model, mosquitoes were offered a blood-meal as previously described and given opportunity to lay eggs once a fortnight, beginning from day 10. The blood-feeding rate was never 100%, meaning that cages contained an assortment of mosquitoes in various physiological states (unfed, blood-fed, gravid). No separation of adult females was made following blood-feeding, though cages were always visually checked prior to sampling to ensure a variety of physiological states were present following the first blood-feeding. Over three replicates, approximately 170-210 mosquitoes were sampled for each sex at each timepoint. Mosquitoes were killed using chloroform and stored at 4°C in falcon tubes containing silica beads (Preservation Equipment, ref. P404-3300) and cotton wool to allow desiccation.

To collect an independent unseen data set for model validation, an additional 202 *Cx. pipiens* females from an independent rearing batch were sampled at seven of the original eight age classes as before, excluding age class 8. Mosquitoes were reared under identical conditions as before. Sampling began

approximately 1 month after the end of sampling the ‘seen’ mosquitoes and was undertaken by two independent investigators.

Mid-infrared spectroscopy

A Bruker Alpha II FT-IR Spectrometer equipped with a diamond ATR crystal (Bruker Platinum ATR Unit A225) using Bruker OPUS software was used to measure each individual mosquito. The head and thorax of each mosquito were placed on the crystal, and an anvil was lowered to hold the specimen in place. Each measurement consisted of 24 averaged scans from 4000-400 cm^{-1} at a resolution of 4 cm^{-1}. The diamond was cleaned using 70% ethanol between measurements, and background scans were taken regularly to reduce noise (*e.g.*, buildup of atmospheric CO_2).

Supervised machine learning

All analyses were conducted in Python v3.10 using scikit-learn [37]. Four datasets were analysed: three datasets (low, medium and high resolution) of female *Cx. pipiens* mosquitoes, ranging from 1-79 days old, and a dataset (one resolution) of male *Cx. pipiens* mosquitoes, ranging from 1-24 days old. A custom script [38] was run on all data to discard low-quality spectra from analyses, *i.e.*, low spectral signal, evidence of atmospheric intrusion of CO_2 or water, or other distortion from the anvil [15].

For each dataset, spectra were split into training (80%) and test (20%) sets. Spectra were mean centred. Initial model selection was conducted by comparing accuracy of six machine learning (ML) algorithms, Logistic Regression (LR), Support Vector Classifier (SVC), Random Forest (RF), k-Nearest Neighbours (KNN), Gradient Boosting Classifier (GB) and Multi-Layer Perceptron Classifier (MLP). Following this, the model with the highest accuracy was chosen for further optimisation. Hyperparameters were optimised using GridSearch. Cross-validation test scores were checked to ensure the number 1-ranked model was not overfitting. Performance of the final optimised model was evaluated using the test set and the blind test set. Model performance was evaluated using recall, precision, F1-score, balanced accuracy, and confusion matrices. To determine which biochemical features played the most significant role in age determination, permutation importance was calculated for each model. Here, the mean decrease in model accuracy following the removal of each individual wavenumber was calculated. The greater the decrease in model accuracy, the higher the importance of this wavenumber to age determination of *Cx. pipiens*.

To provide additional validation of the technique on an independent biological replicate to ensure generalisability of the final optimised model, and that it is not overfitted to the training data, spectra from the unseen dataset were tested on the final optimised models at each resolution and accuracy determined using confusion matrices.

Results

Mid-infrared spectra of *Culex pipiens* mosquitoes varies across vector lifespan

In this study, 1588 and 593 MIRS spectra were obtained for female and male *Cx. pipiens* mosquitoes, respectively, to develop a model of age determination. Low-quality spectra were discarded from analyses (approximately 5% per sex), leaving 1522 and 568 MIRS spectra to be used to develop ML algorithms for age determination of *Cx. pipiens* females and males. Baseline performance of each ML algorithm was determined and for all models, with the MLP model found to have the highest baseline performance (Supplementary Figure 1). Accuracy decreased with increasing resolution in *Cx. pipiens* females; 89% at low resolution, 79% at medium resolution and 66% at high resolution. Baseline accuracy for *Cx. pipiens* males (similar number of age classes to medium resolution model) was 77%.

Hyperparameters of the highest performing MLP models were next optimised using GridSearch (Supplementary Table 1). Cross-validation scores were checked to ensure consistency between data folds (Supplementary Figure 2). Learning curves were checked to ensure models were learning with increased number of training samples (Supplementary Figure 3). Balanced accuracy of the female MLP models decreased with increasing number of age classes; 92.8% at low resolution (two age classes), 88.9% at medium resolution (four age classes) and 79.0% at high resolution (eight age classes). When misclassifications were made by the MLP model, these tended to be for adjacent age classes close (*i.e.*, one higher or lower age category) to the true class (Figure 1). There did not appear to be any systematic bias in either direction. Balanced accuracy of the male MLP model was 82.5%. Precision, recall and f1 scores varied between age classes of final optimised MLP models (Supplementary Table 2).

Permutation importance

The top-50 wavenumbers with the most importance in *Cx. pipiens* age determination were identified by permutation importance and visualised for each model (Figure 2). At high and medium resolutions, model learning appeared to be from a range of biochemical regions, including the C – H stretch and C – O stretch. At low resolution, importance appears to be concentrated between 2162-2100 cm^{-1}, with some influence from the amide I region and C – O stretch. For male *Cx. pipiens*, learning appears to be most influenced by the amide III region. Generally, model learning appears to be mostly from the fingerprint region (1300-600 cm^{-1}) across both sexes and age class resolution in *Cx. pipiens*, with less importance appearing to be derived from the functional group region (4000-1300 cm^{-1}).

