International Journal Software Engineering and Computer Science (IJSECS)
5 (3), 2025, 907-917
Published Online December 2025 in IJSECS (http://www.journal.lembagakita.org/index.php/ijsecs)
P-ISSN: 2776-4869, E-ISSN: 2776-3242. DOI: https://doi.org/10.35870/ijsecs.v5i3.5120.

RESEARCH ARTICLE **Open Access**

Classification Optimization of *Aedes albopictus* and *Culex quinquefasciatus* Mosquito Larvae Using Vision Transformer Method

Abdullah Al Faruq *
Informatics Engineering Study Program, Faculty of Computer Science, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, East Jakarta City, Special Capital Region of Jakarta, Indonesia.
Corresponding Email: alfaruqvrg@gmail.com.

Dadang Iskandar Mulyana
Informatics Engineering Study Program, Faculty of Computer Science, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, East Jakarta City, Special Capital Region of Jakarta, Indonesia.
Email: mahvin2012@gmail.com.

Sopan Adrianto
Informatics Engineering Study Program, Faculty of Computer Science, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, East Jakarta City, Special Capital Region of Jakarta, Indonesia.
Email: sopan@stikomcki.ac.id.

Received: July 22, 2025; Accepted: September 5, 2025; Published: December 1, 2025.

Abstract: Mosquito-transmitted diseases like Dengue Hemorrhagic Fever and Filariasis pose serious health threats throughout tropical regions, particularly in Indonesia. Quick and accurate identification of mosquito larvae plays a crucial role in disease prevention, especially for Aedes albopictus and Culex quinquefasciatus species that act as main disease carriers. Manual identification methods using microscopes or visual guides often struggle with time constraints, accuracy issues, and dependence on trained specialists. Our research focuses on improving the classification of Aedes albopictus and Culex quinquefasciatus mosquito larvae using Vision Transformer (ViT) technology, a deep learning method that has shown strong results in image recognition tasks. We applied the Vision Transformer model to classify mosquito larvae from microscopic field images. The study also tested how different factors impact model performance, such as image clarity, lighting conditions, and image resolution. Our findings show that using Vision Transformer in classification systems produced excellent results, achieving 98.00% accuracy in recall, precision, and F1-score measurements. The research reveals that Vision Transformer methods deliver better accuracy than traditional approaches like Convolutional Neural Networks and can be adapted into working systems for technology and healthcare sectors.

Keywords: Vision Transformer (ViT); Mosquito Larvae Classification; Aedes Albopictus; Culex Quinquefasciatus; Deep Learning.

1. Introduction

Mosquito-borne diseases such as Dengue Hemorrhagic Fever (DHF) and Filariasis remain serious public health challenges in tropical regions, including Indonesia. These two diseases are caused by different mosquito species: *Aedes albopictus* serves as a secondary vector for DHF, while *Culex quinquefasciatus* acts as the primary vector for Filariasis. Both species reproduce through larval stages that develop in aquatic environments, whether clean or polluted water sources. According to Crans (2004), mosquito life cycles involve several critical stages, where the larval phase represents a decisive period that determines adult mosquito populations [1]. Previous research by Akbar & Mulyana (2022) revealed that earlier classification methods using CNN were suboptimal due to inadequate optimization techniques and data enhancement, which affected species recognition accuracy [2]. Therefore, early detection and classification of mosquito larvae becomes a crucial step in controlling mosquito populations and preventing disease transmission.

However, conventional field identification methods still rely heavily on manual observation using microscopes by trained personnel. Earlier studies by De Silva & Jayalal (2020) demonstrated that manual classification by health workers is highly observer-dependent, time-consuming, and can lead to incorrect decision-making and inefficient identification [3]. The process requires time, specialized expertise, and often faces accuracy challenges due to morphological similarities between species during the larval stage. Additionally, the shortage of trained health personnel in endemic areas frequently results in suboptimal monitoring. For instance, in early 2023, DHF cases surged in several regions of Central Java and East Kalimantan, largely attributed to delayed early detection of *Aedes* larval populations, which contributed to increased infection rates in communities. Reinert (2000) noted that approximately 2,500 mosquito species exist worldwide, many of which pose disease risks to humans, including DHF and Filariasis [4].

The complexity of mosquito larvae identification increases due to extremely subtle morphological variations between species, particularly during early instar larval stages. Sesulihatien *et al.* (2020) explained that kinematic characteristics can differentiate between *Aedes* and *Culex* larvae, yet manual observation of movement patterns requires specialized skills and considerable time [5]. Furthermore, Azman & Sarlan (2020) identified that *Aedes* larvae detection and classification systems using deep learning show promising potential but still require further development to achieve optimal accuracy [6]. Another challenge involves the shortage of trained human resources in remote areas, where mosquito-borne disease cases are often higher.

