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.

Image-based classification of abamectin and fipronil exposure in Heteracris littoralis (Orthoptera: Acrididae) embryos using compound eye histology

H. A. Ayad
Cairo University
K. M. Abdel Rahman
khabdelrahmn@sci.cu.edu.eg
Cairo University
M. A. Rushdi
Cairo University
M. H. Annaby
Cairo University

Article

Keywords: Ecotoxicology, insect development, histopathology, image classification, support vector machines

Posted Date: November 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11550796/v1

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

Additional Declarations: No competing interests reported.

Image-based classification of abamectin and fipronil exposure in *Heteracris littoralis* (Orthoptera: Acrididae) embryos using compound eye histology

Hasan A Ayad. Khaled M Abdel Rahman. Mohamed A Rushdi. Mahmoud H Annaby

Abstract

Pesticide exposure is usually assessed by chemical analysis of residues, which says little about biological effect. We asked whether the compound eye of grasshopper embryos carries image features that reflect exposure. Embryos of *Heteracris littoralis* were treated with sublethal concentrations (LC30) of abamectin or fipronil, and ten malformed embryos per treatment and ten untreated controls were sectioned, stained with hematoxylin and eosin and photographed. The compound eye was extracted by graph-cut segmentation, thresholded, sharpened and rotated to a common orientation. Because only 30 embryos were available, each color channel was augmented by bit flipping, giving 1,440 channel images. Histogram of oriented gradients features were used to train five classifiers to assign images to control, abamectin or fipronil groups, with training and test sets drawn from different embryos and the split repeated ten times. Linear support vector machine and one-nearest-neighbor classifiers made no errors, logistic regression and a neural network reached 99.88% and 99.97% test accuracy, and decision trees 97.10%. These results show that the three exposure groups are separable in this small dataset. Validation on independently sampled embryos is needed before any wider use.

Keywords: Ecotoxicology; insect development; histopathology; image classification; support vector machines

Introduction

Pesticide residues in soil are a serious threat to public health, environmental integrity, food security, and long-term ecological sustainability [1–4]. These residues can harm living organisms and disrupt both soil and aquatic ecosystems [5,6]. They may also contaminate the food supply, leading to biomagnification [7]. Additionally, pesticide residues can negatively impact essential soil microorganisms that support soil fertility and health [8,9]. Given these risks, monitoring and managing pesticide contamination is essential to protect human health, ensure sustainable agricultural practices, and protect ecological systems.

Conventional pesticide detection methods, including gas chromatography (GC), liquid chromatography (LC) notably high-performance liquid chromatography (HPLC) [10] and enzyme-linked immunosorbent assays (ELISA) [11], are the most dominant laboratory techniques. These techniques are widely adopted due to their high sensitivity, selectivity, and capacity for multi-residue detection across diverse matrices such as soil, water, and agricultural commodities [12]. Despite their analytical robustness, these methods are constrained by the need for sophisticated instrumentation, labor-intensive sample preparation, and highly skilled personnel [13], which limit their utility for real-time or on-site monitoring. To overcome these challenges, recent advances have introduced alternative spectroscopic and imaging-based techniques that enable rapid, non-destructive analysis. Notable examples include hyperspectral imaging (HSI) [14,15] and fluorescence spectroscopy [16–18], both of which enable rapid, high-throughput screening for pesticide residues. The integration of machine learning (ML) algorithms with these platforms has significantly improved their analytical performance by enabling automated pattern recognition [19]. For instance, in Fourier Transform Infrared (FTIR) spectroscopy, ML models such as support vector machines (SVM) and artificial

neural networks (ANN) are employed to discriminate spectral signatures corresponding to molecular vibrational modes associated with specific pesticide compounds. Similarly, in Near-Infrared (NIR) spectroscopy, ML facilitates the deconvolution of overlapping spectral features and enhances classification accuracy through multivariate calibration techniques like partial least squares regression (PLSR). In hyperspectral imaging, ML enables pixel-level analysis by extracting spatial-spectral correlations, supporting both qualitative classification and quantitative estimation of pesticide residues. Studies in the literature have also shown that combining different types of features during machine learning model training can lead to improved classification performance. For example, Abdelaty et al. [20] integrates spatial vertex-domain features, spectral features, and multilevel features derived from directed graphs, in conjunction with several variants of common spatial patterns.

In addition to advancing detection technologies, there is growing interest in understanding how pesticide exposure affects non-target organisms, especially during sensitive developmental stages. In Egypt, the use of pesticides such as abamectin and fipronil is prevalent in both public health initiatives and agricultural practices. Abamectin operates by interfering with the transmission of electrical signals in nerve and muscle cells [21], while fipronil functions by binding to gamma-aminobutyric acid receptors and glutamate-gated chloride channels, disrupting normal neurological activity [22]. Despite the importance of understanding chemical impacts, research on the effects of these chemicals on grasshopper embryonic development remains limited [23–25].

Grasshoppers deposit their egg-pods few cm deep within the soil, and the application of pesticides can lead to the accumulation of harmful residues that may adversely affect subterranean life [26]. Consequently, grasshopper embryos, which may linger in the soil for extended periods, could experience structural alterations in their sensory organs as a result of exposure to these chemical residues [23,24]. Understanding these implications is crucial for safeguarding not only pest management strategies but also the broader ecological networks that support life.

In the literature, many studies have been conducted to investigate the effects of toxic pesticide exposure on human health and the environment using machine learning models [27,28]. In this article, we present a machine learning-based workflow for detecting and identifying two specific pesticides (abamectin and fipronil) applied to grasshopper embryos. By analyzing slide images of embryonic compound eyes, we extract anatomical features using specialized image descriptors. These features serve as inputs for machine learning models, enabling accurate classification of whether an embryo has been exposed to pesticides, and if so, determining the specific type of pesticide.

The remainder of this paper is structured as follows. The Results section presents the experimental settings and classification results from various machine learning models, and the Discussion section summarizes the main findings. The Methods section explains the laboratory processes and conditions implemented to rear the grasshopper embryos and outlines the data collection process for the histological imagery. It then describes the preprocessing steps to isolate and prepare the compound eye region of interest as the target object in the detection process, specifies the dataset used in the classification process along with the augmentation procedure applied to histological slices of embryonic compound eyes, and describes the feature descriptor employed to encode the histological images for machine learning models.

Results

Classification accuracy of the five classifiers

Table 1 summarizes the average classification accuracy obtained from both cross-validation and testing phases, as well as the average number of misclassified instances for each classifier.

Overall, the five classifiers demonstrated high detection accuracy across all classes. Notably, both the linear SVM and K-nearest neighbors (KNN) classifiers consistently achieved perfect classification performance, with zero misclassifications across all iterations.

It is worthwhile to state that the performance of such decision trees, logistic regression, and neural networks classifiers might vary in terms of accuracy and cost with respect to the training and test dataset partitioning. However, this variation is much limited and as observed after repeating the classification experiment for 10 iterations, no variation is recorded as with the case of KNN and SVM. Additionally, the performance achieved by KNN and SVM outperforms all other classifiers, and the decision trees is generally achieved less than the remaining classifiers.

Confusion matrices and ROC curves

Figures 1–5 present confusion matrices from one experiment, showing actual versus predicted class labels for the three compound eye categories across the five classifiers listed in Table 1. In such figures, **Fi** denotes samples treated with fipronil, **Ab** represents samples treated with abamectin, and **Co** refers to untreated control samples.

Discussion

This study asked whether the compound eye of *H. littoralis* embryos, seen on routine histological slides, can separate embryos exposed to abamectin or fipronil from untreated controls and from each other. Using compound eye segmentation, HOG descriptors and five classical classifiers, with 10 embryos per class, the linear SVM and KNN reported no misclassifications, logistic regression and the neural network came close (test accuracy 99.88% and 99.97%), and decision trees were lowest (97.10%). These values show that the three groups are separable in this dataset. They do not yet show how the workflow would perform on new embryos from new batches.

The compound eye is a sensible target. In earlier work on this species, abamectin and fipronil were associated with disorganized brain and compound eyes, including ommatidia reduced in size and number and irregular in shape, and small optic lobes, but these changes were described for treated embryos as a group and not as signatures of one compound [23]. Abamectin acts on glutamate-gated chloride channels [21], and fipronil on GABA receptors and glutamate-gated chloride channels [22], with its action at the GABA-gated channel characterized in detail [42]. Mapping these mechanisms onto the two image classes is tempting, but our classifiers learned pixel gradient patterns and did not measure channel function, so any such link is a hypothesis. Sublethal exposure can affect arthropods in many ways [43]. The treated embryos analyzed here were also chosen for histology because they were malformed, and we did not measure ommatidial dimensions or compare the groups by blinded morphometry. The results therefore show that selected, affected embryos can be separated by image features, not how severe the damage is or whether any difference is statistically significant.