Validation of age determination models on unseen dataset

To provide additional validation and assess the generalisability of the final optimised model, thereby ensuring that it was not overfitted to the training data, the model was subsequently evaluated using the independent unseen dataset. From the unseen dataset, 194 female MIRS spectra were obtained after discarding of low-quality spectra. There was an expected but not pronounced decrease in accuracy transitioning from the seen to the unseen dataset (Figure 3), with an overall accuracy of 80.9% at low resolution, 46.9% at medium resolution and 34.0% at high resolution, still much higher than random predictions (*i.e.*, 50%, 25% and 12.5%, respectively) suggesting that age-specific spectral features were learned by the model.

Discussion

Characterising the age structure of mosquito populations is key for understanding their transmission potential and accurately assessing disease risk. Here, we demonstrate the successful application of MIRS to age-grade *Culex pipiens*, the main vector of West Nile virus (WNV) and Usutu virus (USUV) in Europe and report, to our knowledge, its first application within the genus *Culex*. Using spectral data collected from a recently established colony of *Cx. pipiens* [34], we have developed three MLP models that can classify adult female mosquitoes with relatively good accuracy into age classes, with variable performance depending on the resolution (low, medium and high age resolution). Further, an additional MLP model was developed that could clearly discriminate three distinct age classes of male *Cx. pipiens* mosquitoes.

MIRS has been previously demonstrated to be effective at discriminating between age classes in several *Anopheles* species [15, 17, 18], *Aedes albopictus* [19] and tsetse flies [20]. Additionally, MIRS has recently been applied to male mosquitoes of *An. gambiae*, *An. coluzzi* [39] and *Ae. albopictus* [19]. Therefore, this work represents a key transition of the application of MIRS-ML from tropical to temperate vector settings. Generally, the final seen models classified female mosquitoes with relatively good accuracy into age classes, being most effective at classifying the extremes of ages for females (≥ 90%). At medium and high resolutions, accuracy was reduced for intermediate age classes, especially at high resolution. Misclassifications were primarily concentrated in neighbouring age classes, suggesting that age-related cuticular changes over short ageing intervals may be too subtle to be reliably detected, resulting in overlap between adjacent age groups.

MIRS has previously been demonstrated to be effective at age grading mosquito vectors at a range of resolutions. The balanced accuracy results obtained here from the seen dataset are similar when compared to previous work which tested efficacy of MIRS at low, medium and high resolution age classes of *Ae. albopictus* [19]. [19] had a greater number of age classes (eleven age classes) at its highest resolution model compared to our study (eight age classes), though *Cx. pipiens* was sampled over a longer lifespan (up to 79 days in *Cx. pipiens* compared to 33 days in *Ae. albopictus*), meaning our study collected samples between longer intervals (11-day intervals in *Cx. pipiens* compared to 3-day intervals in *Ae. albopictus*), which may limit the comparisons to be drawn between these results. The balanced accuracy of the medium resolution model (four age classes) is similar to previous findings on *Anopheles* mosquitoes [17] and tsetse flies [20], both of which were examined using three age classes.

We further validated the three models using unseen data. The overall model accuracy declined when transitioning from the seen to the unseen datasets, and with age resolution, especially for intermediate age classes. The model accuracy at low resolution slightly decreased from 92.8% to 80.9%, and from 88.9% to 46.9% at medium resolution. While the model at high resolution decreased from 79.0% to 34.0%, it could still predict newly emerged (1–3 days old) mosquitoes with 83% accuracy. To gain confidence in the result obtained from the high-resolution model, end-users could test the same spectra

with all models to check for consistent results between model resolutions (*e.g.*, obtaining 1-35d at low resolution, 1-13d at medium resolution and 1-3d at high resolution).

Here, Multi-Layer Perceptron Classifier (MLP) was found to have the highest baseline accuracy at classifying age class of both female and male *Cx. pipiens* mosquitoes. MLP has previously been demonstrated to be an efficient model of age classification using MIRS data for *An. funestus* [18]. By changing the standard scaler option to false during preprocessing, this reduces the noise-to-signal ratio and means that no part of the spectra needed to be removed for analysis (*e.g.*, as in [19]), allowing for the full dataset to be utilised by the MLP model.

Vector competence and EIP are highly heterogeneous and vary dramatically depending on environmental factors such as temperature, but also the specific mosquito species/populations and virus combination, due for instance to genotype by genotype interactions [4–6]. The age classes at which mosquitoes were sampled have direct relevance to arbovirus EIP. Though vector competence studies in *Cx. pipiens* demonstrate high variation in EIP between populations and temperatures tested settings [40–45], recent modelling suggests that *Cx. pipiens* have an EIP of around 12 and 21 days for USUV[46] and WNV[47] respectively at 24°C. Taking these numbers into account, a high resolution model (11-day intervals) may be more necessary to distinguish between transmitting and non-transmitting mosquitoes for USUV, though a medium resolution model (22-day intervals) may be suitable to distinguish between transmitting and non-transmitting mosquitoes for WNV. While the low resolution model can accurately classify a 'young' (1-35d) or 'old' (44-79d) mosquito, the applications of this may be strictly limited to coarser estimates of population age structure. Understanding the age structure of wild *Cx. pipiens* populations is also a useful epidemiological information and can provide insights on the longevity of mosquitoes in natural settings, our understanding of which is currently limited to laboratory studies (*e.g.*, [34]) or mark-release-recapture experiments (*e.g.*, [48]).

