



OPEN Discrimination of inherent characteristics of susceptible and resistant strains of *Anopheles gambiae* by explainable artificial intelligence analysis of flight trajectories

Yasser M. Qureshi¹✉, Vitaly Voloshin^{1,2}, Katherine Gleave³, Hilary Ranson³, Philip J. McCall³, James A. Covington¹, Catherine E. Towers¹ & David P. Towers¹

Understanding mosquito behaviours is vital for the development of insecticide-treated nets (ITNs), which have been successfully deployed in sub-Saharan Africa to reduce disease transmission, particularly malaria. However, rising insecticide resistance (IR) among mosquito populations, owing to genetic and behavioural changes, poses a significant challenge. We present a machine learning pipeline that successfully distinguishes between innate IR and insecticide-susceptible (IS) mosquito flight behaviours independent of insecticidal exposure by analysing trajectory data. Data-driven methods are introduced to accommodate common tracking system shortcomings that occur due to mosquito positions being occluded by the bednet or other objects. Trajectories, obtained from room-scale tracking of two IR and two IS strains around a human-baited, untreated bednet, were analysed using features such as velocity, acceleration, and geometric descriptors. Using these features, an XGBoost model achieved a balanced accuracy of 0.743 and a ROC AUC of 0.813 in classifying IR from IS mosquitoes. SHAP analysis helped decipher that IR mosquitoes tend to fly slower with more directed flight paths and lower variability than IS—traits that are likely a fitness advantage by enhancing their ability to respond more quickly to bloodmeal cues. This approach provides valuable insights based on flight behaviour that can reveal the action of interventions and insecticides on mosquito physiology.

Mosquito-borne diseases are responsible for around 1 million deaths every year¹. These include some of the deadliest viral and parasitic infections affecting humans, such as malaria, dengue, yellow fever, Zika and filariasis². The threat from malaria remains the most significant, with the WHO Africa region accounting for >90% of cases and deaths globally³. In the African region, the main preventative measure is the insecticide-treated net (ITN), the use of which has risen from <5% of households in 2000 to over 50% by 2015³. Although malaria deaths fell by approximately 55% (normalised figures per 100,000 population) in the same period, malaria incidence and mortality rates have largely plateaued since then³. The emergence of resistance to the pyrethroid insecticide used on bednets and its spread across sub-Saharan Africa is believed to be at least partially responsible for the lack of progress in malaria vector elimination targets.

All ITNs currently recommended for use by WHO contain pyrethroid insecticides and resistance to this insecticide class in malaria vectors has rapidly spread since the scale up of ITN use, nearly 25 years ago. Resistance can be caused by physiological changes in the mosquito that reduce the amount of insecticide reaching the target site, reduce the binding to the target, or result in behavioural mechanisms that reduce contact with the insecticide⁴. In the major African malaria vector, *Anopheles gambiae*, multiple mechanisms causing either behavioural or physiological resistance have been reported^{5–7}. Behavioural resistance can arise if changes in the vector population's preferred biting time (night v day) or location (indoor v outdoor) evolve to reduce the likelihood of contact with insecticide on bednets. There have been several reports of shifts in biting

¹School of Engineering, University of Warwick, Coventry CV4 7AL, UK. ²School of Biological and Behavioural Sciences, Queen Mary University of London, London E1 4NS, UK. ³Vector Biology Department, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool L3 5QA, UK. ✉email: Yasser.Qureshi@warwick.ac.uk

patterns^{8–10} following ITN scale up; higher rates of outdoor transmission since the year 2000 are thought to have resulted in millions of additional malaria cases each year¹¹. Nonetheless, the majority of mosquito bites can still be prevented by ITN use. Physiological mechanisms, such as mutations in the sodium channel target or increased rates of detoxification by cytochrome P450s, can cause very high levels of pyrethroid resistance with other, less well-defined mechanisms, further contributing to the phenotype¹². The intensity of pyrethroid resistance, and the underlying mechanisms responsible, differs between populations, sometimes even within a small geographical area¹³.

The link between physiological resistance mechanisms and mosquito behaviour is poorly understood. Fitness effects of specific resistance mechanisms that impact their interaction with ITNs have been reported¹⁴, but limitations in bioassay methods and the ability to measure mosquito behaviour have constrained understanding of the relationship between pyrethroid resistance and the mosquito's interaction with ITNs.

More detailed analyses of vector behaviour has been achieved via video based optical tracking systems including behaviour of mosquitoes at human baited bednets in sub-Saharan Africa¹⁵. That study determined that net contact of less than 1 min per mosquito was sufficient to reduce activity of pyrethroid susceptible mosquitoes to a negligible level 30 min after the mosquitoes were introduced. Insights were gained regarding the preferred location of activity - above the bednet - and direction of arrival - descent from above onto the host from above.

Specific mosquito tracking analyses¹⁶ have used video tracking to compare flight activity of pyrethroid insecticide susceptible (IS; N'goussou and the highly susceptible Kisumu strains) and resistant (IR; VK7 and Banfora strains) strains of *Anopheles gambiae* as they respond to human hosts at different bednets: untreated (UT), Olyset (OL, single pyrethroid active ingredient), Permanet 3 (P3, pyrethroid on all net surfaces, roof with pyrethroid and piperonyl butoxide, PBO) and Interceptor G2 (IG2, pyrethroid and chlorfenapyr on all net surfaces) bednets¹⁶. Significantly higher levels of activity were seen around an untreated net in comparison to the 3 ITNs, at which the total activity, the number and duration of net contacts, were similar at all ITNs for both susceptible and resistant strains. There was a steep decay in mosquito activity after approximately 20 min, indicative of knockdown or mortality, for the IS strains at OL and P3 ITNs, but only with the highly susceptible Kisumu at IG2. A slow decline in activity was measured with OL, P3 and IG2 nets with the IR strains, indicating the slower impact of the second insecticide on the IR strains as they are resistant to the fast acting 'knockdown' effect of the pyrethroid.

In this study we investigated whether a more detailed examination of mosquito flight, using machine learning models, could identify basic or inherent behaviours that distinguish IS and IR populations, prior to insecticide contact. Explainable AI (XAI) is an emerging field that attempts to interpret machine learning models for better transparency and understanding. By uncovering the rationales behind decision-making, XAI can raise confidence in AI systems and enhance their real-world adoption. For example, Ryo et al. successfully employed XAI to interpret ML species distribution models boosting the usability and interpretability of the ecological models for further research¹⁷. That study demonstrated the potential for XAI to assist ecologists identify the underlying factors, connecting the key features that influence species distribution within an ML model.

Here, we applied XAI techniques to detect fundamental differences between insecticide-susceptible (IS) and insecticide-resistant (IR) mosquito strains using trajectory features. This study utilised trajectory data of malaria vectors orienting to a human host within an untreated bednet in order to examine behaviours without any effect of insecticides, i.e. the innate behaviours and inherent flight characteristics of pyrethroid resistant and susceptible strains of primary malaria vector *Anopheles gambiae* s.l.. This approach provides insights into the baseline behavioural and physiological characteristics associated with pyrethroid resistance. These findings establish a foundation for understanding how insecticide resistance may shape mosquito behaviour, with implications for improving vector control strategies.

Methods

Our approach for classifying the mosquito trajectories consists of several key steps. Mosquito trajectories are first split into segments using a sliding window method. These segments can then be used to generate features that describe the flight behaviour of the mosquitoes, such as the mean velocity, turning angle, and angular acceleration. The features are then provided to a machine learning model which is trained to classify segments as IS or IR. The classified segment predictions are combined using a voting method to form whole track predictions. The last stages include model evaluation and model interpretation using XAI techniques.

Throughout this work, a mosquito track was considered as a two-dimensional track or trajectory, T , that can be described as follows, where x_i and y_i correspond to the i th position within a Cartesian coordinate system, and t_i is time:

$$T = \{(x_i, y_i, t_i) \text{ for } i = 0, \dots, N\} \quad (1)$$

Dataset description

The dataset used was generated within laboratory experiments at the Liverpool School of Tropical Medicine (LSTM), UK. In each experiment, three-to-five-day old unfed adult female mosquitoes from one of insecticide-susceptible (Kisumu and N'goussou) or insecticide-resistant (VK7 and Banfora) *An gambiae* strains were tracked around an untreated human-baited bednet for 2 h¹⁶. The Kisumu strain originated from Kenya in 1975 while the N'goussou strain was established in 2006 from Cameroon¹⁸. VK7 and Banfora strains originated from Burkina Faso, established in 2014 and 2015, respectively. Both resistant strains exhibit high pyrethroid resistance due to fixed kdr mutations (L1014F) and metabolic resistance involving elevated P450 expression¹⁸. Further information about the strains used can be found in¹⁸.

To maintain resistance phenotypes in the insecticide-resistant strains, mosquitoes were exposed to 0.05% deltamethrin-treated papers every 3–5 generations using the WHO susceptibility bioassay protocol. Resistance

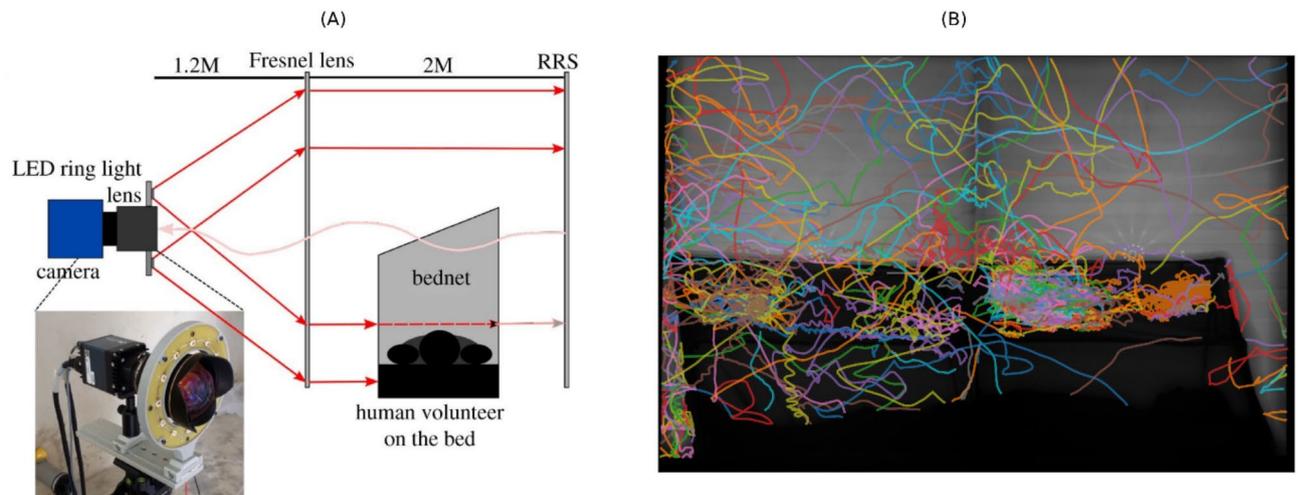


Fig. 1. (A) Schematic of the experimental setup, featuring the camera and imaging lens surrounded by an LED ring light, a Fresnel lens, and a retroreflective screen for tracking mosquito flight trajectories. This figure is adapted from¹⁹. (B) Example of tracking results from a Kisumu trial, where each coloured line represents the flight path of an individual mosquito, visualising the movement patterns of multiple mosquitoes during the first 15 min of the trial.