HOG encodes the local distribution of gradient orientations and tolerates moderate changes in illumination [36], and has been used from pedestrian to tumor detection [36–41]. We did not examine which parts of the 3,780-component vector drove the decisions. The classifier ranking is not surprising for a high-dimensional descriptor and only 15 training embryos per iteration, a setting in which sample size relative to feature dimension is a classic concern [44]. A 1-nearest-neighbor rule scores perfectly only if each test image lies closest to a training image of its own class, which is plausible when images also share staining, illumination and acquisition conditions. This comparison was not designed to rank classical classifiers against deep networks.

Perfect scores call for caution. Bit-flipping produced 1,440 channel images, but all derive from 30 embryos, so the biological sample size is 10 per class, not 480. Augmentation helps only if it creates plausible variation [34,35], and flipping high-order bits changes intensities in ways no stain or camera would. The decisive issue is the unit of partitioning. Five embryos per class were used for training and five different embryos for testing, so derived images of an embryo should stay on the same side of the split. If they fall in both training and test data, or in different cross-validation folds, a classifier can learn to recognize the specimen instead of the exposure. Splitting MRI slices at random rather than by subject inflated test accuracy markedly, even for randomly assigned labels [45], record-wise cross-validation greatly overestimated accuracy compared with subject-wise validation [46], leakage is a recurring cause of over-optimistic results [47], and hierarchical data should be validated in blocks [48]. The Methods describe an embryo-level split but not how the three cross-validation folds were built or nested with the final test, so the estimates hold only if all variants of each embryo stayed together at every step. The 10 iterations resample the same 30 embryos and are not independent replications. The error counts in Table 1 refer to derived test images, so their relation to accuracy and to the 720 test images should be made explicit, with a dispersion measure at the embryo level. Sharing per-embryo predictions, split indices and code would let readers check this [49].

The workflow differs in kind from most published pesticide analyses. Chromatography and immunoassays quantify the compound [10–12], and hyperspectral, fluorescence and near-infrared methods with machine learning read an optical signature of the residue [14–19], with open problems in model performance, data fusion and hardware [50]. Our approach does not measure residues. It classifies images of tissue from known exposure groups, and only that task was evaluated. Calling it detection of pesticide-associated effects needs further biological validation, and identification of residues in soil or crops was not tested. Acetylcholinesterase, a long-used marker of organophosphate and carbamate exposure, also responds to other pollutants [51], so biological indicators can lose specificity. An abnormal compound eye is likely less specific still, because species, developmental stage, dose and other stressors can alter embryonic morphology. It should not be treated as a chemical fingerprint, and it does not replace validated chemical methods.

In summary, HOG features and simple classifiers separated selected abamectin and fipronil exposed embryos from controls, which suggests that compound eye morphology merits study as an indicator of exposure. It does not yet show generalization, pesticide-specific damage, or detection of environmental residues.

Methods

Insect rearing, pesticide exposure and histology

The grasshopper *Heteracris littoralis* was reared in the Department of Entomology, Faculty Science, Cairo University in room temperature and ambient relative humidity. These hoppers were collected from Abu Rawash location: $(30.045,\ 31.094)$, near Cairo University, and reared for many years. They were kept in wooden cages of $30 \times 30 \times 50\ \mathrm{cm}^3$ that were provided with a Lamp (12W) and was lighted for 14 hrs. Insects were fed fresh clover (*Trifolium alexandrinum*) from October till June, then sesban (*Sesbania sesban*). *H. littoralis* is a grasshopper found all over the year round in Egypt and might have some injury to crops [29].

To get eggs, insects were offered cups filled with wet and sterilized sand for egg deposition. Eggs were removed every day morning and were put in a Petri dish with a wet cotton. They were considered 0-day old and were incubated at $27 \pm 1\ ^\circ\mathrm{C}$. For experiments, pesticides (abamectin (1.8%) and fipronil (20%)) were diluted to serial concentrations; then 20 eggs were used for each

concentration to calculate the median lethal concentration (LC_{50}). Eggs were put in a watch glass and received 0.5 ml of the tested pesticide for 2-3 s then were removed in a Petri dish with a clean and wet cotton and was incubated as above. Experiments were replicated three times. Control groups received water only. The development of the eggs was noticed daily for any changes in the eggs. Eggs that failed to have an embryo as those contained ones but did not hatch, were considered dead [23].

Abdel Rahman [23] used the sub-lethal dose (LC_{30}) in histological investigation ($n = 30$); many embryos then had mal-formations. These embryos were counted and were used in the histological examination: for each pesticide we used 10 embryos that suffered deformations. Usually about 30% of these embryos were about to hatch and were suffering from malformations. Of course, many embryos hatched ($\approx 50\%$), (for details see ref. [23]). To make the embryo visible, the egg shell (chorion) was removed using 3% Sodium Hypochlorite. Histological preparation were made by fixing eggs in hot Bouin's solution for 24 h, then they were washed several times in alcohol (70%), then dehydrated in ascending alcoholic series up to 100%, then they were left in ether overnight. Specimens were cleared in cedar wood oil (for 24 h) and were infiltrated in paraffin wax (for 1 h, three changes, melting point 60°). They were mounted, trimmed and were cut 8-10 μm using a microtome and were mounted on a glass slide. Specimens were stained with hematoxylin and eosin and were differentiated in acid-alcohol (1–2 drops of conc. HCl in 100ml 70% alcohol). Finally, Canada balsam was added on the stained specimens and covered with a coverslip. Slides were examined and then photographed with Galaxy M31 Mobile. Statistical analysis and the Probit analysis were carried out with IBM SPSS Statistics Version 22 (IBM Corp. Armonk, NY, USA) [23].

For complete examination of the specimens, groups of treated (LC_{30}) and control embryos ($n = 10$) were cut and processed as above. These specimens were examined and photographed. The chosen sections were about the median part of the compound eye where the optic nerve connects the eye. In these sections, the eye is in its biggest size with full number of ommatidia (the basic unit of the eye, basically formed of: lenses, cone cells, retinula cells that end with nerve endings and connect the brain). Figure 6 shows the Sagittal whole slide image through the compound eye of a fipronil-treated embryo.

In examination, it was difficult to get a precise transverse section through the mid head with two equal eyes as the specimens were slightly oblique; so we used the left one only.

Image preprocessing

In the preceding subsection, the process of imaging and collecting the whole slide images of the grasshopper embryos is described in great technical detail. Three different image sets are collected that pertain to three different groups: the control group refers to embryos not affected by any pesticides, the embryo group affected by abamectin, and the last group affected by fipronil. Figure 7 shows the complete sets of such slides for the three groups. As is evident, the number of slide images for each group is ($n = 10$), which is identical to the number of embryo samples chosen in the experiment described in that subsection.

The primary objective of this section is to systematically extract and enhance the embryonic compound eye region from the slide images, as it represents the area of interest. Each slide image predominantly features three significant regions, which correspond to three primary objects: the embryonic compound eye, the embryonic brain, and the background. While these main regions contain intricate organs as Fig. 6 describes, our focus will be on emphasizing the principal objects rather than the individual components they encompass.

To begin, the first step is to isolate the compound eye region. This area is essential, as it contains not only the vital organs associated with the compound eye but also various pigmented regions that are irrelevant to the compound eye tissue. Isolating this region precisely is critical to the subsequent processes.

Once we have successfully isolated the compound eye region, we turn our attention to segmenting the relevant tissues connected to the compound eyes. This segmentation process allows us to distinctly identify the various structures within the compound eye region, and cut off any other irrelevant tissue region or background objects.

Following the segmentation, we enhance the compound eye region to improve visibility and contrast among different tissue areas. This enhancement is vital, as it facilitates a clear distinction between various tissue types, enabling a more accurate manifestation of their features.

Finally, we perform a rotation of the compound eye region as an essential step for ensuring a consistent orientation angle. This transformation is adopted to normalize the positioning of the compound eyes being extracted. Figure 8 illustrates these processes when the slide image is taken as input and depicts the possible outcome after each process is performed.

In the following subsections, the processes of extraction, segmentation, thresholding, enhancement, and transformation are introduced in much further details to uncover the steps undertaken within each.

Extraction process

Figure 9a presents a section image effectively identifying the compound eye, brain, and background regions. To begin the extraction of the compound eye object region, we create a mask around the compound eye area. By outlining an oval selection around the left compound eye region, we successfully segment this compound eye object and accidentally a portion of the brain, as illustrated in the Fig. 9b.

Segmentation process

In this process of the compound eye's object segmentation, we divide the extracted compound eye into two distinct regions: the foreground (which includes the compound eye's object components manifesting the optic lobes and its constituent neuropils) and the background (comprising brain-connected cells and other background elements). This segmentation process utilizes the graph cuts methodology. This approach represents an image as a graph made up of nodes and edges, transforming the segmentation problem into one of graph partitioning. The goal is to create two subgraphs that correspond to the foreground and background regions.