The MIRS model developed also could detect different age classes of *Cx. pipiens* males and had a similar balanced accuracy result to the medium and high resolution *Cx. pipiens* female models. The lower accuracy of the male model (three age classes) compared to the medium resolution female model (four age classes) may be explained by the lower sample size of data used to train the male model or the shorter time intervals between age classes. MIRS has previously been utilised to determine age class of laboratory-reared male mosquitoes at three age classes in both *Ae. albopictus* [19] and *An. gambiae* [39], with our successful results here in *Culex* closely mirroring these previous findings. While male mosquitoes have less direct epidemiological relevance than females, they can be targeted in vector control strategies, for example through the release of sterile males to suppress populations [35]. Thus, determining population age structure of male mosquitoes can be highly informative to vector control efforts, *e.g.*, validating that larval source management interventions have been effective, or ensuring sterile males have lived sufficiently long enough to mate with wild females.

While this study presents critical proof-of-principle for the application of MIRS age-grading technique for *Culex* mosquitoes, it has some limitations. First, all data and resultant models were derived entirely from

a mosquito laboratory colony (*e.g.*, reduced genetic diversity, constant humidity and temperature conditions), potentially limiting the transferability of the technique from lab to field. Indeed, this effect was noted when transitioning from the seen to the unseen dataset, exemplifying the need to assemble larger MIRS datasets for model training. Whilst mosquitoes were sampled at a variety of physiological stages (unfed, blood-fed, gravid) to ensure population variability at each age class, incorporating genetically diverse populations of *Cx. pipiens* could improve the applicability of the model to other regions of Europe, or beyond. *Culex* lifespan itself may additionally vary between different populations and temperatures [33]. The models developed could be improved for field applicability by incorporating more data for ML, in addition to systematically including mosquitoes with more diverse blood-feeding histories, *e.g.*, an older mosquito with only one previous blood-feed or an intermediate-aged mosquito with several previous blood-feeds. MIRS has previously been validated for age determination in *Anopheles* in semi-field settings by taking scans of field-caught larvae reared to adult stage in laboratory settings and laboratory colony mosquitoes reared in semi-field conditions [17]. Similar efforts would need to be made to validate the application of MIRS to field populations of *Culex.* Vector normalisation during data preprocessing could additionally account for variability between operators taking MIRS scans. Importantly, this work demonstrates the validation of the MIRS-ML approach on unseen data sampled independently of the training data, representing a key step before applying the approach to semi-field and field populations of *Cx. pipiens* vectors.

As this is the first application of MIRS to study biological characteristics of *Culex* mosquitoes, there is a great deal of potential for continued investigation. In *Anopheles* mosquitoes, MIRS has been demonstrated to be effective for cryptic species identification [15, 17] and in pathogen detection [49]. MIRS is a destructive technique, meaning that in practice, the head & thorax would be required to take a scan, leaving only the abdomen left to be used for pathogen detection. For example, it could be of interest to use MIRS to identify overwintered females, which would be of an older age than was measured here (up to 79 days old), have vastly different blood-feeding histories (*e.g.*, [28]), and could also be overwintering arboviruses (*e.g.*, USUV [50]). A model could alternatively be trained by experimentally inducing diapause in *Cx. pipiens* mosquitoes (*e.g.*, [32]) to capture diapause related changes in mosquito cuticle.

Conclusions

In conclusion, the work presented here has demonstrated that there is variation in MIRS spectra obtained from *Cx. pipiens* mosquito cuticles at distinct age classes, allowing the age grading of lab-reared mosquitoes at different age classes, each with its own relevance to the epidemiology of *Culex*-borne diseases currently circulating in Europe. By estimating the age of wild *Culex* vectors, inferences can be made on the longevity of mosquitoes in field settings and predict vectorial capacity of the population. Models were developed for both female and male mosquitoes, meaning that these findings could compliment well existing entomological surveillance programmes. Models were found to be learning from regions of biochemical relevance as evidenced by permutation importance. The results presented here have been validated by an independent biological replicate, however this clearly has demonstrated

the shortcomings of the technique and the difficulty in scaling to high resolution estimates of age. Future work should enhance models by incorporating more samples, including field-collected ones for training. An improved understanding of *Cx. pipiens* age structure could not only allow for enhanced evaluation of vector control but also provide key epidemiological insights on mechanisms of arboviral transmission.

Abbreviations

MIRS – Mid-infrared spectroscopy

EIP – Extrinsic incubation period

USUV – Usutu virus

WNV – West Nile virus

ML – Machine learning

LR – Logistic Regression

SVC – Support Vector Classifier

RF – Random Forest

KNN – k-Nearest Neighbours

GB – Gradient Boosting Classifier

MLP – Multi-Layer Perceptron Classifier

Declarations

Acknowledgments

We thank David Kerrigan for administrative support, Margaret McFadyen for support in the insectaries and Andrea Pallikara for excellent technical assistance in the laboratory. We thank Iván Casas Gómez-Uribarri and Mia Quinn for providing training, guidance and support on the use of the Spectrometer. We thank all members of the Mosquito Scotland team for their support.

Conflict of interest

The authors declare no conflict of interest. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data availability

Raw data are provided with this paper and are available in the University of Glasgow repository http://dx.doi.org/10.5525/gla.researchdata.2318. Scripts created in the preparation of this manuscript can be accessed via GitHub.