Strain	Insecticide-resistance status	Number of trials	Number of tracks	Average track duration (s)	Average velocity (mm/s)
Kisumu	Susceptible	5	7631	22.39 (1.00-698.11)	344.29 (5.20-2290.84)
N'goussu	Susceptible	4	6125	21.40 (1.00-1010.03)	453.78 (3.54-2602.18)
Banfora	Resistant	4	7189	24.26 (1.00-475.42)	389.57 (0.74-2478.15)
VK7	Resistant	4	5862	17.04 (1.00-457.77)	429.54 (4.10-2102.89)

Table 1. Information on the trajectories of each mosquito strain.

profiles were monitored through annual bioassays assessing susceptibility to multiple insecticides, including pyrethroids, carbamates, and organophosphates, ensuring the strains retained their resistance traits over time. Additionally, molecular genotyping was routinely performed to verify the presence of *kdr* (L1014F) and metabolic markers, including elevated cytochrome P450 expression, confirming stability of resistance mechanisms¹⁸.

All experiments, also known as trials, were conducted between June 2019 and February 2020 using a custom built free-flight climate-controlled testing room (7 × 4.8 m in area, 2.5 m high). The experiments were performed in the afternoon to coincide with the 'night' phase of the mosquito's circadian rhythm when they would be host-seeking in the wild. Mosquitoes were tracked using paired identical recording systems, where each recording system used one camera (12 MPixel Ximea CB120RG-CM with a 14 mm focal length lens), aligned with a single Fresnel lens (1400 × 1050 mm and 3 mm thick, 1.2 m focal length; NTKJ Co., Ltd, Japan) placed approximately 1210 mm away. Retroreflective screens (2.88 m², coated with high-gain sheeting) were positioned behind the bednet to maximise light capture and improve contrast for video tracking¹⁹. The system operated under infrared illumination (850 nm wavelength) to enable night-phase tracking without visual disruption, capturing trajectories at 50 frames per second (fps). The setup is telecentric and produces 2D data on mosquito flight, the 2 cameras are necessary to increase the field of view whilst maintaining spatial resolution. Further information on the experimental set up is outlined in¹⁶ and the extraction of trajectories from video recordings is outlined in¹⁹. Figure 1 illustrates the experimental setup and example tracking results. Table 1 summarises the dataset used, where a strict limit on the track duration is set (above 1 s).

The primary objective of this work is to explore differences between IR and IS mosquito flight behaviour, i.e. to consider Banfora and VK7 data together as IR and Kisumu and N'goussu as IS, as this will give more generic understanding and a larger dataset for the computational model. The outcomes of the IR vs. IS model are analysed using the SHAPs XAI toolset²⁰ discussed in Sect "Evaluation and Interpretation". Further comparisons have been explored to investigate strain-specific behavioural variations and ML model performances are reported for pairwise comparisons (Kisumu vs. N'goussu; Banfora vs. VK7) as well as a multiclass model considering the 4 mosquito strains against each other. The multiclass approach provides additional insights into strain-specific differences that may not be captured in the binary classification, enabling a more granular analysis of behavioural patterns across strains.

Data processing

The dataset contains trajectories of varying durations, which can have an impact on the features generated and, in turn, affect the classification by the machine learning models. This variation in duration arises due to natural mosquito behaviour and the recording process. Small obstructions in the camera view, the mosquitoes sitting or walking on the bednet, as well as mosquitoes leaving the camera view, can break tracks into various lengths. Therefore, the trajectories in the dataset may only represent a portion of any mosquito's flight path, and the reason there are different durations for each trajectory. To address this issue, trajectories were split into fixed duration segments using a moving window approach to unify track duration and eliminate duration bias for the machine learning model²¹.

The window size and overlap between consecutive segments becomes a hyperparameter of the pipeline. Longer segments provide more information per segment but makes the model larger and increases the risk of overfitting. Larger overlap creates more segments per track improving per trajectory prediction but generates a lot of very similar segments again causing a risk of the model overfitting. To obtain a suitable window size and overlap, these parameters were optimised as described in Sect. "Hyperparameter tuning". The windowing approach ensures that each segment had the same duration, and thus the same amount of information, whilst excluding the length of the track as a feature (directly and indirectly) and removing the possibility for data leak. As a result of this process, only trajectories that have a length greater than or equal to the optimised segment length are preserved to train the ML model.

As a result of the recording process, some positions within the mosquito tracks are missed. This can be due to mosquitoes flying in areas that were entirely obscured or encountering regions with poor contrast. To maintain the continuity of the tracks, linear interpolation was applied to fill in the gaps²². However, in some instances, the tracks exhibited substantial gaps introducing bias into the data. Across all tracks, there was a median gap size of 0.02 s which is expected for a 50 fps camera system. However, there was a mean gap size of 0.07 s with a standard deviation of 0.16 s. To address this issue and minimise its impact, we introduced a segment quality metric. It evaluated a track segment based on its information content, assigning higher scores to segments with more consecutively interpolated positions, as they contribute relatively little information. To establish a threshold for the metric, the mutual information of the tracks with the target variable (IR/IS status) at various thresholds of the segment quality metric was computed, where segments below this threshold were used, and the weighted average of the maximum thresholds that obtained the maximum mutual information for each trajectory feature was used as the final threshold. Track segments with segment quality scores larger than this threshold were removed. Various segment quality scores and thresholding techniques were assessed and are outlined in the supplementary information.

Feature extraction

The extraction of meaningful features from the mosquito track segments is an essential step in the analysis of their flight patterns. The features can be broadly categorised into two types: shape descriptors and kinematic features. Shape descriptors capture the geometric characteristics of the tracks, such as the curvature of the path. On the other hand, kinematic features are based on time-dependent characteristics describing the movement dynamics, such as the speed, acceleration, and turning angle of the mosquito. After extraction of these characteristics, statistics (mean, standard deviation etc.) are computed for each feature over each segment and used as features for the model. As mentioned in the previous section, some of the positions within the tracks are interpolated so contain artificially generated positions. Feature statistics are calculated at positions containing real (not interpolated) data within a segment. The calculations for the features of flight are detailed in²¹ for 3D trajectories and²³ for 2D trajectories and they are also provided in the supplementary.

Dataset partitioning

The dataset was partitioned into two subsets: a tuning set and a modelling set. The tuning set was designated for feature selection and hyperparameters tuning. It comprises 2 trials for each strain, resulting in a total of 8 trials. Meanwhile, the modelling set, with a total of 9 trials, is dedicated to machine learning model training and evaluation. Splitting the dataset by trials prevents data leakage and simulates a real-world scenario where an entire trial is tested. This division allows for robust model development and comprehensive evaluation of the machine learning models through cross-validation as well as elimination of data leaks.

Feature selection

Having extracted meaningful features from trajectory segments (see Sect. "Feature extraction" for feature extraction), the Mann-Whitney U-test, a non-parametric statistical test, was employed to select relevant features using SciPy²⁴. To mitigate against group testing p-value inflation, a family-wise error rate (FWER) controlling procedure utilising Bonferroni correction was used. Features that demonstrated statistically significant differences between groups with rejection of the null hypothesis at an FWER < 0.05 were selected. Additionally, a Spearman correlation matrix was computed to identify collinearity between features. Pairs of features with high correlation ($\rho > 0.85$) were flagged, and one feature from each pair was removed to minimise redundancy, prevent multicollinearity, and enhance model stability and generalisability.

Classification model

The primary objective of this study was to accurately differentiate between insecticide-susceptible and insecticide-resistant mosquito species using an ML classifier. To achieve this, logistic regression, random forest, and XGBoost classification^{25,26} models were evaluated. In order to generate whole track predictions, every track segment was classified independently and the mode of the segment binary predictions for each track used as the prediction for the complete track.

Prior to training the model, several processing steps were taken. Each feature was standardised through Z-score normalisation where the mean and standard deviation values were calculated from the training set. Then to balance the distribution of segments in each class, the training set was oversampled using SMOTE (Synthetic Minority Oversampling Technique)²⁷. These steps mitigate the models fitting to the imbalance of the dataset or to prioritise some features due to difference in the magnitude of their values.

Evaluation and interpretation

The performance of the proposed framework was evaluated on the modelling dataset using a cross-validation approach. The dataset split in each iteration of this process is known as a fold, with each fold consisting of a different combination of two insecticide-resistant (IR - one from each IR strain) and two insecticide-susceptible (IS - one from each IS strain) trials in the training set, with the remainder of the trials used for model testing. This approach is necessary to maintain consistency in training across different strains while ensuring robustness in performance evaluation. Although the number of track segments may differ between folds, the ratio of IR to IS segments is balanced using SMOTE to prevent bias and improve generalisability. Overall, there were 24 folds. When evaluating the performance of the training set, it should be noted that only those segments that were not generated via the oversampling technique (i.e., segments that constitute a genuine track) in the training set were used to derive the scores for the final track prediction.

Various performance metrics are calculated using Scikit-learn²⁵ including accuracy, balanced accuracy, ROC AUC (area under the receiver operator characteristic curve) score, PR AUC (area under the precision-recall curve) score, F1 score, precision score, recall score, Matthew Correlation Coefficient (MCC), Cohen kappa coefficient and log-loss score. The performance metrics are calculated on the whole track predictions where the arithmetic mean, and the minimum and maximum of performances across all folds is provided.

Balanced accuracy provides an average of recall across classes, making it particularly useful for imbalanced datasets, with values ranging from 0 to 1 where higher values indicate better performance. The F1 score balances precision and recall, providing a single metric that is particularly useful when false positives and false negatives carry similar costs. It ranges from 0 to 1, with higher values indicating better performance, especially in scenarios with class imbalances. The ROC AUC score measures a model's ability to distinguish between classes, with a score of 0.5 representing random guessing and values closer to 1 indicating strong discriminatory power. ROC curves illustrate the trade-off between sensitivity (true positive rate) and specificity (1 - false positive rate), where curves closer to the top-left corner represent better model performance. Together, these metrics offer comprehensive insights into a model's ability to classify data accurately and handle class imbalances. Further information on each performance metric, alongside their equations, is detailed in the supplementary. To determine whether the model's balanced accuracy was significantly greater than chance (0.5 for binary classification or 0.25 for multiclass classification), a Wilcoxon signed-rank test was performed on the per-fold differences. To control the FWER across multiple models, a Bonferroni correction was applied by multiplying each p-value by the number of models tested.

Confusion matrices were also used to highlight model performance by displaying the normalised (by true class) percentages of true and false predictions for IR vs. IS classifications. High diagonal values indicate accurate predictions, while off-diagonal values represent misclassifications. Under random chance, predictions are equally distributed across classes, so each row would show 50% in each cell. This normalisation simplifies the assessment of model performance within each class and highlights patterns in misclassification.