Let $G = (V, E)$ be the simple, non-directed graph corresponding to the extracted compound eye shown in Fig. 10a, where V is the set of all nodes (pixels), and $E \subseteq V \times V$ is a set of ordered pairs of elements from V representing the edges connecting nodes. The graph G is characterized by a matrix W, whose element $w(i, j)$ specifies the weight of the connecting edge between nodes i and j. The problem is to partition V into subsets A (nodes of subgraph representing the foreground), and B (nodes of subgraph representing the background), such that $A \cup B = V$, and $A \cap B = \phi$. The weight function represents how the two nodes are close in proximity and similar in intensity. The following w function measures the weights on edges,

$$w(i,j) = \begin{cases} e^{-\frac{[I(n_i)-I(n_j)]^2}{\sigma_I^2}} e^{-\frac{dist(n_i,n_j)}{\sigma_d^2}} & \text{if } dist(n_i, n_j) \le r \\ 0 & \text{otherwise} \end{cases} \qquad (1)$$

where $I(n_i)$ is the n_i's intensity, σ_I is a constant measure of variation in intensity, σ_d is a constant measure of variation in distance, $\text{dist}(n_i, n_j)$ is a distance function between nodes, r is a radial constant that determines the extent to which we recognize and value similarity.

Based on the *min-cut theorem*, the partitioning between $\mathcal{A}$ and $\mathcal{B}$ is optimal when the cut severs edges with minimal total weights. Define any cut between $\mathcal{A}$ and $\mathcal{B}$ as the total weight of the edges connecting nodes in $\mathcal{A}$ to nodes in $\mathcal{B}$ that, if removed, would fully disconnect $\mathcal{A}$ from $\mathcal{B}$. This is formally expressed as:

$$\text{cut}(\mathcal{A}, \mathcal{B}) = \sum_{\mathcal{U} \in \mathcal{A},\, \mathcal{V} \in \mathcal{B}} w\,(\mathcal{U}, \mathcal{V}). \qquad (2)$$

To compute the minimal graph cut while accounting for the association of each partition to the overall graph, we adopt the normalized cut definition:

$$\text{Ncut}(\mathcal{A}, \mathcal{B}) = \frac{\text{cut}(\mathcal{A}, \mathcal{B})}{\text{cut}(\mathcal{A}, \mathcal{V})} + \frac{\text{cut}(\mathcal{A}, \mathcal{B})}{\text{cut}(\mathcal{B}, \mathcal{V})}, \qquad (3)$$

where $\mathcal{V}$ denotes the full set of graph vertices.

Finding a normalized cut is a crucial step in graph segmentation, as highlighted in ref. [30], where it is framed as a generalized eigenvalue problem. This method not only enhances efficiency but also effectively divides the graph into two distinct subregions (subgraphs) for the foreground and background nodes. Figure 10b shows the segmented compound eye object after separating its foreground region or cuttings its representative subgraph from the background region or its representative subgraph.

Thresholding process

After applying graph-cut segmentation to isolate the compound eye's object region, it becomes necessary to further segment only the compound eye's tissue pixels, excluding the pigmented areas within the compound eye's object. A thresholding operation proves essential for effective separation in image processing [31]. By closely examining the visual characteristics of the compound eye's object tissues alongside the surrounding pigment areas, we notice that the pigments exhibit significantly higher luminance levels compared to the compound eye tissues. This key insight suggests that converting the RGB color space into the Luminance-Chrominance (L, Cb, Cr) color space will enhance the thresholding effectiveness. By focusing on the luminance channel L, we can achieve remarkably improved results in distinguishing between these features. It is worthwhile mentioning that the Cr and Cb chrominance (color) components correspond to cyan-red and green-blue color oppositions, respectively [32,33].

Figure 11 illustrates four thresholded subfigures that effectively highlight the segmented region of the compound eye object. In Fig. 11a, we see that pixels with a luminance channel value above $L = 175$ are filtered out, while Fig. 11d eliminates pixels exceeding a luminance value of $L = 100$. The two middle subfigures feature maximum luminance limits of $L = 150$ (Fig. 11b) and $L = 125$ (Fig. 11c), respectively. Notably, as the maximum luminance limit decreases, we witness a progressive removal of pigment pixels, which leads to a corresponding reduction in compound eye tissues. This demonstrates the critical relationship between luminance thresholds and the effectiveness of tissue segmentation.

In our processing, we employed a luminance value ranging from $L = 100$ to $L = 175$ to perform thresholding on the extracted region of the compound eye. The threshold level is based on the darkness of the intensity values of the imaged embryonic tissues.

By establishing the threshold within this particular range, we effectively eliminate the majority of pigment pixels, which are typically brighter and less relevant to our main objective. At the same time, this choice allows us to retain most of the tissue pixels, which are essential for our study. Consequently, this balance ensures that we can accurately focus on the characteristics of the compound eye's structure while minimizing interference from unwanted pigmentation.

Enhancement process

Following the segmentation of the compound eye's object region, we finally isolate the region of interest, i.e., the embryonic compound eye tissue region. To further highlight the component features of such a region, we sharpen them via the unsharp masking filter. This filter first blurs the original image. Then, it subtracts the blurred image from the original image to extract the difference between the original and the blurred images, referred to by the mask. This mask is added to the original image to strengthen the edges and contours between region tissues.

Accordingly, let $\mathcal{C}$ denote the original image of the embryonic compound eye region, and let $\bar{\mathcal{C}}$ represent its blurred counterpart. The *sharpening mask*, $\mathcal{M}$, is computed as the difference between the original and blurred images:

$$\mathcal{M} = \mathcal{C} - \bar{\mathcal{C}}. \qquad (4)$$

The *sharpened image*, $\mathcal{S}$, is then obtained by enhancing the original image using the computed mask, scaled by a weighting factor α:

$$\mathcal{S} = \mathcal{C} + \alpha\mathcal{M}, \qquad (5)$$

where $\alpha \in \mathbb{R}^+$ controls the contribution of the sharpening mask to the final result.

Figure 12a shows the compound eye's tissue image to sharpen. The smoothed tissue image can be shown in Fig. 12b which is subtracted from Fig. 12a to output the boundaries of the tissue region in Fig. 12c. The sharpened image in Fig. 12d resulted after adding Fig. 12c to Fig. 12a.

Transformation process

This process focuses on normalizing the orientation of the segmented compound eyes within the dataset to ensure uniformity across all samples. To achieve this, we rotate the eye regions so that they align horizontally, which enhances symmetry in how the region of interest is positioned throughout the entire dataset.

As depicted in Fig. 13a, we begin by specifying an axis line that traverses the tissue region, connecting the extreme points of the eye tissue. This axis line serves as a reference for determining the rotation angle, which is measured with respect to the positive x-axis. Once we have calculated the appropriate rotation angle, we apply a transformation to adjust the orientation of the tissue region accordingly. This process is illustrated in Fig. 13b, where the rotated position of the compound eyes can be clearly observed, demonstrating enhanced symmetry and consistency across the dataset.

The transformation process marks the completion of the preprocessing stage, resulting in a finalized dataset of compound eye images. Figure 14 presents the images from all three classes after undergoing this process.

Dataset and augmentation

The preprocessed embryonic compound eyes consist our dataset we use in the classification task. Each color image has three channels that are used to form the initial construction of our dataset. Since each class contributes 10 embryonic compound eye samples, and each has three grayscale channels, then initially we have a total of 90 dataset grayscale images. As this number is too small to make the

classification problem work, we expand this initial dataset by augmenting each grayscale channel image to enlarge the dataset size and introduce some variations within each class of images.

We propose a robust image augmentation strategy designed to substantially expand our dataset and mitigate the challenges posed by limited image availability. This preprocessing step is essential for enhancing model performance, particularly in scenarios where data scarcity could hinder generalization. Previous studies [34,35] have demonstrated the effectiveness of various augmentation techniques such as rotation, flipping, scaling, and color adjustments in increasing dataset diversity and improving model robustness. These transformations enable the generation of multiple variants from a single image, thereby enriching the training set.

However, unlike conventional approaches that simulate numerous imaging conditions, our method assumes that the input slide image will undergo specific preprocessing prior to detection. Consequently, we eliminate the need to account for transformations like scaling, rotation, and flipping. Instead, we focus solely on bit flipping as a means of augmenting the dataset. This targeted approach simplifies the augmentation process while still enhancing data variability in a manner relevant to our application.

We detail the image augmentation technique implemented in this work, which is based on systematically flipping specific bits in the color channels of input images.

Let $I \in D$ denote an RGB compound eye image from the color embryonic image dataset D. Each image I comprises three grayscale channels, denoted as $\{I_R, I_G, I_B\}$, corresponding to the red, green, and blue components, respectively. Each channel I_C, where $C \in \{R, G, B\}$, is represented as an 8-bit grayscale image in which each pixel intensity is encoded using an 8-bit binary sequence. The intensity at a given pixel location (i, j) in channel I_C can be expressed as:

$$I_C(i,j) = (b_7 b_6 b_5 b_4 b_3 b_2 b_1 b_0), b_k \in \{0,1\}, k \in [0,7], \qquad (6)$$

where b_k is the k-th bit.