Funding

RMC was supported by a PhD studentship from the NorthWestBio Doctoral Training Partnership funded by the UK Biotechnology and Biological Sciences Research Council. This study was also supported in part by the UKRI Defra grant BB/X018113/1 to HMF and EP and by the UK Medical Research Council (MC_UU_0034/8 to EP).

Ethics declaration

Not applicable.

Consent to Participate declarations

Not applicable.

Consent to Publish declaration

Not applicable.

Author contributions

R.M.C.: Conceptualisation, Data curation, Formal analysis, Investigation, Methodology, Visualisation, Writing – original draft, Writing – review & editing

M.P-B.: Formal analysis, Investigation, Methodology, Validation, Writing – original draft, Writing – review & editing

J-P.P.: Conceptualisation, Investigation, Resources, Writing – review & editing

S.K.: Investigation, Writing – review & editing

J.T.A.D.: Supervision, Writing – review & editing

K.M.: Supervision, Writing – review & editing

S.B.: Supervision, Formal analysis, Writing – review & editing

R.N.C.: Supervision, Writing – review & editing

H.M.F.: Supervision, Funding acquisition, Writing – review & editing

F.B.: Conceptualisation, Supervision, Methodology, Writing – original draft, Writing – review & editing

E.P.: Conceptualisation, Supervision, Funding acquisition, Project administration, Writing – original draft, Writing – review & editing

References

1. Brady OJ, Godfray HCJ, Tatem AJ, Gething PW, Cohen JM, Ellis McKenzie F, et al. Vectorial capacity and vector control: Reconsidering sensitivity to parameters for malaria elimination. Trans R Soc Trop Med Hyg. Oxford University Press; 2016;110:107–17. https://doi.org/10.1093/trstmh/trv113
2. Lewis J, Gallichotte EN, Randall J, Glass A, Foy BD, Ebel GD, et al. Intrinsic factors driving mosquito vector competence and viral evolution: a review. Front. Cell. Infect. Microbiol. Frontiers Media SA; 2023. https://doi.org/10.3389/fcimb.2023.1330600
3. Johnson BJ, Hugo LE, Churcher TS, Ong OTW, Devine GJ. Mosquito Age Grading and Vector-Control Programmes. Trends Parasitol. Elsevier Ltd; 2020. p. 39–51. https://doi.org/10.1016/j.pt.2019.10.011
4. Lewis J, Gallichotte EN, Randall J, Glass A, Foy BD, Ebel GD, et al. Intrinsic factors driving mosquito vector competence and viral evolution: a review. Front Cell Infect Microbiol. 2023;13. https://doi.org/10.3389/fcimb.2023.1330600
5. Tabachnick W. Nature, Nurture and Evolution of Intra-Species Variation in Mosquito Arbovirus Transmission Competence. Int J Environ Res Public Health. 2013;10:249–77. https://doi.org/10.3390/ijerph10010249
6. Lambrechts L. Quantitative genetics of Aedes aegypti vector competence for dengue viruses: towards a new paradigm? Trends Parasitol. 2011;27:111–4. https://doi.org/10.1016/j.pt.2010.12.001
7. Lopez K, Irwin P, Bron GM, Paskewitz S, Bartholomay L. Ultra-low volume (ULV) adulticide treatment impacts age structure of Culex species (Diptera: Culicidae) in a West Nile virus hotspot. J Med Entomol. Oxford University Press; 2023;60:1108–16. https://doi.org/10.1093/jme/tjad088
8. Emidi B, Kisinza WN, Mosha FW. Impact of non-pyrethroid insecticide treated durable wall lining on age structure of malaria vectors in Muheza, Tanzania. BMC Res Notes. BioMed Central Ltd.; 2017;10. https://doi.org/10.1186/s13104-017-3078-7
9. Ratovonjato J, Randrianarivelojosia M, Rakotondrainibe ME, Raharimanga V, Andrianaivolambo L, Le Goff G, et al. Entomological and parasitological impacts of indoor residual spraying with DDT, alphacypermethrin and deltamethrin in the western foothill area of Madagascar. Malar J [Internet]. 2014;21. http://www.malariajournal.com/content/13/1/21
10. Tyndale-Biscoe M. Age-grading methods in adult insects: A review. Bull. Entomol. Res. 1984. p. 341–77. https://doi.org/10.1017/S0007485300015637
11. Rosay B. Anatomical Indicators for Assessing Age of Mosquitoes: Changes in Ovarian Follicles. Ann Entomol Soc Am [Internet]. 1969;62:605–11. https://academic.oup.com/aesa/article/62/3/605/24083