SHapley Additive ExPlanations (SHAP)²⁰ values are a powerful tool for interpreting machine learning models by quantifying how much each feature contributes to a prediction. SHAP values explain whether a feature pushes the model's output higher (toward classifying a mosquito as IR) or lower (toward classifying it as IS) compared to the baseline prediction. In the context of studying insecticide resistance in mosquitoes, SHAP values were calculated for track segments to reveal behavioural differences between IR and IS strains. Positive SHAP values indicate that a feature contributes toward a prediction of IR, while negative SHAP values suggest that the feature pushes the prediction toward IS.

SHAP provides various visualisations to aid interpretation. SHAP summary plots provide an overview of feature importance and direction of influence, showing which features most strongly differentiate the two classes. SHAP bar plots rank features by their average absolute impact, highlighting the most influential predictors across the dataset. SHAP scatter plots further explore individual feature effects, displaying relationships between feature values and their contributions to predictions, which can uncover non-linear patterns or interactions. Together, these visualisations provide a detailed and interpretable analysis of the features driving model decisions, helping to characterise behavioural patterns associated the classes.

Hyperparameter tuning

Hyperparameter tuning was employed to identify the best parameters for the model while minimising the risk of overfitting. Each machine learning model has various parameters that require tuning, as well as the window size and overlap length used to split tracks to segments. The parameters were tuned together in a cross-validated grid search approach attempting to maximise balanced accuracy. Tuning was conducted using two trials from each strain resulting in four IR trials and four IS trials. Each fold in the cross-validated grid search included 1 of each strain in the training set with the remainder in the test set. This ensures that the selected parameters are optimised based on an unseen dataset over a 5-fold cross validation technique. A full description of the parameter ranges and step sizes, as well as the full set of optimised hyperparameters identified for each model can be found in the supplementary.

Classification task and model	Window size (s)	Window overlap (s)	Final number of tracks	Number of segments
IR vs. IS, XGBoost	8	7.5	10,514	695,509
IR vs. IS, Random Forests	7.5	6.5	10,943	357,120
IR vs. IS, Logistic Regression	9.5	9	9,589	664,369
IS vs. IS, Kisumu vs. N'goussu, XGBoost	1.5	0.5	11,402	192,150
IR vs. IR, Banfora vs. VK7, XGBoost	6	5.5	6,017	352,761
Multiclass, XGBoost	6	5.5	11,887	683,967

Table 2. Windowing parameters for each model alongside the final number of tracks and segments.

Performance metric	Logistic regression	Random forests	XGBoost
Balanced accuracy	0.723 (0.682–0.759)	0.721 (0.680–0.772)	0.743 (0.688–0.786)
ROC AUC score	0.788 (0.734–0.831)	0.806 (0.760–0.859)	0.813 (0.757–0.866)
F1 score (IR)	0.686 (0.642–0.739)	0.687 (0.638–0.747)	0.702 (0.646–0.757)
F1 score (IS)	0.746 (0.675–0.797)	0.730 (0.658–0.776)	0.765 (0.725–0.802)
Recall (IR)	0.725 (0.631–0.798)	0.758 (0.657–0.888)	0.739 (0.628–0.883)
Recall (IS)	0.720 (0.591–0.779)	0.684 (0.543–0.780)	0.746 (0.626–0.852)
Precision (IR)	0.656 (0.540–0.750)	0.637 (0.522–0.746)	0.682 (0.574–0.801)
Precision (IS)	0.780 (0.706–0.869)	0.792 (0.698–0.914)	0.794 (0.694–0.918)
PR AUC (IR)	0.714 (0.621–0.788)	0.750 (0.693–0.800)	0.735 (0.681–0.793)
PR AUC (IS)	0.812 (0.747–0.883)	0.841 (0.779–0.919)	0.840 (0.777–0.917)
Cohen Kappa coefficient	0.435 (0.351–0.508)	0.425 (0.327–0.528)	0.472 (0.373–0.556)
Matthew correlation coefficient	0.440 (0.361–0.509)	0.436 (0.356–0.538)	0.480 (0.374–0.556)
Log loss	0.567 (0.523–0.635)	0.565 (0.539–0.605)	0.552 (0.497–0.594)

Table 3. Performance of each machine learning model applied to independent test data for IR (Banfora & VK7) vs. IS (Kisumu & N'goussu), with the arithmetic mean and the range provided across all folds.

Results

Data processing

Trajectories were split into segments to mitigate the impact of duration imbalance between tracks. Through hyperparameter tuning, various window sizes and overlap lengths for the windowing technique were tested. Each model performed best with different window parameters. Table 2 displays the window parameters selected after tuning and the effect they have on the number of tracks (tracks shorter than the segment length are discarded) and segments within the dataset partitioned for modelling. Note that the numbers of tracks and segments were after segment quality filtering.

Model evaluation – classification of insecticide resistance status

The performance of the classification of insecticide resistance status, IR or IS, is provided in Table 3. The arithmetic mean and the range (minimum and maximum) of scores across all folds are provided in brackets. Wilcoxon signed-rank test comparing each model's per-fold balanced accuracy to 0.5 indicated that all three models were statistically significant ($p = 7.15 \times 10^{-7}$). The confusion matrix for each model is displayed in Fig. 2 which shows the percentage of the predictions over all folds, and the ROC curves are depicted in Fig. 3 featuring each folds' curve.

The SHAP plots for the best performing fold for the XGBoost model applied to independent test data are provided below. This includes SHAP summary plot, Fig. 4, and SHAP bar plot, Fig. 5.

Velocity in the vertical direction is one of the strongest contributors to the model. There are 5 vertical-velocity features, ranked 1, 2, 6, 11 and 27 in the SHAPs bar plot. These are the 3rd quartile (identifying 75% of the population have lower values), 1st quartile (25% of the population), standard deviation, kurtosis and skewness respectively of the vertical-velocity distributions within a track segment. Figure 6 displays the SHAP scatter plots for these vertical velocity features. A histogram comparison of vertical velocities for all IR and IS tracks, Fig. 7, reveals distinct distributions that are consistent with the SHAP scatter plots.

Fractal dimension is another strong contributor in the separation between IR and IS, ranked 3rd in the SHAP bar plots. This feature is a measure of the linearity or complexity of the trajectory. Figure 8 shows the probability density histogram of fractal dimension for trajectory segments with the XGBoost segment length (8 s) for each strain separately.

Model evaluation – exploring mosquito strains

In addition to the analysis above, further target classifications were explored to assess the differences between the mosquito strains. Namely, whether in-class mosquito strains were separable and how different each strain was from one another.

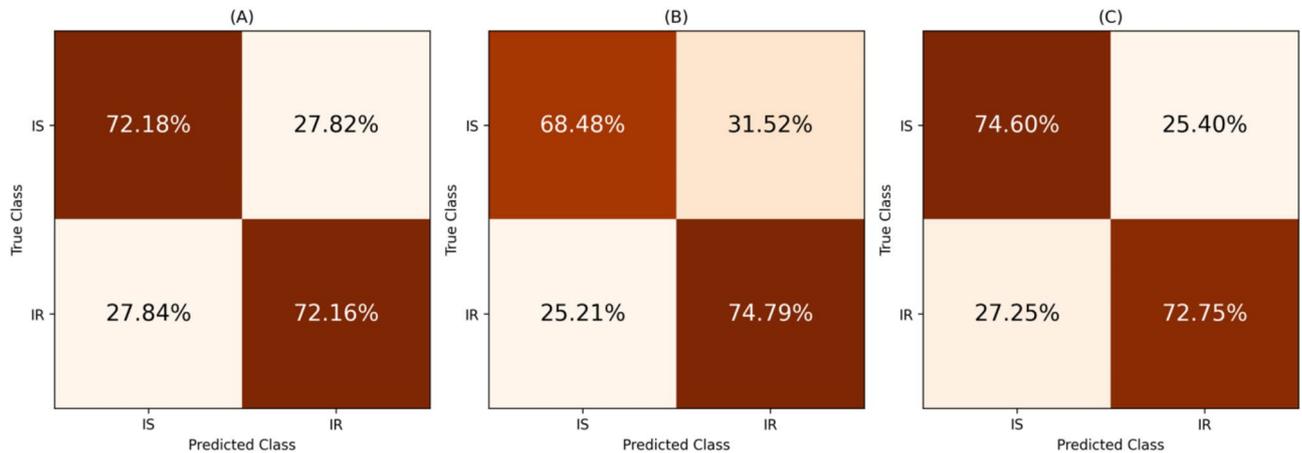


Fig. 2. Confusion matrices from independent test data displaying the percentages of the predictions over all folds for IR vs. IS. In the figure, (A) displays the logistic regression model, (B) is the random forests model and (C) is the XGBoost model. The matrices compare predicted classes (columns) with true classes (rows); diagonal cells represent correct classifications, while off-diagonal cells indicate misclassifications.

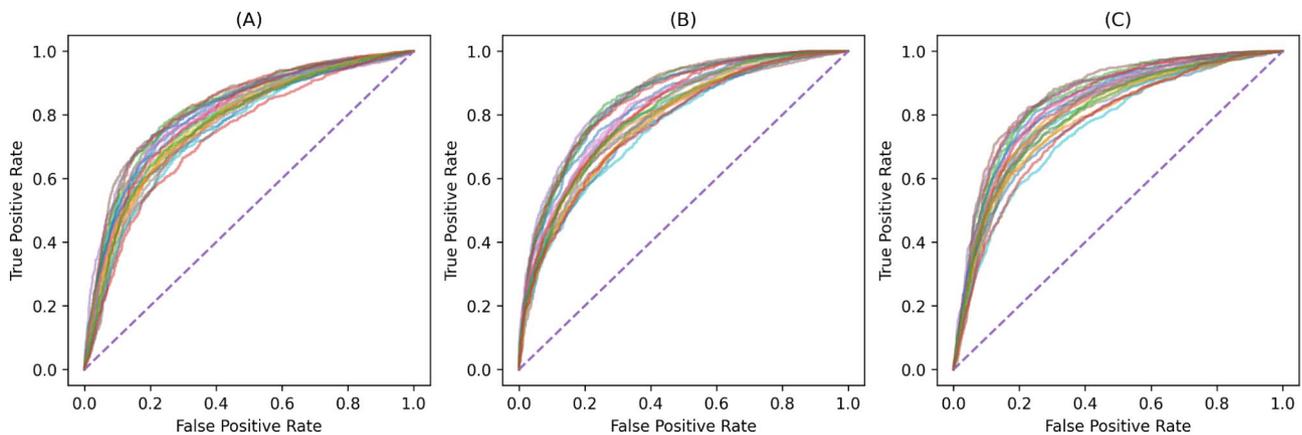


Fig. 3. ROC Curves from independent test data for each IR vs. IS model, where (A) displays the logistic regression model, (B) is the random forests model and (C) is the XGBoost model. The ROC curve plots the true positive rate (sensitivity) against the false positive rate (1 – specificity) at different thresholds; curves closer to the top-left indicate stronger performance. Each line represents a single fold’s ROC curve.