Given that $\neg$ represents the bitwise NOT (negation) operation, the following steps describe the construction of augmented channel images by selectively flipping individual or multiple bits in I_C:

1. **Single-bit flipping:** Construct 8 altered channel images $I_{C_0}, I_{C_1}, \ldots, I_{C_7}$, where the k-th bit b_k is flipped:

$$\begin{aligned} I_{C_0}(i,j) &= (b_7 b_6 b_5 b_4 b_3 b_2 b_1 \neg \boldsymbol{b_0}), \\ I_{C_1}(i,j) &= (b_7 b_6 b_5 b_4 b_3 b_2 \neg \boldsymbol{b_1} b_0), \\ &\vdots \\ I_{C_7}(i,j) &= (\neg \boldsymbol{b_7} b_6 b_5 b_4 b_3 b_2 b_1 b_0). \end{aligned} \qquad (7)$$

2. **Two-bit flipping:** Construct 4 altered channel images $I_{C_{10}}, I_{C_{32}}, \ldots, I_{C_{76}}$, where two adjacent bits are flipped:

$$\begin{aligned} I_{C_{10}}(i,j) &= (b_7 b_6 b_5 b_4 b_3 b_2 \neg \boldsymbol{b_1} \neg \boldsymbol{b_0}), \\ I_{C_{32}}(i,j) &= (b_7 b_6 b_5 b_4 \neg \boldsymbol{b_3} \neg \boldsymbol{b_2} b_1 b_0), \\ &\vdots \\ I_{C_{76}}(i,j) &= (\neg \boldsymbol{b_7} \neg \boldsymbol{b_6} b_5 b_4 b_3 b_2 b_1 b_0). \end{aligned} \qquad (8)$$

3. **Four-bit flipping:** Construct 2 altered channel images $I_{C_{3210}}, I_{C_{7654}}$, where four consecutive bits are flipped:

$$\begin{aligned} I_{C_{3210}}(i,j) &= (b_7 b_6 b_5 b_4 \neg \boldsymbol{b_3} \neg \boldsymbol{b_2} \neg \boldsymbol{b_1} \neg \boldsymbol{b_0}), \\ I_{C_{7654}}(i,j) &= (\neg \boldsymbol{b_7} \neg \boldsymbol{b_6} \neg \boldsymbol{b_5} \neg \boldsymbol{b_4} b_3 b_2 b_1 b_0). \end{aligned} \qquad (9)$$

4. **Full-bit flipping:** Construct 1 altered channel image $I_{C_{76543210}}$, where all 8 bits are flipped as $I_{C_{76543210}}(i,j) = (\neg \boldsymbol{b_7} \neg \boldsymbol{b_6} \neg \boldsymbol{b_5} \neg \boldsymbol{b_4} \neg \boldsymbol{b_3} \neg \boldsymbol{b_2} \neg \boldsymbol{b_1} \neg \boldsymbol{b_0})$.

It is evident that each step negates adjacent bits in groups of one, two, four, and eight bits, respectively. The rationale behind this operation is to simulate potential alterations that may arise during the imaging process. Minor alterations can occur due to the flipping of the least significant bits. While small changes may have minimal impact individually, altering multiple bits simultaneously, especially those with higher significance, can produce noticeable effects. Figure 15 presents the complete set of channels for one of the treated classes with abamectin, along with their flipped-bit counterparts. Variations between original image channels and their flipped-bit versions range from subtle, almost imperceptible differences to significant alterations.

The augmentation process followed the aforementioned steps, expanding each channel by a factor of 15. This means that each channel was augmented with its flipped-bit counterpart, along with an additional 15 relevant flipped-bit variations. The dataset consists of 10 embryo samples per class, with three classes in total. Since each color image comprises three component channels, the dataset initially contains 90 channels across all classes. Through augmentation, this yields 1,350 flipped bit channels. Including the original 90 channels, the final dataset comprises a total of 1,440 image channels, 480 channels per class. Figure 8 illustrates the preprocessing pipeline, where the compound eye slide image is first decomposed into its three grayscale color channels. Each channel is then subjected to augmentation by generating multiple bit-flipped variations.

Feature extraction

The Histogram of Oriented Gradients (HOG) feature descriptor was originally introduced for pedestrian detection in ref. [36]. Since then, it has been widely adopted across various domains, including face detection [37], lung classification and detection [38], rapid ship detection [39], marine oil spill identification [40], and breast cancer prediction [41]. The HOG feature is utilized in our context to detect and classify the treated and untreated (control) embryos. By detecting the HOG feature in the embryonic eyes of each class, a distinguishing decision can be made on which class this embryonic eye belongs. This feature is constructed based on computing the distribution of gradients over a spectrum of gradient directions in local spatial blocks. The behind idea introduced in ref. [36] is that the shape and appearance information can be captured via such a distribution rather than the exact location of each gradient.

The construction of a Histogram of Oriented Gradients (HOG) feature descriptor involves several steps. First, the input grayscale image is divided into small spatial regions called cells, which are arranged into larger overlapping blocks. Within each cell, the gradient magnitude and orientation are computed at every pixel. These gradients are then accumulated into a one-dimensional histogram based on orientation bins, weighted by their corresponding magnitudes. To account for variations in illumination and contrast, each block of cells is normalized using a suitable norm (e.g., L2-norm), which enhances robustness to lighting changes. The final HOG descriptor is obtained by concatenating all normalized block histograms across the image.

The Histogram of Oriented Gradients (HOG) descriptor is constructed as follows:

1. **Gradient computation:** For a grayscale image $I(x,y)$, compute the horizontal and vertical gradients:

$$\begin{aligned} G_x &= I(x+1,y) - I(x-1,y), \\ G_y &= I(x,y+1) - I(x,y-1) \end{aligned} \qquad (10)$$

Then compute the gradient magnitude $G(x, y)$ and orientation $\theta(x, y)$ as follows:

$$G(x, y) = \sqrt{{G_x}^2 + {G_y}^2}$$

$$\theta(x, y) = \tan^{-1}(G_y, G_x)$$

2. **Cell histograms:** Divide the image into cells of size $N \times N$. For each cell c, construct a histogram of gradient orientations with B bins:

$$h_b = \sum_{(x,y)\in c} \delta_b\left(\theta(x, y)\right) \cdot M(x, y) \qquad (11)$$

where $\delta_b(\cdot)$ is a binning function for orientation bin b.

3. **Block normalization:** Group adjacent cells into overlapping blocks (e.g., 2×2 cells). Let v be the concatenated cell histograms in a block. Normalize using:

$$v' = \frac{v}{\sqrt{\| v \|_2^2 + \epsilon^2}} \qquad (12)$$

4. **Descriptor formation:** The final HOG feature vector is formed by concatenating all normalized block vectors:

$$H = [v_1', v_2', \ldots, v_K'] \qquad (13)$$

The configuration parameters used to compute the HOG descriptor in our experiments are as follows: each cell consists of 8×8 pixels; each block contains 2×2 cells with a stride of one cell between adjacent blocks; the gradient orientations are quantized into 9 bins uniformly spaced over the range $0°$ to $180°$; and the L2-Hys normalization scheme is applied, which first performs L2-norm normalization of the block vector followed by clipping values above a threshold (typically 0.2) and renormalizing to unit length.

Classification experiments

We addressed a classification task involving three distinct classes of embryonic compound eyes using five machine learning classifiers: decision trees, linear SVM, logistic regression, K-nearest neighbors (KNN, $k = 1$), and neural networks. As described in "Dataset and augmentation" in the Methods, the dataset was evenly split into training and testing sets, each containing five unique embryonic samples per class. To ensure generalizability and avoid overfitting, test samples were drawn exclusively from embryos not present in the training set. Performance was evaluated over 10 independent iterations, each with a different permutation of sample assignments. Reported results represent the mean classification accuracy across these iterations.

During the training phase, five embryonic compound eyes are randomly selected from each class. Each selected sample contributes 48 grayscale images, derived from three original color channels and their respective bit-flipped counterparts, resulting in a total of 720 training images across all classes (5 samples * 48 images/sample * 3 colors/image). The remaining embryonic samples are reserved for testing and processed in the same manner, yielding an additional 720 images for evaluation.

To optimize computational efficiency, all images are resized to 64×128 pixels prior to feature extraction. The Histogram of Oriented Gradients (HOG) descriptor is then computed for each image in both the training and testing sets. This descriptor captures local gradient orientations and spatial distributions, effectively encoding shape and edge information. Each resulting HOG feature vector consists of 3,780 components, serving as classifier inputs.

To prevent overfitting and ensure robust model performance, the training data is further divided into three folds for cross-validation during the model training phase. After training, each of the five classifiers decision trees, linear SVM, logistic regression, KNN, and neural networks is evaluated on the testing set, where each image is represented by its corresponding HOG feature vector. This pipeline ensures a consistent and rigorous evaluation of classification accuracy across varying data partitions.

As described above, the training and testing procedures were repeated across 10 independent iterations. In each iteration, a different random selection of embryonic samples was used to form the training and testing sets, ensuring robust performance evaluation across varying data splits.

Acknowledgements

The authors are appreciated to Dr. Elnems, head of the EPMM for supplying the pesticides.

Author contributions

Clinical trial number: Not applicable.

Authors' contributions: H. A. Ayad: conceptualization, methodology, machine-learning modeling, formal analysis, writing — original draft. K. M. Abdel Rahman: conceptualization, experimental design, insect rearing, pesticide exposure experiments, data collection, writing — review and editing. M. A. Rushdi: image acquisition, image analysis, feature extraction, writing — review and editing. M. H. Annaby: supervision, methodology, validation, writing — review and editing. All authors read and approved of the final manuscript.