12. Blom R, Spitzen J, de Haan T, Koenraadt CJM. Phenotypical aspects of Culex pipiens biotype pipiens during diapause: Lipid utilization, body size, insemination, and parity. J Insect Physiol. Elsevier Ltd; 2024;159. https://doi.org/10.1016/j.jinsphys.2024.104714
13. Anagonou R, Agossa F, Azondékon R, Agbogan M, Oké-Agbo F, Gnanguenon V, et al. Application of Polovodova's method for the determination of physiological age and relationship between the level of parity and infectivity of Plasmodium falciparum in Anopheles gambiae s.s, south-eastern Benin. Parasit Vectors. BioMed Central Ltd.; 2015;8. https://doi.org/10.1186/s13071-015-0731-7
14. Goh B, Ching K, Soares Magalhães RJ, Ciocchetta S, Edstein MD, Maciel-De-freitas R, et al. 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 Negl. Trop. Dis. Public Library of Science; 2021. https://doi.org/10.1371/journal.pntd.0009218
15. González Jiménez M, Babayan SA, Khazaeli P, Doyle M, Walton F, Reedy E, et al. Prediction of mosquito species and population age structure using mid-infrared spectroscopy and supervised machine learning. Wellcome Open Res. F1000 Research Ltd; 2019;4:76. https://doi.org/10.12688/wellcomeopenres.15201.1
16. Suarez E, Nguyen HP, Ortiz IP, Lee KJ, Kim SB, Krzywinski J, et al. Matrix-assisted laser desorption/ionization-mass spectrometry of cuticular lipid profiles can differentiate sex, age, and mating status of Anopheles gambiae mosquitoes. Anal Chim Acta. 2011;706:157–63. https://doi.org/10.1016/j.aca.2011.08.033
17. Siria DJ, Sanou R, Mitton J, Mwanga EP, Niang A, Sare I, et al. Rapid age-grading and species identification of natural mosquitoes for malaria surveillance. Nat Commun. Nature Research; 2022;13. https://doi.org/10.1038/s41467-022-28980-8
18. Mwanga EP, Siria DJ, Mshani IH, Mwinyi SH, Abbasi S, Jimenez MG, et al. Rapid classification of epidemiologically relevant age categories of the malaria vector, Anopheles funestus. Parasit Vectors. BioMed Central Ltd; 2024;17. https://doi.org/10.1186/s13071-024-06209-5
19. Foti M, Micocci M, Pazmiño-Betancourth M, Casas Gomez-Uribarri I, Serini P, Caputo B, et al. Towards scalable age-grading of Aedes albopictus mosquito using mid-infrared spectroscopy and machine learning. Sci Rep. Nature Research; 2025;15. https://doi.org/10.1038/s41598-025-24404-x
20. Pazmiño-Betancourth M, Casas Gómez-Uribarri I, Mondragon-Shem K, Babayan SA, Baldini F, Rafuse Haines L. Advancing age grading techniques for Glossina morsitans morsitans , vectors of african trypanosomiasis, through mid-infrared spectroscopy and machine learning . Biol Methods Protoc. Oxford University Press (OUP); 2024; https://doi.org/10.1093/biomethods/bpae058
21. Madhav M, Blasdell KR, Trewin B, Paradkar PN, López-Denman AJ. Culex-Transmitted Diseases: Mechanisms, Impact, and Future Control Strategies using Wolbachia. Viruses. Multidisciplinary Digital Publishing Institute (MDPI); 2024. https://doi.org/10.3390/v16071134
22. Gobbo F, Chiarello G, Sgubin S, Toniolo F, Gradoni F, Danca LI, et al. Integrated One Health Surveillance of West Nile Virus and Usutu Virus in the Veneto Region, Northeastern Italy, from 2022

to 2023. Pathogens. Multidisciplinary Digital Publishing Institute (MDPI); 2025;14. https://doi.org/10.3390/pathogens14030227

23. Stiles PC, Balatsos G, Liyanage P, Baron JN, Sakellariou Sofianou M, Bisia M, et al. Three-year assessment of integrated vector management in Attica, Greece: results from surveillance and control activities. Commun Biol. Nature Research; 2025;8. https://doi.org/10.1038/s42003-025-08825-y
24. Münger E, Atama NC, van Irsel J, Blom R, Krol L, van Mastrigt T, et al. One Health approach uncovers emergence and dynamics of Usutu and West Nile viruses in the Netherlands. Nature Communications . Nature Research; 2025;16. https://doi.org/10.1038/s41467-025-63122-w
25. Brugman VA, Hernández-Triana LM, Medlock JM, Fooks AR, Carpenter S, Johnson N. The role of Culex pipiens L. (Diptera: Culicidae) in virus transmission in Europe. Int. J. Environ. Res. Public Health. MDPI; 2018. https://doi.org/10.3390/ijerph15020389
26. Becker N, Jöst A, Weitzel T. The culex pipiens complex in Europe. J Am Mosq Control Assoc. 2012;28:53–67. https://doi.org/10.2987/8756-971X-28.4s.53
27. Wehmeyer ML, Jaworski L, Jöst H, Șulești T, Rauhöft L, Afonso SMM, et al. Host attraction and host feeding patterns indicate generalist feeding of Culex pipiens s.s. and Cx. torrentium. Parasites and Vectors . BioMed Central Ltd; 2024;17. https://doi.org/10.1186/s13071-024-06439-7
28. Blom R, Krol L, Langezaal M, Schrama M, Trimbos KB, Wassenaar D, et al. Blood-feeding patterns of Culex pipiens biotype pipiens and pipiens/molestus hybrids in relation to avian community composition in urban habitats. Parasit Vectors. BioMed Central Ltd; 2024;17. https://doi.org/10.1186/s13071-024-06186-9
29. Dohm DJ, Sardelis MR, Turell MJ. Experimental Vertical Transmission of West Nile Virus by Culex pipiens (Diptera: Culicidae) [Internet]. J. Med. Entomol. 2002. https://academic.oup.com/jme/article/39/4/640/903080
30. Schilling M, Lawson B, Spiro S, Jagdev M, Vaux AGC, Bruce RC, et al. Vertical transmission in field-caught mosquitoes identifies a mechanism for the establishment of Usutu virus in a temperate country. Sci Rep. Nature Research; 2025;15. https://doi.org/10.1038/s41598-025-09335-x
31. Dahl E, Öborn L, Sjöberg V, Lundkvist Å, Hesson JC. Vertical Transmission of Sindbis Virus in Culex Mosquitoes. Viruses. MDPI; 2022;14. https://doi.org/10.3390/v14091915
32. Linthout C, Münger E, Blom R, Pijlman GP, Oude Munnink BB, Koenraadt CJM. Overwintering and vertical transmission of Usutu virus in *Culex pipiens* , *biotype pipiens* mosquitoes contribute to annual re-emergence. EMI: Animal & Environment [Internet]. 2026;2. https://doi.org/10.1080/29986990.2026.2633032
33. Moser SK, Barnard M, Frantz RM, Spencer JA, Rodarte KA, Crooker IK, et al. Scoping review of Culex mosquito life history trait heterogeneity in response to temperature. Parasit. Vectors. BioMed Central Ltd; 2023. https://doi.org/10.1186/s13071-023-05792-3
34. Parvy J-P, Brass DP, Preston B, Carmichael RM, Kerrigan D, Kirby G, et al. Colonisation and Temperature-Driven Variation in Life-History Traits of the West Nile Virus Vector Culex pipiens at Its