Classifying between insecticide susceptible (IS) strains

The differences within the insecticide susceptible class were explored, where a binary classification was attempted between the IS strains, Kisumu and N’goussu, using the XGBoost classifier. Table 4 displays the performance of this task. The Wilcoxon signed-rank test indicated that the balanced accuracy observed across folds does not deviate significantly from random performance (0.5), achieving a p-value of 0.19. The confusion matrix is displayed in Fig. 9 (A) which displays the percentage of the predictions over all folds. Additionally, the ROC curve is depicted in Fig. 9 (B) featuring each fold’s curve.

Classifying between insecticide resistant (IR) strains

Differences within the IR class were also explored using binary classifiers, involving the Banfora and VK7 strains. Table 5 displays the performance of this task. The Wilcoxon signed-rank test revealed that the balanced accuracy across folds wasn’t statistically significantly higher than chance level (0.5), with a p-value of 0.75. The confusion matrix is displayed in Fig. 10 (A) which shows the percentage of the predictions over all folds. Additionally, the ROC curve is in Fig. 10 (B) featuring each fold’s curve.

Classifying between each strain

To assess the differences between each strain, we used a multiclass classification approach. The proposed pipeline remained mostly unchanged, however in this case the Mann Whitney U-test was performed pairwise for all strains and the multiclass XGBoost model was used. The results of the model are displayed in Table 6. The performance of each individual strain is also provided in Table 7, by calculating the model accuracy for each

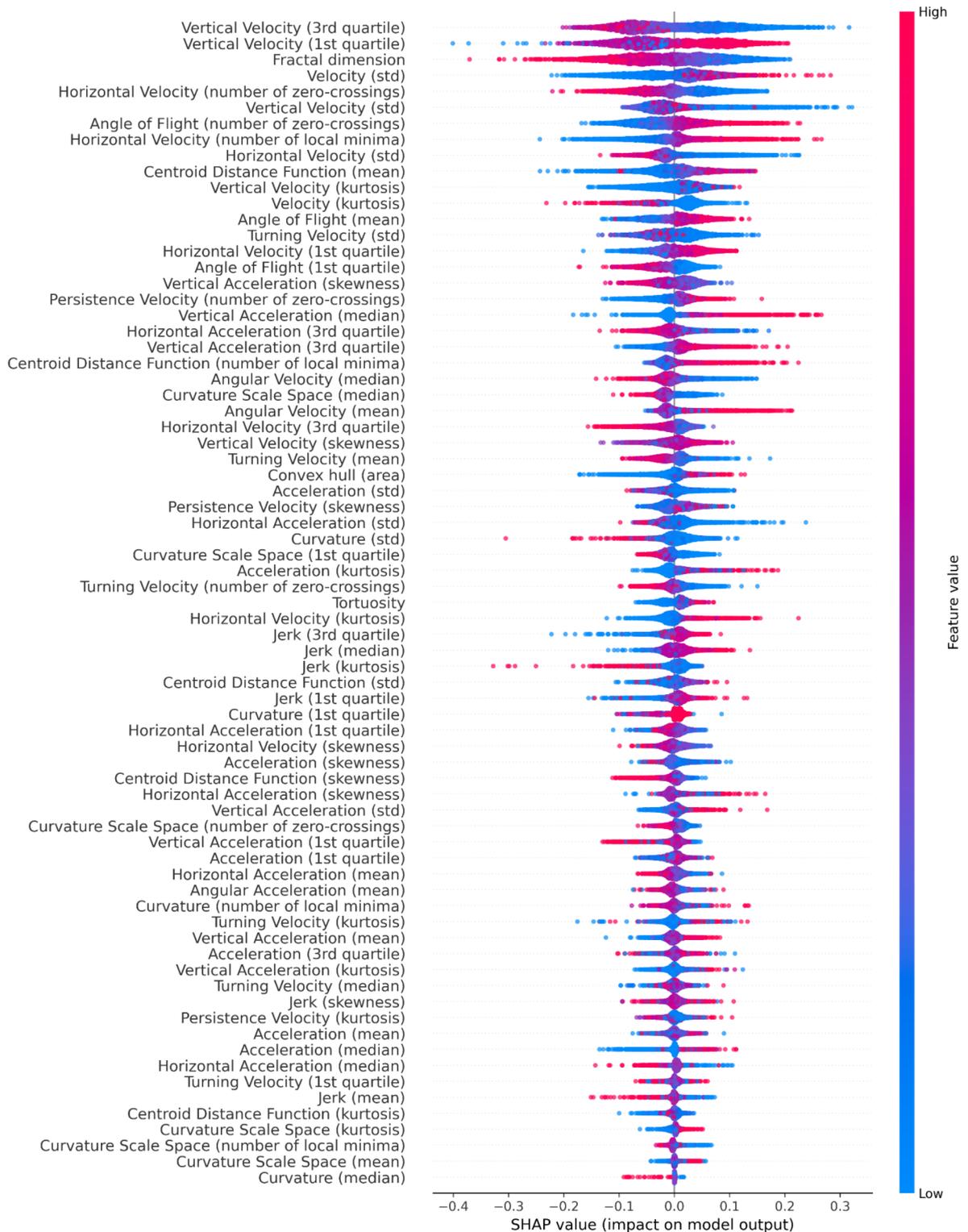


Fig. 4. SHAP summary plot for the best IR vs. IS XGBoost model fold applied to independent test data, displaying the feature contributions to model predictions. Features are ranked by their mean absolute SHAP, indicating their overall importance. Each dot represents a track segment with position on the x-axis showing the SHAP value (contribution) and its colour representing the feature value (low to high). Positive SHAP values push the predictions towards the IR class, while negative SHAP values push the predictions towards the IS class. Wider spreads indicate greater variability in feature influence across segments. This visualisation highlights both the magnitude and direction of feature effects, helping to interpret behavioural differences between strains.

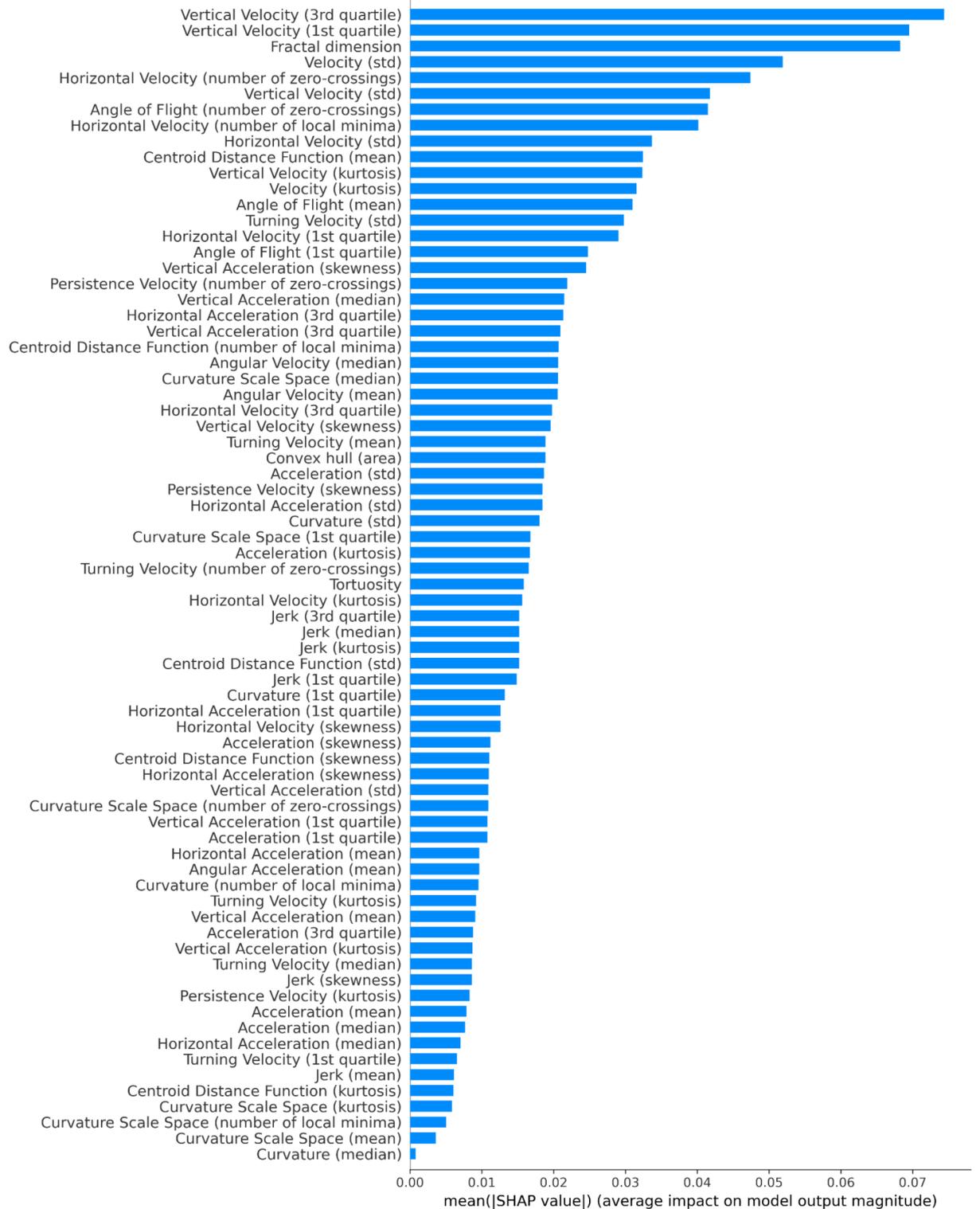


Fig. 5. SHAP bar plot for the best IR vs. IS XGBoost model fold applied to independent test data, showing the mean absolute SHAP values for each feature. Features are ranked by their average contribution to the model's predictions, with higher values indicating greater importance.