Funding information: No fund was offered

Data availability:

The authors declare that the data supporting the findings of this study are available within the paper. Should any raw data files be needed in another format they are available from the corresponding author upon reasonable request.

Additional information

Competing interests: The authors declare no competing interests.

Code availability: No codes

Use of AI-assisted tools

During the preparation of this work, the authors used AI tools (Rubriq for Microsoft Word and Paperpal) for language refinement. The final content was reviewed and edited by the authors, who take full responsibility for the final content.

References

1. Durham, W. F. Significance of pesticide residues to human health. *J. Dairy Sci.* **54**, 701–706 (1971). https://doi.org/10.3168/jds.S0022-0302(71)85911-8
2. Navarro, S., Vela, N. & Navarro, G. Review. An overview on the environmental behaviour of pesticide residues in soils. *Span. J. Agric. Res.* **5**, 357–375 (2007). https://doi.org/10.5424/sjar/2007053-5344
3. Sefiloglu, F. O., Tezel, U. & Balcıoğlu, I. A. Validation of an analytical workflow for the analysis of pesticide and emerging organic contaminant residues in paddy soil and rice. *J. Agric. Food Chem.* **69**, 3298–3306 (2021).

4. Kalia, A. & Gosal, S. K. Effect of pesticide application on soil microorganisms. *Arch. Agron. Soil Sci.* **57**, 569–596 (2011). https://doi.org/10.1080/03650341003787582

5. Aktar, W., Sengupta, D. & Chowdhury, A. Impact of pesticides use in agriculture: their benefits and hazards. *Interdiscip. Toxicol.* **2**, 1–12 (2009). https://doi.org/10.2478/v10102-009-0001-7

6. Zhang, W. Global pesticide use: profile, trend, cost and benefit. *CABI Agric. Biosci.* **1**, 1–11 (2018). https://doi.org/10.1186/s43170-018-0001-7

7. Sánchez-Bayo, F. & Wyckhuys, K. A. G. Worldwide decline of the entomofauna: a review of its drivers. *Biol. Conserv.* **232**, 8–27 (2019). https://doi.org/10.1016/j.biocon.2019.01.020

8. Köhler, H.-R. & Triebskorn, R. Wildlife ecotoxicology of pesticides: can we track effects to the population level and beyond? *Science* **341**, 759–765 (2013). https://doi.org/10.1126/science.1237591

9. Pathak, V. M. et al. Current status of pesticide effects on environment, human health and its eco-friendly management as bioremediation: a comprehensive review. *Front. Microbiol.* **13**, 962619 (2022). https://doi.org/10.3389/fmicb.2022.962619

10. Wang, X., Meng, X., Wu, Q., Wang, C. & Wang, Z. Solid phase extraction of carbamate pesticides with porous organic polymer as adsorbent followed by high performance liquid chromatography-diode array detection. *J. Chromatogr. A* **1600**, 9–16 (2019). https://doi.org/10.1016/j.chroma.2019.04.031

11. Yin, X. et al. A sensitive and specific enzyme-linked immunosorbent assay for the detection of pymetrozine in vegetable, cereal, and meat. *Food Chem.* **418**, 135949 (2023). https://doi.org/10.1016/j.foodchem.2023.135949

12. Rissato, S. R., Galhiane, M. S., de Almeida, M. V., Gerenutti, M. & Apon, B. M. Multiresidue determination of pesticides in honey samples by gas chromatography–mass spectrometry and application in environmental contamination. *Food Chem.* **101**, 1719–1726 (2007). https://doi.org/10.1016/j.foodchem.2005.10.034

13. Thorat, T., Patle, B. K., Wakchaure, M. & Parihar, L. Advancements in techniques used for identification of pesticide residue on crops. *J. Nat. Pestic. Res.* **4**, 100031 (2023). https://doi.org/10.1016/j.napere.2023.100031

14. Li, Q. et al. Rapid pesticide residues detection by portable filter-array hyperspectral imaging. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **330**, 125703 (2025). https://doi.org/10.1016/j.saa.2025.125703

15. Martínez Gila, D. M., Bonillo Martínez, D., Satorres Martínez, S., Cano Marchal, P. & Gámez García, J. Non-invasive detection of pesticide residues in freshly harvested olives using hyperspectral imaging technology. *Smart Agric. Technol.* **9**, 100644 (2024). https://doi.org/10.1016/j.atech.2024.100644

16. Vodova, M. et al. Detection of pesticides in food products using paper-based devices by UV-induced fluorescence spectroscopy combined with molecularly imprinted polymers. *Food Chem.* **380**, 132141 (2022). https://doi.org/10.1016/j.foodchem.2022.132141

17. Guo, Z. et al. Detection of pesticide in water using two-dimensional fluorescence correlation spectroscopy and n-way partial least squares. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **229**, 117981 (2020). https://doi.org/10.1016/j.saa.2019.117981

18. Mool-am-kha, P. et al. A fluorescence biosensor for organophosphorus pesticide detection with a portable fluorescence device-based smartphone. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **327**, 125330 (2025). https://doi.org/10.1016/j.saa.2024.125330

19. Lapcharoensuk, R., Fhaykamta, C., Anurak, W., Chadwut, W. & Sitorus, A. Nondestructive detection of pesticide residue (chlorpyrifos) on bok choi (*Brassica rapa* subsp. *chinensis*) using a portable NIR spectrometer coupled with a machine learning approach. *Foods* **12** (2023). https://doi.org/10.3390/foods12050955

20. Abdelaty, M. M., Rushdi, M. A., Rasmy, M. E. & Annaby, M. H. Graph vertex and spectral features for EEG-based motor imagery classification. *Comput. Biol. Med.* **189**, 109944 (2025). https://doi.org/10.1016/j.compbiomed.2025.109944

21. Wolstenholme, A. J. & Rogers, A. T. Glutamate-gated chloride channels and the mode of action of the avermectin/milbemycin anthelmintics. *Parasitology* **131**, S85–S95 (2005). https://doi.org/10.1017/S0031182005008218

22. Raymond-Delpech, V., Matsuda, K., Sattelle, B. M., Rauh, J. J. & Sattelle, D. B. Ion channels: molecular targets of neuroactive insecticides. *Invertebr. Neurosci.* **5**, 119–133 (2005). https://doi.org/10.1007/s10158-005-0004-9

23. Abdel Rahman, K. M. Effects of abamectin and fipronil insecticides on the brain and compound eyes of the embryo of *Heteracris littoralis* (Rambur) (Orthoptera: Acrididae). *Int. J. Trop. Insect Sci.* **43**, 1237–1241 (2023). https://doi.org/10.1007/s42690-023-00989-6

24. Abdel Rahman, K. M. & Mostafa, M. Embryonic development of the locust *Anacridium aegyptium* Linnaeus (Orthoptera: Acrididae) and toxicological effects of abamectin on its development. *J. Basic Appl. Zool.* **86**, 111 (2025). https://doi.org/10.1186/s41936-025-00515-9

25. Bode, K. et al. A locust embryo as predictive developmental neurotoxicity testing system for pioneer axon pathway formation. *Arch. Toxicol.* **94**, 4099–4113 (2020). https://doi.org/10.1007/s00204-020-02929-6

26. Badawy, M. I. Use and impact of pesticides in Egypt. *Int. J. Environ. Health Res.* **8**, 223–239 (1998). https://doi.org/10.1080/09603129873507

27. Pan, D. et al. Effects of neonicotinoid pesticide exposure in the first trimester on gestational diabetes mellitus based on interpretable machine learning. *Environ. Res.* **273**, 121168 (2025). https://doi.org/10.1016/j.envres.2025.121168

28. Huang, X., Wang, H., Wu, Z. & Lu, G. Classification and regression machine learning models for predicting mixed toxicity of carbamazepine and its transformation products. *Environ. Res.* **271**, 121089 (2025). https://doi.org/10.1016/j.envres.2025.121089

29. Ibrahim, M. M. Development and survival of the grasshopper *Heteracris littoralis* Rambur on a restricted diet (Orthoptera: Acrididae). *Z. Angew. Entomol.* **90**, 22–25 (1980). https://doi.org/10.1111/j.1439-0418.1980.tb03498.x

30. Shi, J. & Malik, J. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**, 888–905 (2000). https://doi.org/10.1109/34.868688

31. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **9**, 62–66 (1979). https://doi.org/10.1109/TSMC.1979.4310076

32. Busin, L., Vandenbroucke, N. & Macaire, L. Color spaces and image segmentation. *Adv. Imaging Electron Phys.* **151**, 65-168 (2008).