Northern European Limit. Parasit Vectors. 2026;in revision.

35. Suryaningtyas NH, Nurcahyo RW, Ernawan B, Sasmita HI, Sunaryo S, Ramadhani T, et al. Dispersal and survival of gamma-irradiated Culex quinquefasciatus: Implications for sterile insect technique applications. Open Vet J. Faculty of Veterinary Medicine, University of Tripoli; 2025;15:3054–62. https://doi.org/10.5455/OVJ.2025.v15.i7.16
36. Huho BJ, Ng'habi KR, Killeen GF, Nkwengulila G, Knols BG, Ferguson HM. A reliable morphological method to assess the age of male Anopheles gambiae. Malar J. 2006;5:62. https://doi.org/10.1186/1475-2875-5-62
37. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research [Internet]. 2011;12:2825–30. http://jmlr.org/papers/v12/pedregosa11a.html
38. González-Jiménez M. Bad Blood: FTIR Spectral Analysis Tool. University of Glasgow; 2025.
39. Sanou R, Mwanga EP, Sow BBD, Maïga H, Wiminga F, Siria DJ, et al. Age-grading and species identification of male mosquito Anopheles gambiae s.l. using mid-infrared spectroscopy and machine learning. Sci Rep [Internet]. 2026; https://doi.org/10.1038/s41598-026-46306-2
40. Pilgrim J, Metelmann S, Widlake E, Seechurn N, Vaux A, Mansfield KL, et al. UK mosquitoes are competent to transmit Usutu virus at native temperatures. One Health [Internet]. 2024;19:100916. https://doi.org/10.1016/j.onehlt.2024.100916
41. Krambrich J, Bole-Feysot E, Höller P, Lundkvist Å, Hesson JC. Vector competence of Swedish Culex pipiens mosquitoes for Usutu virus. One Health. Elsevier B.V.; 2024;18. https://doi.org/10.1016/j.onehlt.2024.100707
42. Vogels CBF, Fros JJ, Göertz GP, Pijlman GP, Koenraadt CJM. Vector competence of northern European Culex pipiens biotypes and hybrids for West Nile virus is differentially affected by temperature. Parasit Vectors. BioMed Central Ltd.; 2016;9. https://doi.org/10.1186/s13071-016-1677-0
43. Jansen S, Heitmann A, Uusitalo R, Korhonen EM, Lühken R, Kliemke K, et al. Vector Competence of Northern European Culex pipiens Biotype pipiens and Culex torrentium to West Nile Virus and Sindbis Virus. Viruses. MDPI; 2023;15. https://doi.org/10.3390/v15030592
44. Vogels CBF, Göertz GP, Pijlman GP, Koenraadt CJM. Vector competence of northern and southern European Culex pipiens pipiens mosquitoes for West Nile virus across a gradient of temperatures. Med Vet Entomol. Blackwell Publishing Ltd; 2017;31:358–64. https://doi.org/10.1111/mve.12251
45. Soto A, De Coninck L, Devlies A-S, Van De Wiele C, Rosales Rosas AL, Wang L, et al. Belgian Culex pipiens pipiens are competent vectors for West Nile virus while Culex modestus are competent vectors for Usutu virus. PLoS Negl Trop Dis. 2023;17:e0011649. https://doi.org/10.1371/journal.pntd.0011649
46. Paton RS, Vollans M, Glenn L, Fyles M, Vaux AGC, Medlock JM, et al. A faster extrinsic incubation period could explain Usutu leading West Nile in temperate Europe. bioRxiv. 2026; https://doi.org/https://doi.org/10.64898/2026.04.02.716093

47. Vollans M, Day J, Cant S, Hood J, Kilpatrick AM, Kramer LD, et al. Modelling the temperature dependent extrinsic incubation period of West Nile Virus using Bayesian time delay models. Journal of Infection. W.B. Saunders Ltd; 2024;89. https://doi.org/10.1016/j.jinf.2024.106296
48. Jones CE, Lounibos LP, Marra PP, Kilpatrick AM. Rainfall Influences Survival of *Culex pipiens* (Diptera: Culicidae) in a Residential Neighborhood in the Mid-Atlantic United States. J Med Entomol. 2012;49:467–73. https://doi.org/10.1603/ME11191
49. Mwanga EP, Kweyamba PA, Siria DJ, Mshani IH, Mchola IS, Makala FE, et al. Reagent-free detection of Plasmodium falciparum malaria infections in field-collected mosquitoes using mid-infrared spectroscopy and machine learning. Sci Rep. Nature Research; 2024;14. https://doi.org/10.1038/s41598-024-63082-z
50. Koenraadt CJM, Münger E, Schrama MJJ, Spitzen J, Altundag S, Sikkema RS, et al. Overwintering of Usutu virus in mosquitoes, The Netherlands. Parasites and Vectors . BioMed Central Ltd; 2024;17. https://doi.org/10.1186/s13071-024-06620-y