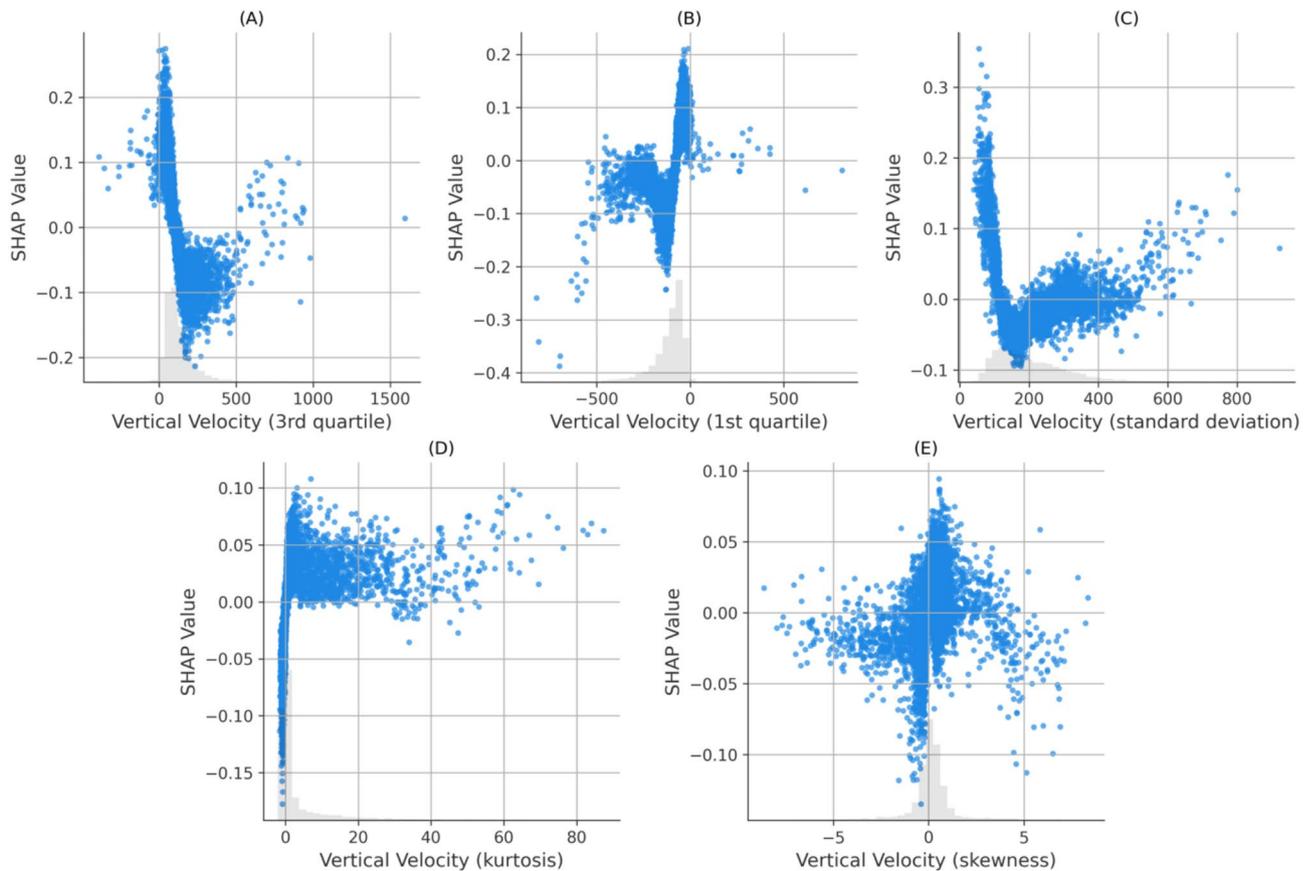


Fig. 6. SHAP scatter plots illustrating the influence of vertical velocity features on model predictions for IR vs. IS classification. Each plot shows the SHAP values (y-axis) against the feature values (x-axis), revealing the relationship between feature magnitude and its contribution to predictions. Positive SHAP values indicate a push toward the IR class, while negative values push toward the IS class. (A) is the 3rd quartile of vertical velocity, (B) is the 1st quartile of vertical velocity, (C) is the standard deviation of vertical velocity, (D) is the kurtosis of vertical velocity, and (E) is the skewness of vertical velocity.

strain at each fold. The Wilcoxon signed-rank test demonstrated that the balanced accuracy across folds was significantly higher than the random-chance baseline (0.25), with a p-value of 7.15×10^{-7} . The confusion matrix is displayed in Fig. 11 which displays the sum of the predictions over all folds.

Discussion

ML models have been successful in classifying different mosquito strains based on innate flight trajectory features. The classification outcomes for the IR vs. IS model with 2 mosquito strains in each class show the strongest performance using the non-linear XGBoost classifier with balanced accuracy and ROC AUC of 0.743 and 0.813 respectively. The model's performance is consistent across all folds of the cross-validation process, see Fig. 3. Importantly, this classification was achieved without relying on data reflecting mosquito responses to insecticides. Instead, it leveraged inherent differences in baseline behaviours between the classes. Responses of the same four IS and IR strains were investigated previously as they fed through ITNs and untreated (UT) nets, which showed that the blood volume flow rate of the susceptible strain increased by 35% in presence of insecticide, but the resistant strain was already at the higher rate for both UT and ITNs²⁸. These prior findings align with the results presented here, reinforcing that IR strains exhibit distinct behavioural patterns even in the absence of insecticide stimuli.

Whilst exploring the differences between IR and IS mosquito classes, the differences within the classes was examined. Within the susceptible class, the classifier identified notable differences between the Kisumu and N'goussu strains with an average balanced accuracy of 0.655 and a ROC AUC score of 0.726. Similarly, the classifier identified differences between the resistant strains, VK7 and Banfora, with an average balanced accuracy and ROC AUC score of 0.662 and 0.717, respectively. This illustrates that even though these strains are similar in their insecticide resistance, there are still identifiable differences between them. Yet, despite the differences within the classes, they still share common characteristics that enable the separation between IR and IS classes. A multiclass approach was also explored to identify distinctiveness of all four strains. The performance of this approach was slightly better than chance (score of 0.25) with an average balanced accuracy of 0.496. This may indicate that while some behavioural differences exist, there are common behaviours across

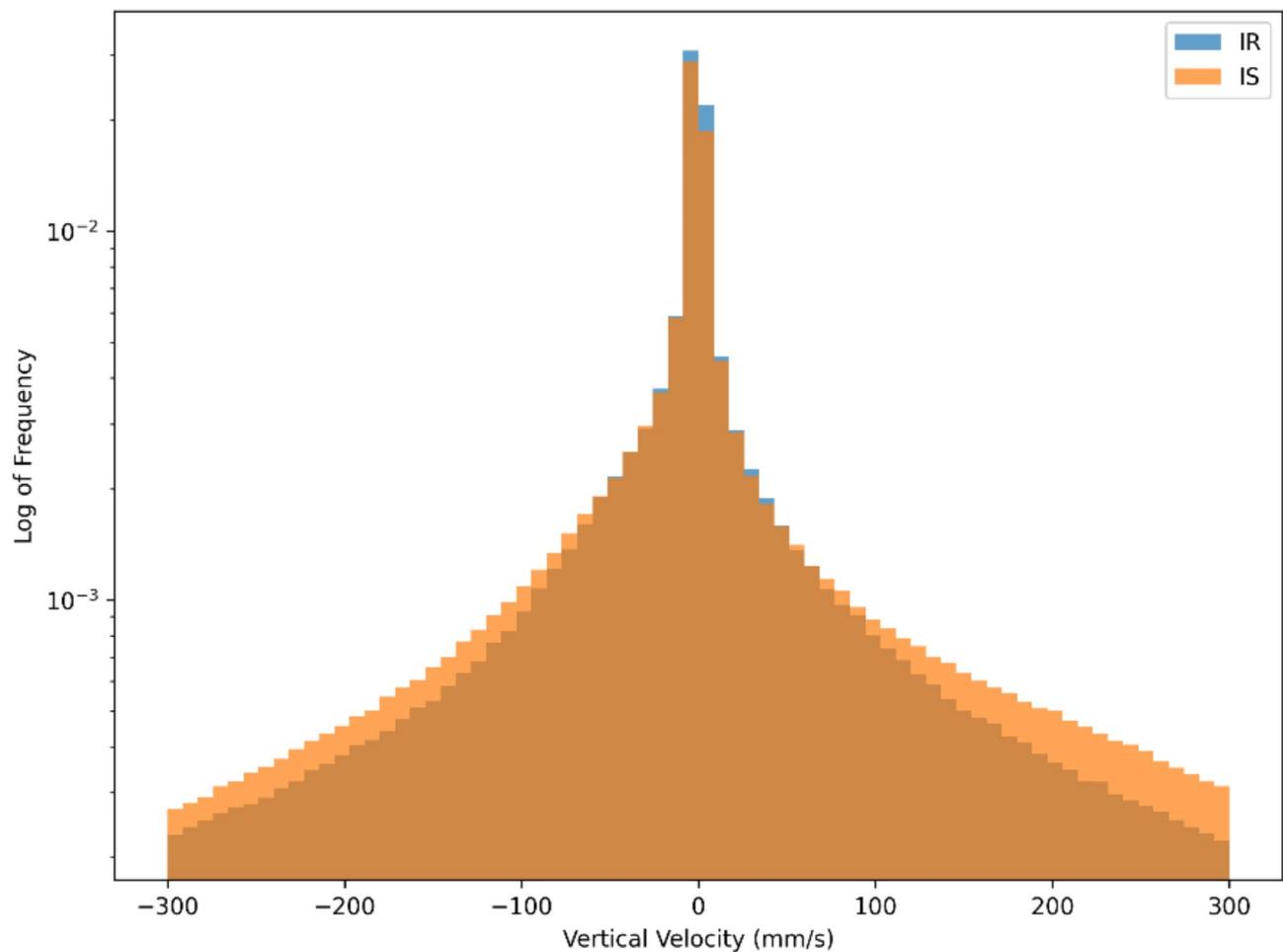


Fig. 7. Histogram of vertical velocities for IR and IS mosquito trajectories. The velocity range is restricted within the range -300 mm/s to 300 mm/s to exclude outliers and extreme values. The plot uses logarithmic scaling for frequency.

some or all mosquito strains. Nevertheless, an interesting feature of the multiclass results was the stronger separability of the Kisumu strain as well as the poor separability of VK7 (recall that features are evaluated over fixed duration segments and do not include activity levels). It is conjectured that the lower activity of VK7 may lead to its behaviour in short trajectory segments being encompassed by the other strains. Additionally, it was observed that predictions for the N'goussu strain were closely split between Kisumu and N'goussu, suggesting potential behavioural similarities or overlaps between these two strains. This aligns with the earlier results where differences within the susceptible and resistant classes were noted, reinforcing that despite shared class characteristics, individual strains exhibit distinct behaviours. In comparison, the stronger separability within the IR and IS classes may reflect genetic differences between the two IS strains and two IR strains. The susceptible strains are from different sides of the African continent (Kisumu originates from Kenya whereas N'goussu is from Cameroon) and were colonised more than 30 years apart. The two IR strains are both from southwest Burkina Faso, were both colonised in 2015 and both show similar phenotypic levels of pyrethroid resistance¹³. However, the pyrethroid resistance in VK7 (2014) is largely conferred by target site resistance and elevated cytochrome P450 activity and whilst the Banfora strain has both of these resistance mechanisms there is also an indication that increased rates of respiration, and potentially changes in the microbiome also contribute to the resistance phenotype²⁹. The strongest performance is from the IR vs. IS model utilising information from 2 genetically distinct IS and IR strains suggesting that there are common behavioural traits that evolve with IR strains that may associate with fitness benefits as described above.