33. Poynton, C. *Digital Video and HD: Algorithms and Interfaces* (Elsevier, 2012).

34. Goceri, E. Medical image data augmentation: techniques, comparisons and interpretations. *Artif. Intell. Rev.* **56**, 12561–12605 (2023). https://doi.org/10.1007/s10462-023-10453-z

35. Shorten, C. & Khoshgoftaar, T. M. A survey on image data augmentation for deep learning. *J. Big Data* **6**, 60 (2019). https://doi.org/10.1186/s40537-019-0197-0

36. Dalal, N. & Triggs, B. Histograms of oriented gradients for human detection. in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* vol. 1 886–893 (IEEE, 2005). https://doi.org/10.1109/CVPR.2005.177

37. Zhu, Q., Yeh, M.-C., Cheng, K.-T. & Avidan, S. Fast human detection using a cascade of histograms of oriented gradients. in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)* vol. 2 1491–1498 (IEEE, 2006). https://doi.org/10.1109/CVPR.2006.119

38. Lian, Q.-L. et al. Identification of lung tumors in nude mice based on the LIBS with histogram of orientation gradients and support vector machine. *IEEE Access* **11**, 141915–141925 (2023). https://doi.org/10.1109/ACCESS.2023.3342105

39. Feng, J., Li, B., Tian, L. & Dong, C. Rapid ship detection method on movable platform based on discriminative multi-size gradient features and multi-branch support vector machine. *IEEE Trans. Intell. Transp. Syst.* **23**, 1357–1367 (2022). https://doi.org/10.1109/TITS.2020.3024919

40. Xu, J. et al. Oil slick identification in marine radar image using HOG, random forest, and PSO. *IEEE Geosci. Remote Sens. Lett.* **21**, 1–5 (2024). https://doi.org/10.1109/LGRS.2024.3431043

41. Youssef, D., Atef, H., Gamal, S., El-Azab, J. & Ismail, T. Early breast cancer prediction using thermal images and hybrid feature extraction-based system. *IEEE Access* **13**, 29327–29339 (2025). https://doi.org/10.1109/ACCESS.2025.3541051

42. Cole, L. M., Nicholson, R. A. & Casida, J. E. Action of phenylpyrazole insecticides at the GABA-gated chloride channel. *Pestic. Biochem. Physiol.* **46**, 47–54 (1993). https://doi.org/10.1006/pest.1993.1035

43. Desneux, N., Decourtye, A. & Delpuech, J.-M. The sublethal effects of pesticides on beneficial arthropods. *Annu. Rev. Entomol.* **52**, 81–106 (2007). https://doi.org/10.1146/annurev.ento.52.110405.091440

44. Raudys, S. J. & Jain, A. K. Small sample size effects in statistical pattern recognition: recommendations for practitioners. *IEEE Trans. Pattern Anal. Mach. Intell.* **13**, 252–264 (1991). https://doi.org/10.1109/34.75512

45. Yagis, E. et al. Effect of data leakage in brain MRI classification using 2D convolutional neural networks. *Sci. Rep.* **11**, 22544 (2021). https://doi.org/10.1038/s41598-021-01681-w

46. Saeb, S., Lonini, L., Jayaraman, A., Mohr, D. C. & Kording, K. P. The need to approximate the use-case in clinical machine learning. *GigaScience* **6**, gix019 (2017). https://doi.org/10.1093/gigascience/gix019

47. Kapoor, S. & Narayanan, A. Leakage and the reproducibility crisis in machine-learning-based science. *Patterns* **4**, 100804 (2023). https://doi.org/10.1016/j.patter.2023.100804

48. Roberts, D. R. et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **40**, 913–929 (2017). https://doi.org/10.1111/ecog.02881

49. Haibe-Kains, B. et al. Transparency and reproducibility in artificial intelligence. *Nature* **586**, E14–E16 (2020). https://doi.org/10.1038/s41586-020-2766-y

50. He, H. et al. Spectroscopic and imaging technologies combined with machine learning for intelligent perception of pesticide residues in fruits and vegetables. *Foods* **14**, 2679 (2025). https://doi.org/10.3390/foods14152679

51. Lionetto, M. G., Caricato, R., Calisi, A., Giordano, M. E. & Schettino, T. Acetylcholinesterase as a biomarker in environmental and occupational medicine: new insights and future perspectives. *Biomed. Res. Int.* **2013**, 321213 (2013). https://doi.org/10.1155/2013/321213

Authors and Affiliations:

Hasan A Ayad[1]. Khaled M Abdel Rahman[2]. Mohamed A Rushdi[3]. Mahmoud H Annaby[1].

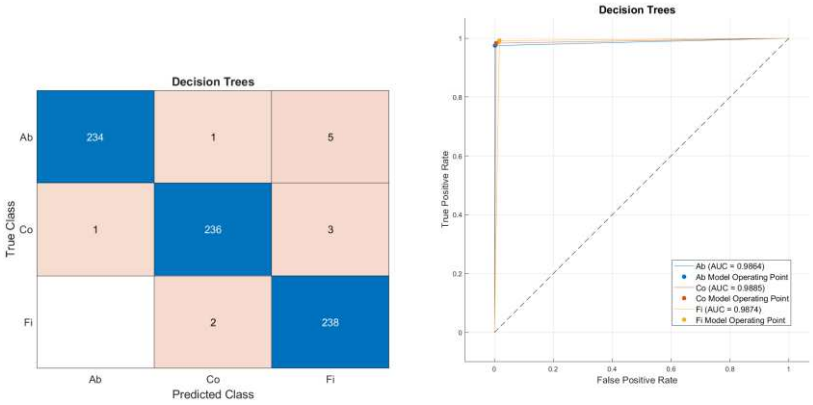
✉ K M Abdel Rahman khabdelrahmn@sci.cu.edu.eg

[1] Department of Mathematics, Faculty of Science, Cairo University, Giza 12613, Egypt

[2] Department of Entomology, Faculty of Science, Cairo University, Giza 12613, Egypt

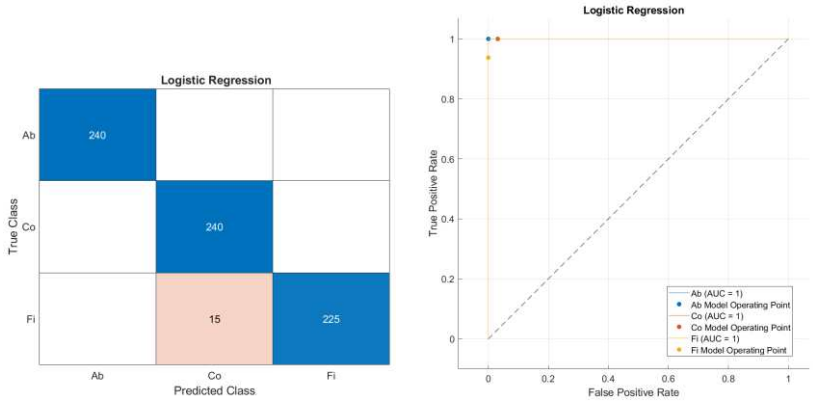
[3] Department of Biomedical Engineering and Systems, Faculty of Engineering, Cairo University, Giza 12613, Egypt

Figure legends



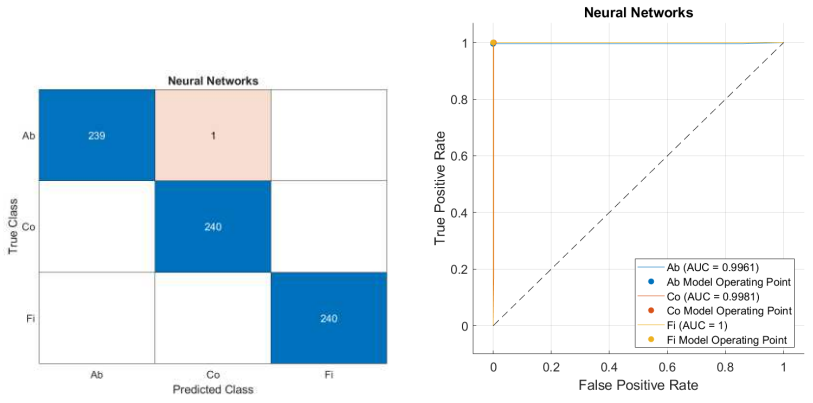
(a) **(b)**

Figure 1. Confusion matrix (a) and area under the receiver operating characteristic (ROC) curve (AUC) (b) illustrating the performance of the decision tree model on the test set, based on a random partitioning of the dataset into training and testing subsets. Ab, abamectin-treated; Co, untreated control; Fi, fipronil-treated.



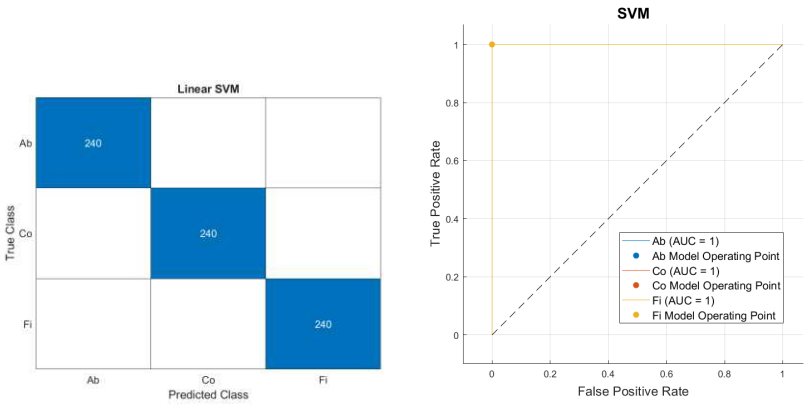
(a) **(b)**

Figure 2. Confusion matrix (a) and area under the receiver operating characteristic (ROC) curve (AUC) (b) illustrating the performance of the logistic regression model on the test set, based on a random partitioning of the dataset into training and testing subsets. Ab, abamectin-treated; Co, untreated control; Fi, fipronil-treated.