Figures

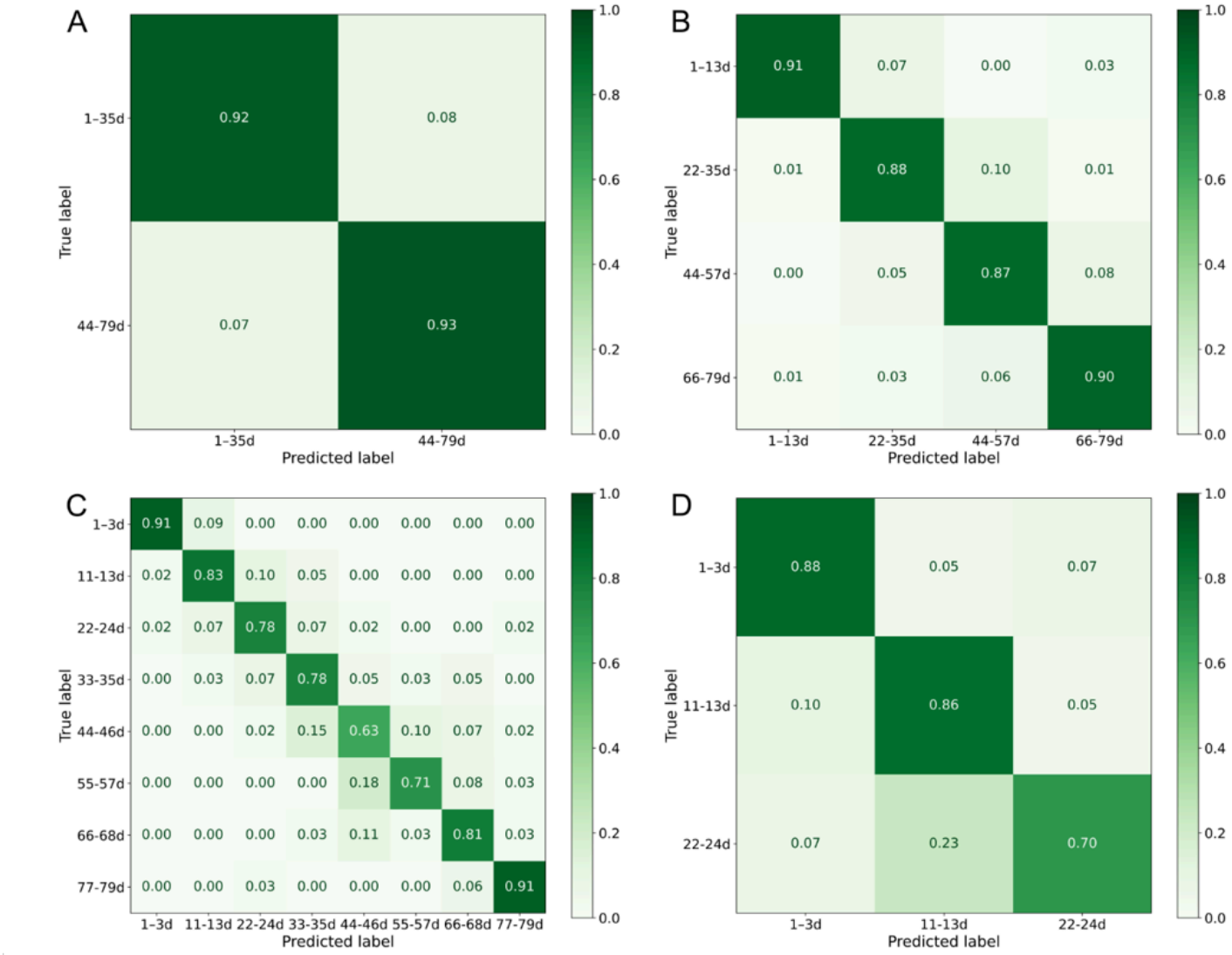


Figure 1

Confusion matrices showing the proportion of accurate (diagonal) classifications of *Cx. pipiens* mosquito age class using the final ML-optimised MLP algorithm in females at (A) low resolution (two age classes); (B) medium resolution (four age classes); (C) and high resolution (eight age classes), (D) and in males (three age classes).