The SHAPs analysis of the XGBoost IR vs. IS model has identified the features which most strongly differentiate IR and IS behaviours. Interestingly, Fig. 5 shows that there are non-negligible contributions for the vast majority of the features in the model – indicating that all features play a minor role in classification. IR behaviours are characterised by moving more slowly in the vertical direction than IS and hence expending less energy in flight. This slower vertical movement might allow IR mosquitoes to modulate their trajectories more effectively in response to environmental cues, although this hypothesis was not explicitly tested here. This also implies that on average an IR mosquito has lower momentum than IS and can change direction more quickly. However, it is important to note that the differences observed between IR and IS are relatively subtle and do not

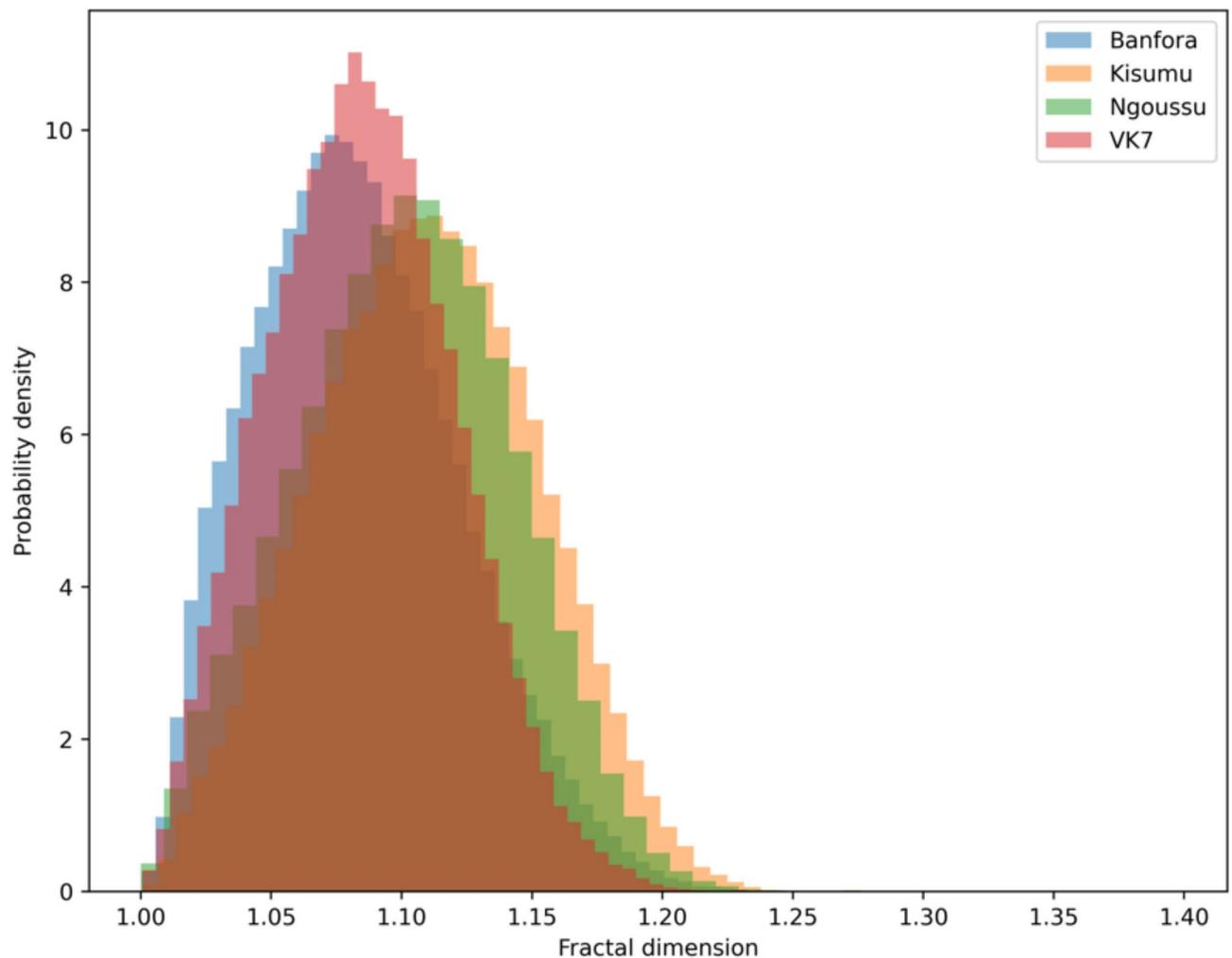


Fig. 8. Probability density histogram of fractal dimension of trajectory segments, based on the XGBoost segment length (8 s), with each strain highlighted separately. The IR strains (Banfora in blue and VK7 in red) demonstrate lower fractal dimension values than the IS strains (Kisumu in orange and N'goussu in green). Interestingly, despite containing different strains, the inter-resistance classes show similarities in fractal dimension values. IR strains show more directed flight in comparison to IS.

Model	Balanced accuracy	ROC AUC score
XGBoost	0.655 (0.616–0.691)	0.726 (0.688–0.768)

Table 4. Performance of the XGBoost model applied to independent test data when classifying between the IS mosquito strains (Kisumu and N'goussu), with the average and the range provided across all folds.

necessarily imply that IR mosquitoes are easier to predict or more vulnerable to attack. Instead, their slower speeds may simply reflect an alternative host-seeking strategy. For example, if the mosquito identifies blood meal cues such as CO₂ or heat, IR behaviours indicate that it is flying more slowly and hence can adjust its trajectory more easily to follow those cues. The skewness of the distributions indicate that IR tend to fly down (negative velocities) more slowly than flying up in comparison to IS. In other words, IR fly more slowly when host seeking downwards, which may give more opportunity to detect and follow bloodmeal cues. Previous work identified that mosquitoes can detect a surface when some 30 to 45 mm away due to the induced velocity and pressure changes around the mosquito³⁰ observations that are consistent with experimental findings around a human baited bednet¹⁵. This mechanism would lead to a reduction in velocity when flying downwards for all mosquitoes, however, the model here is comparing IR to IS strains. Hence, these results imply either an active response by the mosquito to the signals received when descending towards a surface or a more global behavioural shift or both. The skewed vertical velocity distribution also implies that IR fly upwards (positive velocity) quicker. In making these interpretations it is important to note that in these experiments the mosquitoes have not blood fed – hence

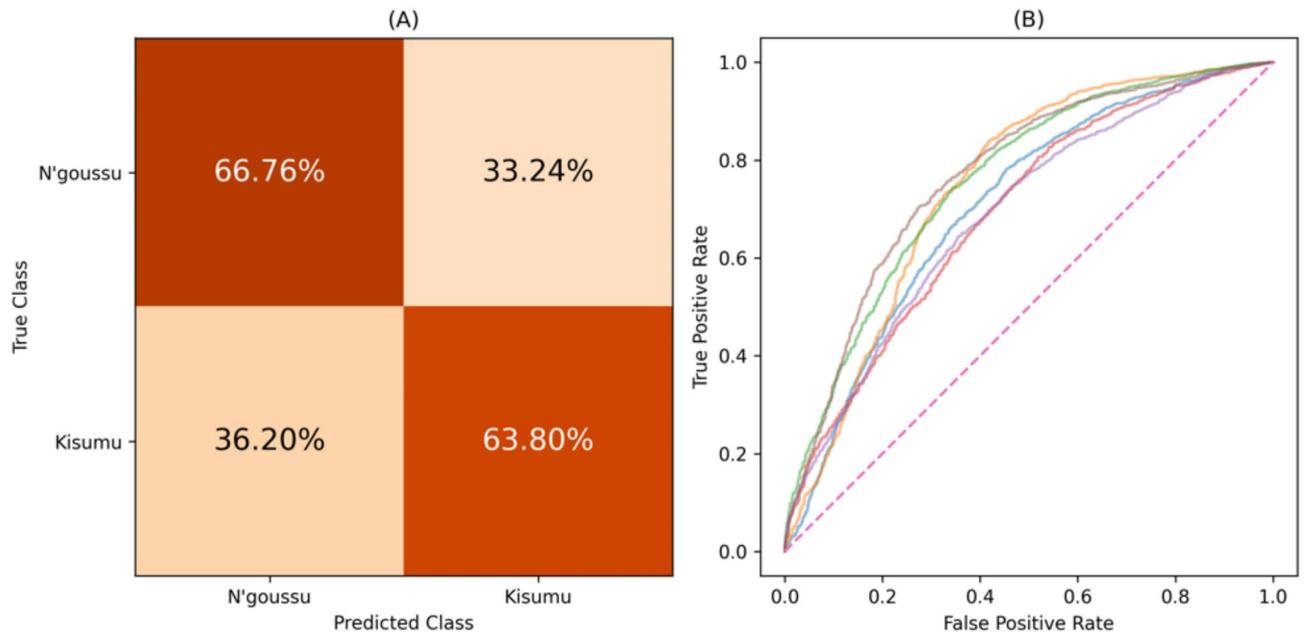


Fig. 9. (A) Confusion matrix displaying the percentages of the predictions over all folds for the IS mosquito strains (Kisumu vs. N'goussu). (B) Figure displaying each folds' ROC curve for the Kisumu vs. N'goussu mosquito strains. Both graphs show results from independent test data.

Model	Balanced accuracy	ROC AUC score
XGBoost	0.662 (0.606–0.725)	0.717 (0.633–0.809)

Table 5. Performance of the XGBoost model applied to independent test data when classifying between the IR mosquito strains (Banfora and VK7), with the average and the range provided across all folds.

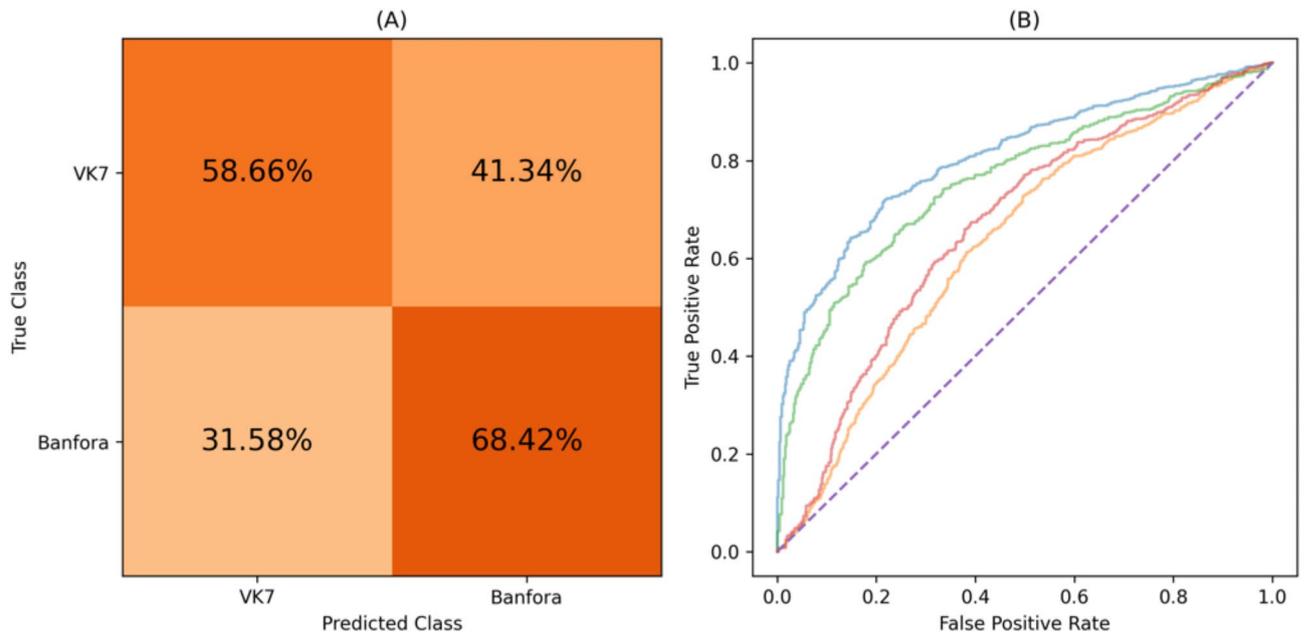


Fig. 10. (A) Confusion matrix displaying the percentages of the predictions over all folds for the IR mosquito strains (Banfora vs. VK7). (B) Figure displaying each folds' ROC curve for the Banfora vs. VK7 mosquito strains. Both graphs show results from independent test data.