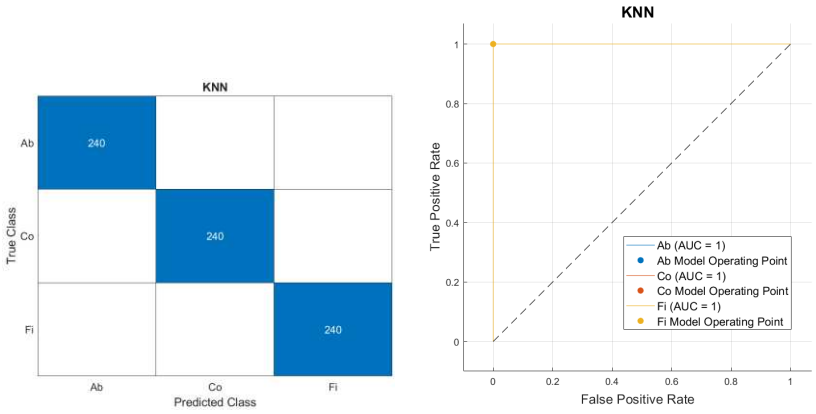
(a) **(b)**

Figure 3. Confusion matrix (a) and area under the receiver operating characteristic (ROC) curve (AUC) (b) illustrating the performance of the neural network model on the test set, based on a random partitioning of the dataset into training and testing subsets. Ab, abamectin-treated; Co, untreated control; Fi, fipronil-treated.



(a) **(b)**

Figure 4. Confusion matrix (a) and area under the receiver operating characteristic (ROC) curve (AUC) (b) illustrating the performance of the SVM model on the test set, based on a random partitioning of the dataset into training and testing subsets. Ab, abamectin-treated; Co, untreated control; Fi, fipronil-treated.



(a) **(b)**

Figure 5. Confusion matrix (a) and area under the receiver operating characteristic (ROC) curve (AUC) (b) illustrating the performance of the K-nearest neighbor model on the test set, based on a random partitioning of the dataset into training and testing subsets. Ab, abamectin-treated; Co, untreated control; Fi, fipronil-treated.

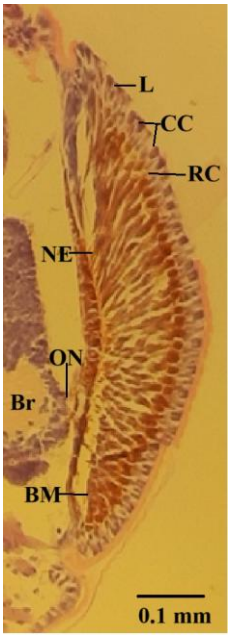


Figure 6. Transverse section in a normal well-developed compound eye to show the normal histological structure. Basement membrane (BM); brain (Br); crystalline cone cells (CC); lenses (L); nerve endings (NE); retinula cells (RC); optic nerve (ON).



(a) **(b)** **(c)**

Figure 7. Whole slide images of grasshopper embryo heads. Images were captured at the median section of the compound eyes to ensure consistency across samples. (a) Control group: embryos not exposed to any pesticide treatment. (b) Abamectin-treated group. (c) Fipronil-treated group.

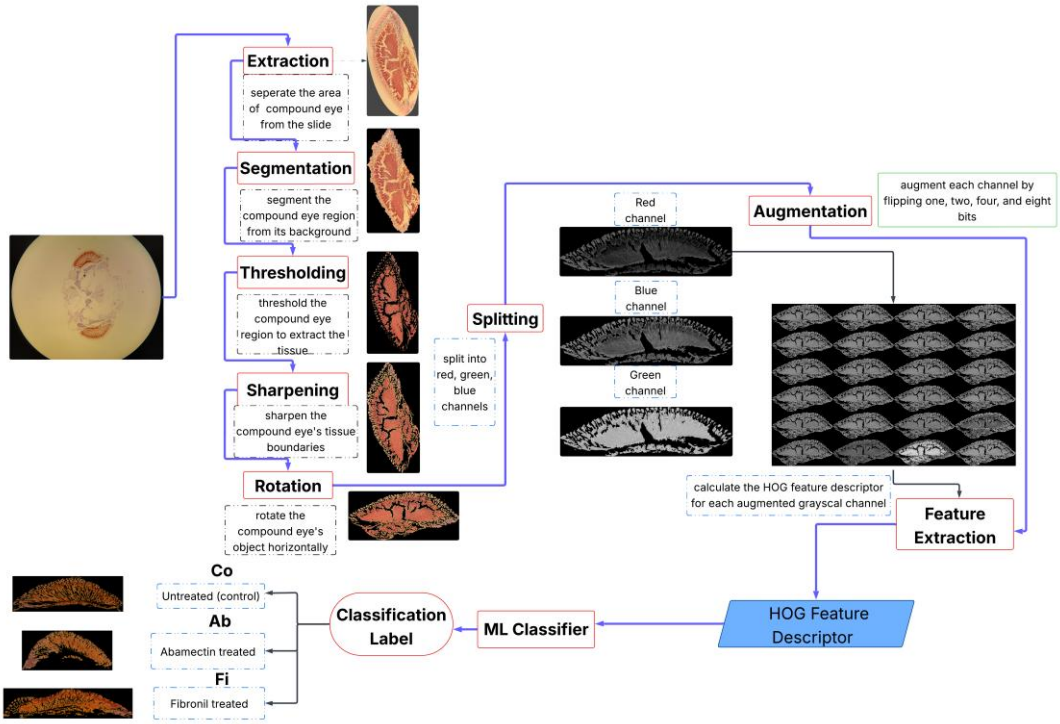
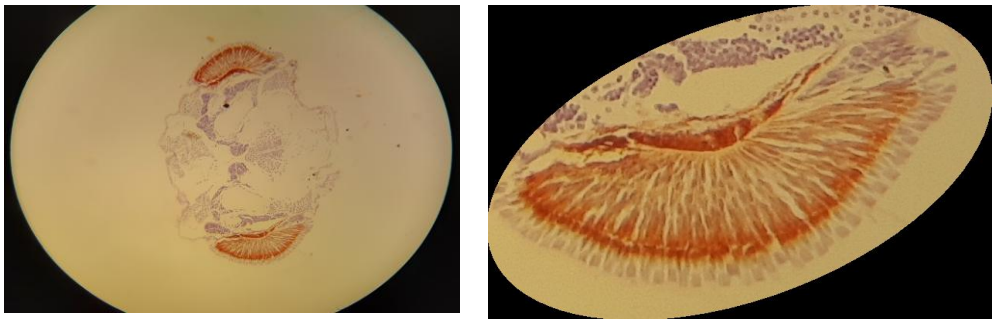
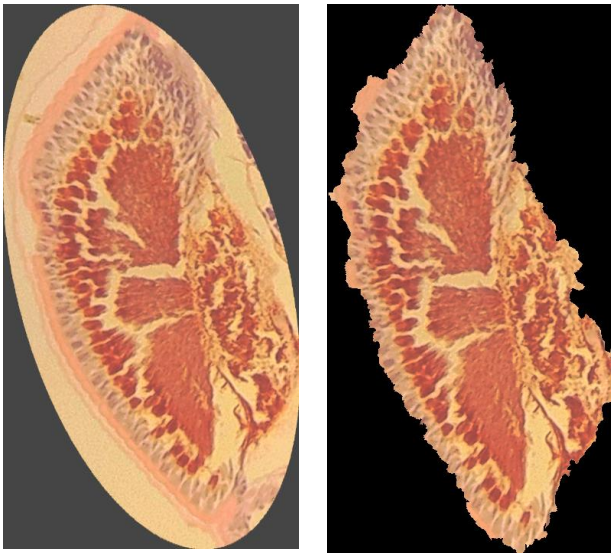


Figure 8. Workflow of the proposed system for grasshopper embryo classification from histological slides, with three classes: untreated, abamectin treated, and fipronil treated.