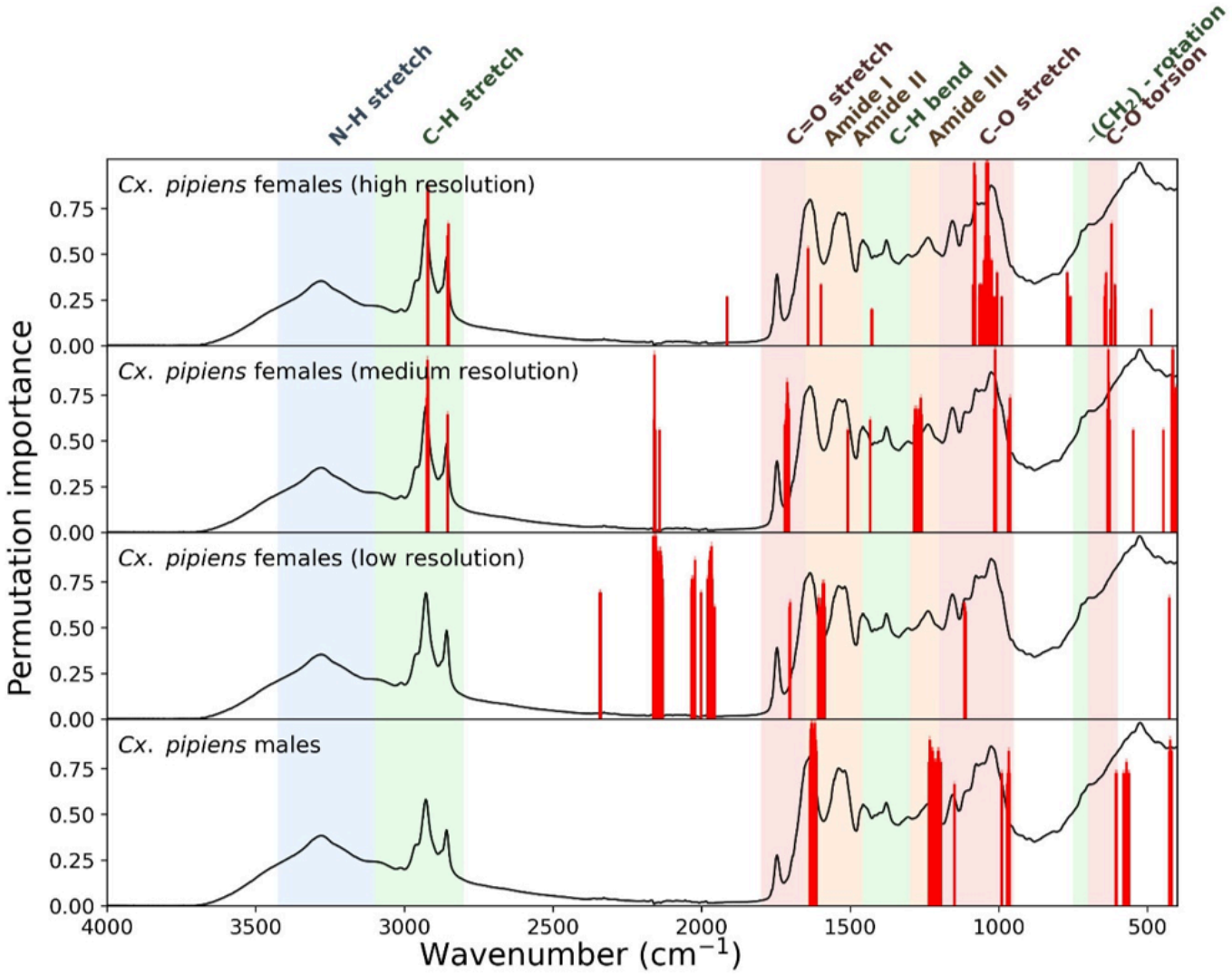


Figure 2

Mean mid-infrared spectra of *Cx. pipiens* female mosquitoes, overlaid with the top-50 most important wavenumbers (red vertical lines) for age determination identified by permutation importance, presented per model. Biochemical functional groups associated with each region are highlighted and indicated at the top of the plot.

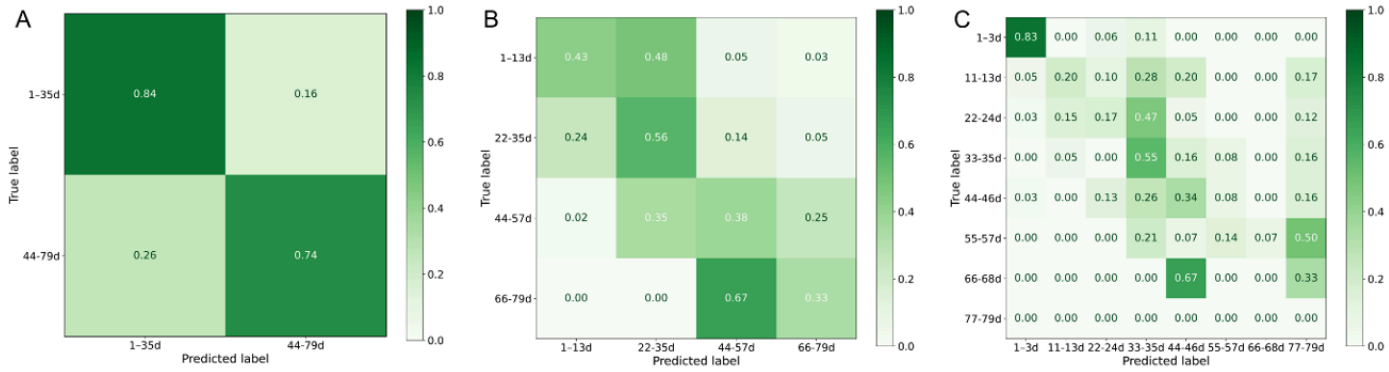


Figure 3

Confusion matrices showing the proportion of accurate (diagonal) classifications of *Cx. pipiens* mosquito age class using the final ML-optimised MLP algorithm in females on the unseen data at (A) low resolution (two age classes); (B) medium resolution (four age classes); (C) and high resolution (eight age classes).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Carmichaeletal2026MIRSSIPVSubmission1.docx