Model	Balanced accuracy (training)	Balanced accuracy (testing)
XGBoost	0.928 (0.906–0.952)	0.496 (0.454–0.553)

Table 6. Performance metrics for the multiclass classification model applied to independent test data with the average and the range (minimum and maximum) performance across all folds.

Mosquito strain	Accuracy
Kisumu	0.635 (0.515–0.845)
Banfora	0.607 (0.422–0.813)
N’goussu	0.368 (0.264–0.471)
VK7	0.354 (0.240–0.460)

Table 7. Multiclass model accuracy one-vs-all of each strain applied to independent test data with the average accuracy and the range across all folds provided in brackets.

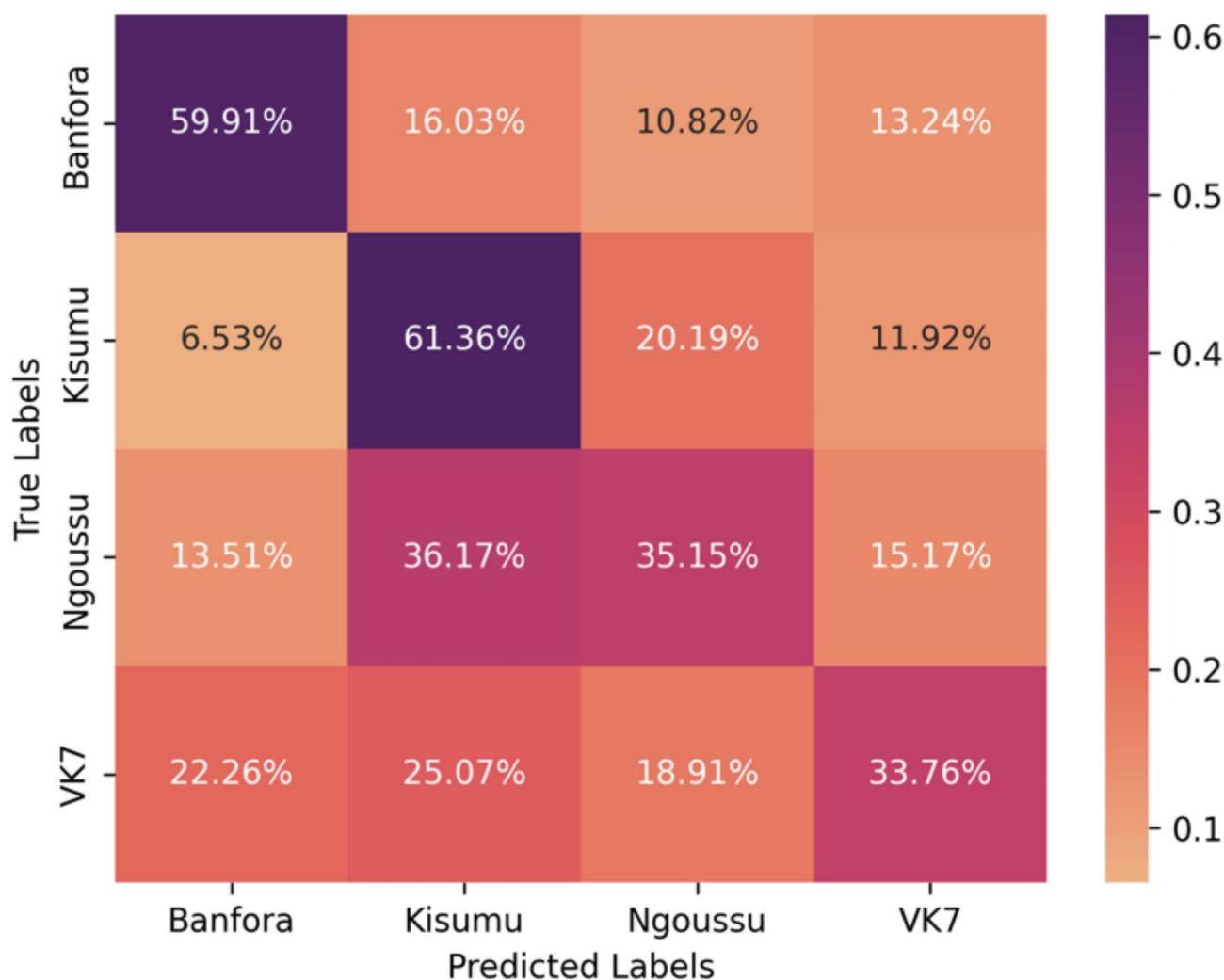


Fig. 11. Confusion matrix displaying the performance of the multiclass classifier applied to independent test data. The colour and the value of each square represents the percentages of the predictions over all folds.

these are innate host seeking characteristics not influenced by taking on a blood meal and the considerable increase in mass that involves.

Horizontal-velocity features are largely similar to those for the vertical-velocity component in the standard deviation (ranked 9), 1st quartile (ranked 15) and 3rd quartile (ranked 26). IR behaviours have a narrower

distribution of horizontal-velocities – implying that IR mosquitoes spend more time at slower speed than IS. The number of zero values and zero crossings (ranked 5) and number of local minima (ranked 8) introduce a further perspective. The number of zero crossings feature is more difficult to interpret as it includes the number of zero values of the horizontal-velocity within an 8 second segment. IR behaviours are associated through the model with low numbers of zero crossings or zero values, however, they have higher numbers of local minima than IS. This suggests that IR tracks are more consistent in direction compared to IS tracks. However, the higher number of local minima indicates a ‘jerky’ motion, characterised by frequent speeding up and slowing down. In the context of mosquito behaviour, a low amplitude ‘dither’ could enable the mosquito to sample CO₂ and thermal fields and maintain directed flight towards the highest concentration of these cues and hence a potential bloodmeal – see supplementary. While studies have demonstrated that mosquitoes utilise CO₂ and thermal cues for host-seeking³¹, the specific role of ‘dither’ motion in facilitating this process has not been explicitly investigated. This presents an intriguing avenue for future research to explore how such flight patterns influence mosquito behaviour.

These differences in velocity features may be dependent on structure and performance of musculature, as well as a myriad of other parameters that may be optimal for these features in IS but not in IR. This could mean that there is a cost to being resistant, which leads to slower speeds and the inability to compete. There is evidence^{32–34} which supports the notion that the mechanisms underpinning resistance may require additional energy, leading to a resource-based trade-off. This trade-off could explain the lower speeds and reduced variability in velocity features as observed in the SHAP analysis, Fig. 6. These findings are further supported by raw trajectory data as shown in the histogram of positional velocity features, Fig. 7. This lower variability in flight characteristics of IR strains could also be explained by the genetic bottleneck caused when selecting for resistance (e.g., fast blood-feeding). The resistant population may therefore be more genetically uniform showing less variation in features, e.g. vertical and horizontal velocity standard deviations, in comparison to IS mosquitoes. In the presence of ITNs, these characteristics become advantageous as they are selected for survival. However, in the absence of insecticide, these characteristics of IR could become a burden and hence would be lost.

Fractal dimension, ranked third in the model, measures the linearity or complexity of a trajectory. Values closer to 1 indicate more direct paths, while values approaching 2 signify complex paths with increased activity. SHAP analysis shows that IR has lower values than IS, approximately 1.04 for IR and 1.15 for IS. This trend aligns with many other features, consistently showing that IR tracks are more linear and direct compared to IS tracks. This potentially indicates that IR trajectories are more efficient or goal-oriented whilst IS tracks reflect more exploratory or less directed behaviour. Notably, the higher fractal dimensions in IS mosquitoes could also indicate a greater degree of baseline unpredictability or protean behaviour. This unpredictability might act as a defensive mechanism, making IS mosquitoes more difficult to attack successfully, as suggested by³⁵, who demonstrated that mosquitoes employ unpredictable escape maneuvers to evade threats. Figure 8 further illustrates that both IR strains, VK7 and Banfora, have distributions with similar lower values than the two IS strains, Kisumu and N’goussu. However, linearity does not imply smoothness, as IR trajectories can be both linear in their overall direction and oscillatory in their local patterns, reflecting fine-scale adjustments during flight.

Several features describe the shape of a trajectory. The centroid distance function (CDF) quantifies the deviations from the centroid, with a larger mean CDF indicating a wider or more variable track. SHAP analysis reveals that IR mosquitoes have a higher mean CDF (ranked 10), covering a larger area than IS, consistent with a larger convex hull (the 2D envelope of all positions in a segment) for IR. IR also has more local minima in CDF (ranked 22), suggesting less smooth and more oscillatory trajectories which may indicate bouncing behaviours. IS tends to stay closer to the centroid with fewer oscillations. The mean change in flight angle (ranked 13) is higher for IR, consistent with wider directional changes, while the 1st quartile (ranked 16) indicates a broader range for IR. IR has more zero-crossings in angle change (ranked 7), suggesting more consistent directional patterns supported by a higher number of local minima in CDF and higher tortuosity (ranked 37).

The analysis of angular velocity reveals that IR mosquitoes have a lower median angular velocity (ranked 23), suggesting smoother movements, but a higher mean angular velocity (ranked 25), indicating occasional sharp turns. This discrepancy suggests that while IR trajectories are generally smooth, they can still make sharp turns, likely due to their lower overall speed, which enhances manoeuvrability compared to IS. Curvature Space Space (CSS) shows higher median CSS values for IS trajectories, implying sharper turns with smaller radii, consistent with IS exhibiting more complex, curved trajectories. In contrast, IR tracks are more direct with fewer pronounced turns, though occasional sharp turns may still occur.