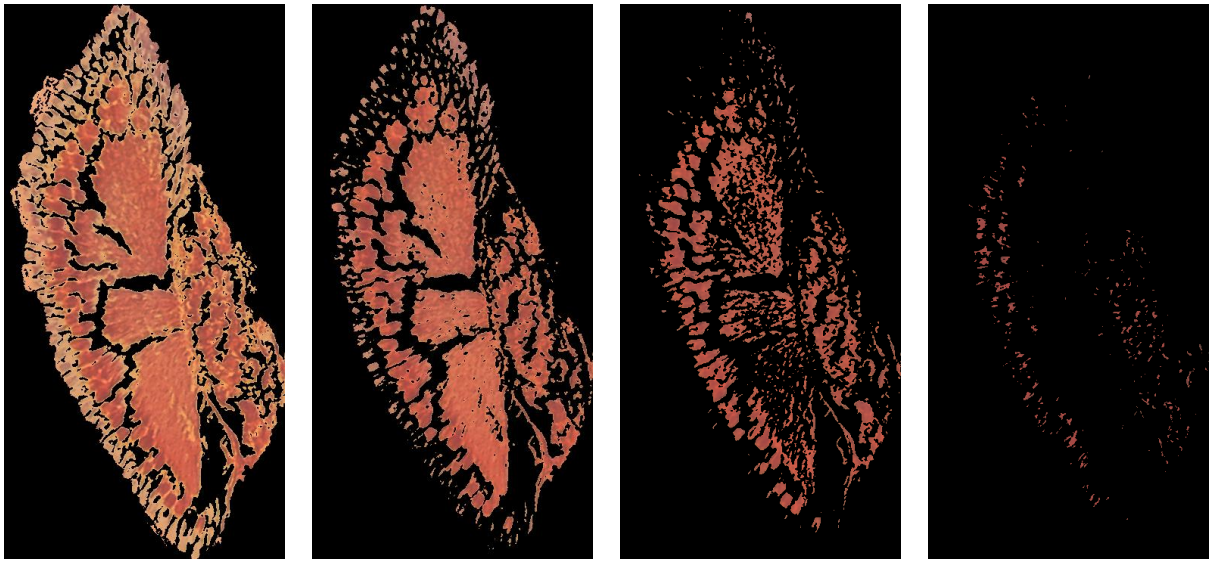
(a) **(b)**

Figure 9. Extraction of the left compound-eye by segmenting an oval area around the eye's region. (a) A slide head image with compound eyes, brain, and background parts. (b) The extracted left compound eye area from (a) with oval masking.



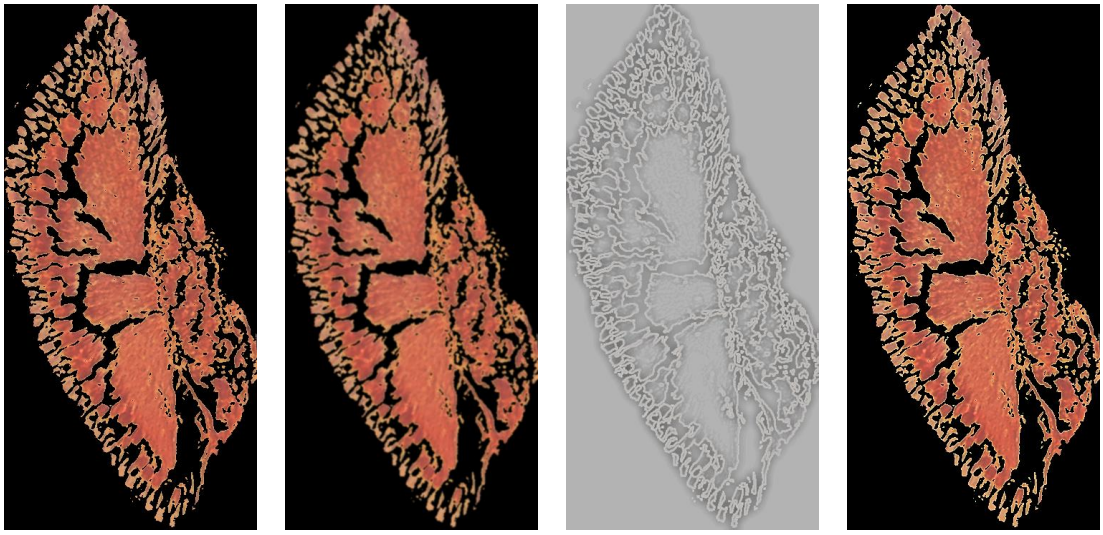
(a) (b)

Figure 10. The segmentation of the region including only the compound-eye parts from the area of the compound-eye. (a) The extracted area of the left compound-eye. (b) The segmented compound-eye region.



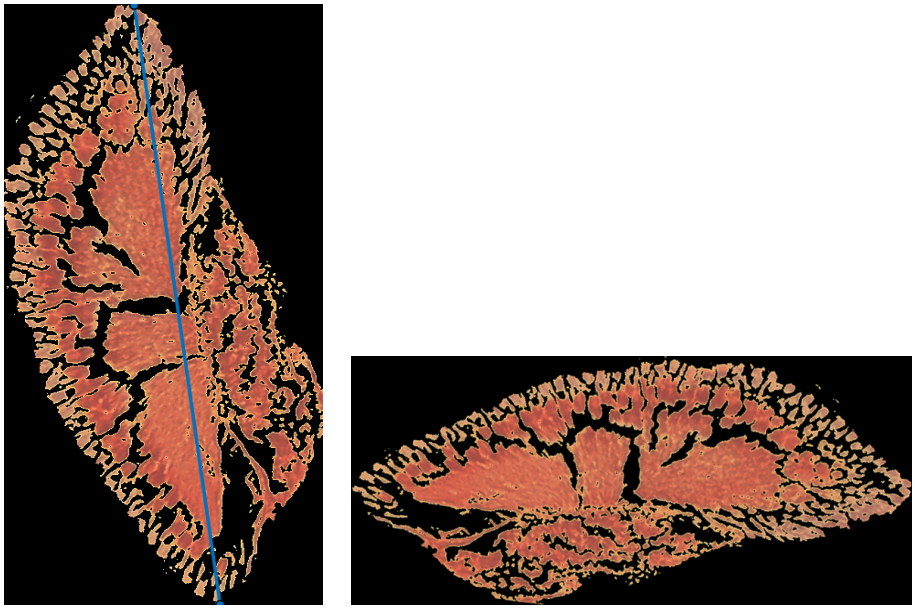
(a) (b) (c) (d)

Figure 11. Thresholded compound-eye region with various luminance threshold levels. (a) Threshold level = 175. (b) Threshold level = 150. (c) Threshold level = 125. (d) Threshold level = 100.



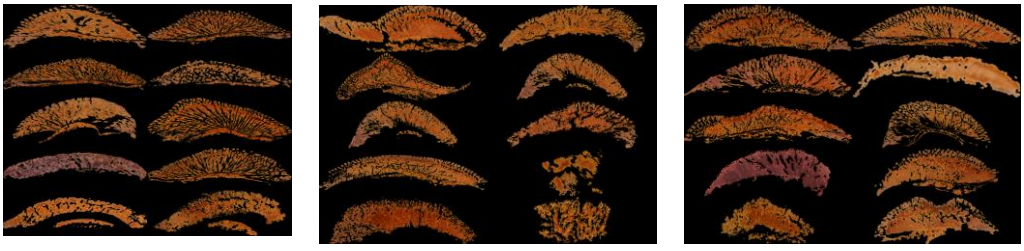
(a) (b) (c) (d)

Figure 12. Enhancing the compound-eye through sharpening its tissues via the unsharp mask, which extracts the contours of the eye, then adds these contours after scaling to the original eye. (a) Original compound-eye. (b) Blurred compound-eye. (c) Contours of compound-eye. (d) Sharpened compound-eye.



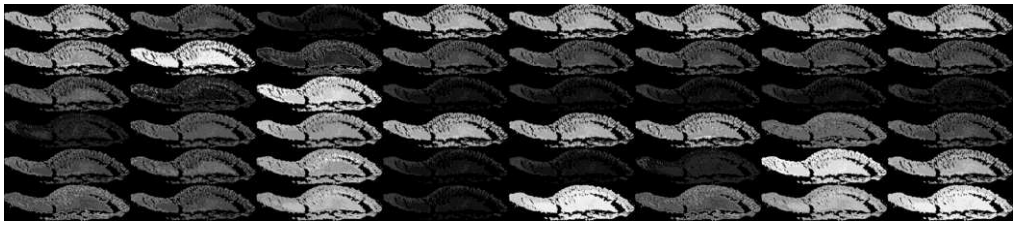
(a) (b)

Figure 13. The rotation transform on compound-eyes makes the orientation of the eye parallel to the x-axis. (a) A compound-eye sample before rotation. (b) The compound-eye after applying the rotation transform.



(a) (b) (c)

Figure 14. Preprocessed compound eye images representing the three experimental classes: (a) Untreated (Control), (b) Abamectin-treated, and (c) Fipronil-treated.

664 

665 **Figure 15.** Augmented image set corresponding to an abamectin-treated sample compound eye.

Tables

Table 1. Comparison of classification performance across five machine learning models.

Classifier	Cross-Validation Accuracy (%)	Error	Test Set Accuracy (%)	Error
Decision Trees	96.10	28.10	97.10	20.90
SVM	100.00	0.00	100.00	0.00
Logistic Regression	96.89	22.40	99.88	0.90
K-Nearest Neighbors (KNN)	100.00	0.00	100.00	0.00
Neural Networks	99.92	0.20	99.97	0.20

Accuracy and error are reported as averages over 10 iterations for both cross-validation and test sets. Error is the average number of misclassified instances.