In terms of accelerations, the vertical-axis component features generally support the discussion regarding preferential IR behaviours in vertical-velocity. The vertical-acceleration median (ranked 19) shows IR behaviours with an upwards bias whereas IS is zero to slightly downwards. Furthermore, the 3rd quartile of the vertical-acceleration distribution (ranked 21) is at higher positive values for IR than IS – supporting the ability of IR mosquitoes to generate stronger upward accelerations for escape. The significant horizontal-axis acceleration features are the 3rd quartile (ranked 20) which shows IS having larger values than IR and standard deviation (ranked 32) which shows IS having a broader distribution of values than IR. The horizontal-axis features are less clearly associated with fitness characteristics, and it is notable that these features are further down the ranking than for the vertical-axis and hence have lower influence on classification.

IR vs. IS models were also determined using logistic regression and random forest classifiers, whilst the XGBoost classifier gave the best performance with an average balanced accuracy of 0.743 and ROC AUC of 0.813. The worst performing classifier was logistic regression with an average balanced accuracy of 0.723 and a ROC AUC score of 0.788. This difference demonstrates that there are complex non-linear relationships across features when attempting to separate the classes. However, the performance of even the simpler linear models suggests that certain features exhibit linear separability. All models displayed stable results across folds with the

range of performance metrics being consistently good (see Fig. 3 (B) and (C) for the ROC curves which exhibit the expected slight variation between folds). This highlights that there are no abnormal trials in the dataset; the experiments were conducted in environmentally controlled laboratories minimising the variability of external effects. The small variation across folds also demonstrates the reliability and potentially high generalisation capability of this pipeline for classification tasks of trajectories. The balanced accuracy of 0.743 is particularly impressive given the biological complexity and variability inherent in mosquito behaviour, suggesting the model effectively captures meaningful and subtle differences between IR and IS strains.

There are some limitations to this study that should be addressed. One such limitation is that the behavioural differences identified may not be applicable to all IR and IS strains as only 4 mosquito strains were considered. Nevertheless, the classification task over all 4 strains revealed a large in-class diversity, and there were much stronger differences between IR and IS that were captured. Further analysis of the SHAP data could be conducted to examine consistency in trends across the folds. Future work should explore behaviours when exposed to a variety of ITN insecticides and more genetically distinct IR and IS mosquitoes.

In summary, this paper demonstrates the ability of data-driven machine learning classifiers to distinguish behaviours of pairs of IR and IS mosquito strains. SHAPs analysis identified that IR mosquitoes exhibit more directed flight together with a low amplitude ‘dither’ enabling the mosquito to sample the concentration of attractants and maintain flight towards the highest concentration of these cues and hence a potential bloodmeal. IR strains also fly slower on average vertically and horizontally – meaning they have less inertia and can change direction more easily in response to cues or threats. This approach could potentially be used to assess mosquito susceptibility status, providing a complementary tool for insecticide resistance monitoring.

Data availability

The data presented in this study are available on request from LSTM. Please contact Professor Philip McCall (Philip.McCall@lstmed.ac.uk) for access. The code used in this paper has been deposited and made publicly available on the authors’ GitHub repository: <https://github.com/yasserqureshi/ir-vs-is>.

Received: 31 October 2024; Accepted: 18 February 2025

Published online: 25 February 2025

References

1. Vector-borne Diseases [Internet]. American Mosquito Control Association. <https://www.mosquito.org/vector-borne-diseases/> (Accessed 28 Mar 2024). (2024).
2. Mosquito-borne diseases. | World Mosquito Program. <https://www.worldmosquitoprogram.org/en/learn/mosquito-borne-diseases> (Accessed 28 Mar 2024). (2024).
3. World malaria report 2023. *Licence: CC BY- NC-SA 3.0 IGO* (World Health Organization, 2023).
4. Gattton, M. L. et al. The importance of mosquito behavioural adaptations to malaria control in Africa. *Evolution* **67** (4), 1218–1230 (2013).
5. Martinez-Torres, D. et al. Molecular characterization of pyrethroid knockdown resistance (*kdr*) in the major malaria vector *Anopheles gambiae* S.s. *Insect Mol. Biol.* **7** (2), 179–184 (1998).
6. Balabanidou, V. et al. Cytochrome P450 associated with insecticide resistance catalyzes cuticular hydrocarbon production in *Anopheles gambiae*. *Proc. Natl. Acad. Sci. U. S. A.* **113** (33), 9268–9273 (2016).
7. Edi, C. V. et al. CYP6 P450 Enzymes and ACE-1 Duplication Produce Extreme and Multiple Insecticide Resistance in the Malaria Mosquito *Anopheles gambiae*. *PLoS Genet.* **10**(3):e1004236. (2014).
8. Nzioki, I. et al. Differences in malaria vector biting behavior and changing vulnerability to malaria transmission in contrasting ecosystems of Western Kenya. *Parasit. Vectors.* **16** (1), 376 (2023).
9. Odero, J. I. et al. Early morning anopheline mosquito biting, a potential driver of malaria transmission in Busia County, Western Kenya. *Malar. J.* **23** (1), 66 (2024).
10. Sanou, A. et al. Insecticide resistance and behavioural adaptation as a response to long-lasting insecticidal net deployment in malaria vectors in the cascades region of Burkina Faso. *Sci. Rep.* **11** (1), 17569 (2021).
11. Sherrard-Smith, E. et al. Mosquito feeding behavior and how it influences residual malaria transmission across Africa. *Proc. Natl. Acad. Sci. U S A.* **116** (30), 15086–15095 (2019).
12. Ingham, V. A., Grigoraki, L. & Ranson, H. Pyrethroid resistance mechanisms in the major malaria vector species complex. *Entomol. Gen.* 515–526. (2023).
13. Williams, J. et al. Sympatric populations of the *Anopheles gambiae* complex in Southwest Burkina Faso evolve multiple diverse resistance mechanisms in response to intense selection pressure with pyrethroids. *Insects* **13** (3), 247 (2022).
14. Chandre, F. et al. Modifications of pyrethroid effects associated with *Kdr* mutation in *Anopheles gambiae*. *Med. Vet. Entomol.* **14** (1), 81–88 (2000).
15. Parker, J. E. A. et al. Infrared video tracking of *Anopheles gambiae* at insecticide-treated bed Nets reveals rapid decisive impact after brief localised net contact. *Sci. Rep.* **5** (1), 13392 (2015).
16. Gleave, K. et al. Impacts of dual active-ingredient bed Nets on the behavioural responses of pyrethroid resistant *Anopheles gambiae* determined by room-scale infrared video tracking. *Malar. J.* **22** (1), 132 (2023).
17. Ryo, M. et al. Explainable artificial intelligence enhances the ecological interpretability of black-box species distribution models. *Ecography* **44** (2), 199–205 (2021).
18. Williams, J. et al. Characterisation of *Anopheles* strains used for laboratory screening of new vector control products. *Parasit. Vectors.* **12** (1), 522 (2019).
19. Voloshin, V. et al. Diffuse retro-reflective imaging for improved video tracking of mosquitoes at human baited bednets. *R. Soc. Open. Sci.* **7** (5), 191951 (2020).
20. Lundberg, S. M. et al. A Unified Approach to Interpreting Model Predictions. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, editors. *Advances in Neural Information Processing Systems* [Internet]. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf (2017).
21. Qureshi, Y. M. et al. Finding a husband: using explainable AI to define male mosquito flight differences. *Biology* **12** (4), 496 (2023).
22. Angarita-Jaimes, N. C. et al. A novel video-tracking system to quantify the behaviour of nocturnal mosquitoes attacking human hosts in the field. *J. Royal Soc. Interface.* **13** (117), 20150974 (2016).
23. Qureshi, Y. M., Voloshin, V., Towers, C. E., Covington, J. A. & Towers, D. P. Double vision: 2D and 3D mosquito trajectories can be as valuable for behaviour analysis via machine learning. *Parasit. Vectors.* **17** (1), 282 (2024).
24. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods.* **17** (3), 261–272 (2020).

25. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12** (85), 2825–2830 (2011).
26. Chen, T., Guestrin, C. & XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco California USA: ACM. 785–94. <https://doi.org/10.1145/2939672.2939785> (Accessed 3 Jan 2025). (2016).
27. Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. SMOTE: synthetic minority Over-sampling technique. *Jair* **16**, 321–357 (2002).
28. Barreaux, P., Ranson, H., Foster, G. M. & McCall, P. J. Pyrethroid-treated bed Nets impair blood feeding performance in insecticide resistant mosquitoes. *Sci. Rep.* **13** (1), 10055 (2023).
29. Ingham, V. A. et al. Integration of whole genome sequencing and transcriptomics reveals a complex picture of the reestablishment of insecticide resistance in the major malaria vector *Anopheles coluzzii*. *PLoS Genet.* **17** (12), e1009970 (2021).
30. Nakata, T. et al. Aerodynamic imaging by mosquitoes inspires a surface detector for autonomous flying vehicles. *Science* **368** (6491), 634–637 (2020).
31. Coutinho-Abreu, I. V., Riffell, J. A. & Akbari, O. S. Human attractive cues and mosquito host-seeking behavior. *Trends Parasitol.* **38** (3), 246–264 (2022).
32. Suh, P. F. et al. Impact of insecticide resistance on malaria vector competence: a literature review. *Malar. J.* **22** (1), 19 (2023).
33. Rivero, A., Magaud, A., Nicot, A. & Vézilier, J. Energetic cost of insecticide resistance in *Culex pipiens* mosquitoes. *Jnl Med. Entom.* **48** (3), 694–700 (2011).
34. Garland, T. Trade-offs. *Curr. Biol.* **24** (2), R60–R61 (2014).
35. Cribellier, A., Camilo, L. H., Goyal, P. & Muijres, F. T. Mosquitoes escape looming threats by actively flying with the bow wave induced by the attacker. *Curr. Biol.* **34** (6), 1194–1205e7 (2024).

Author contributions

Conceptualization, Y.M.Q., V.V., C.E.T., J.A.C. and D.P.T.; methodology, Y.M.Q., V.V., C.E.T., J.A.C., and D.P.T.; software, Y.M.Q.; validation, Y.M.Q., V.V., C.E.T., J.A.C., D.P.T., K.G., H.R., and P.J.M.; formal analysis, Y.M.Q.; investigation, Y.M.Q.; resources, Y.M.Q.; data curation, Y.M.Q.; writing—original draft preparation, Y.M.Q.; writing—review and editing, V.V., C.E.T., J.A.C., D.P.T., K.G., H.R., and P.J.M.; visualization, Y.M.Q.; supervision, V.V., C.E.T., J.A.C. and D.P.T.; project administration, Y.M.Q., V.V., C.E.T., J.A.C. and D.P.T.; funding acquisition, Y.M.Q. All authors have read and agreed to the published version of the manuscript.

Funding

Y.M.Q. was supported by the EPSRC, grant number EP/T51794X/1. The mosquito tracking data was generated under research funded by the Bill & Melinda Gates Foundation [under Grant Agreement Nos. OPP1159078 and INV010445].

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-91191-w>.

Correspondence and requests for materials should be addressed to Y.M.Q.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025