With advances in artificial intelligence technology, Computer Vision and Deep Learning approaches have begun to be utilized for automating biological object identification processes, including mosquito larvae. Saeed *et al.* (2021) demonstrated the use of Convolutional Neural Networks (CNN) for mosquito larvae detection with encouraging results [7]. However, conventional CNN architectures still have limitations in capturing long-range dependencies and global patterns in images. One cutting-edge architecture showing superior performance in image classification is the Vision Transformer (ViT). Unlike traditional convolutional models such as CNN, ViT processes images as small patches and applies self-attention mechanisms to understand visual patterns holistically. Vision Transformers have proven effective in various medical and biological applications. Krishnan & Krishnan (2021) showed that ViT can achieve excellent performance in COVID-19 detection using chest X-ray images [8]. Meanwhile, Li *et al.* (2021) developed a medical diagnosis platform based on Vision Transformer for coronavirus detection, demonstrating high accuracy [9]. ViT capability to recognize spatial structures and textures makes it highly suitable for classifying microscopic objects like mosquito larvae. Wang *et al.* (2022) developed PVT v2 (Pyramid Vision Transformer), showing significant improvements in Vision Transformer baselines [10].

Recent research by Surya *et al.* (2022) conducted evaluations of various deep learning architectures for mosquito larvae classification, concluding that transformer-based approaches show tremendous potential [11]. Adhane *et al.* (2021) also employed deep convolutional neural networks for *Aedes albopictus* mosquito classification with satisfactory results [12]. However, specific application of Vision Transformers for mosquito larvae classification remains scarce and requires deeper investigation. The current study aims to develop an automated classification system to distinguish between *Aedes albopictus* and *Culex quinquefasciatus* mosquito larvae using Vision Transformer methodology. Through implementing such technology, the larvae identification process is expected to become faster, more accurate, and independent of field expert personnel. The system not only enhances efficiency in disease vector monitoring but also has potential as a long-term solution in modern, sustainable technology-based communicable disease control strategies in Indonesia.

Previous research by Liu *et al.* (2021) stated that Swin Transformer has key differences from classical Vision Transformer (ViT), particularly in using shifted windows for self-attention and hierarchical structures [13]. Nevertheless, the scarcity of research using these methods for biological object classification applications, especially mosquito larvae, opens significant opportunities for further investigation. Njaime *et al.* (2024) recently developed automated classification systems for mixed populations of *Aedes aegypti* and *Culex quinquefasciatus* under field conditions, indicating that such technology is becoming mature for practical implementation [14]. The primary goal of the present research involves developing a mosquito larvae classification system that leverages Vision Transformer advantages in capturing global patterns and long-range dependencies in microscopic images. The study is expected to provide reliable technological solutions to support disease vector control programs in Indonesia while reducing dependence on manual expert personnel who are in short supply.

2. Related Work

2.1 Mosquito Classification and Detection Systems

Automated mosquito identification has gained considerable attention from researchers worldwide, with various methodologies being developed to tackle species classification challenges. Mohommed *et al.* (2020) created DenGue CarB, a machine learning-based system for mosquito identification and classification [19]. Their work proved that automated approaches could successfully distinguish between different mosquito species under controlled laboratory settings. Nevertheless, the system struggled with performance consistency when faced with varying environmental conditions and diverse image qualities. Taking a different approach, González-Pérez *et al.* (2021) developed the Mosquito Sound-Based Identification System (MSBIS) to identify *Aedes aegypti*, *Culex quinquefasciatus*, and *Anopheles stephensi* through smartphone audio recordings [20]. The method utilized distinct wing-beat frequencies characteristic of each species. While showing promise for field deployment, the audio-based technique required quiet environments and faced challenges in naturally noisy habitats. Deep learning has transformed mosquito larvae classification significantly. Azman & Sarlan (2020) built the Aedes Larvae Classification and Detection (ALCD) system using advanced neural network architectures, showing marked improvements over conventional computer vision methods [6]. Their research revealed how deep networks excel at extracting features needed to distinguish subtle morphological differences between larvae species. Saeed *et al.* (2021) applied Convolutional Neural Networks for mosquito larvae detection, producing encouraging results in automated identification [7]. More recently, Njaime *et al.* (2024) created automated classification systems capable of handling mixed populations of *Aedes aegypti* and *Culex quinquefasciatus* in real field conditions [14]. Their research tackled practical challenges by testing system robustness across natural environments with fluctuating lighting and background interference. The study proved that deep learning methods could maintain high accuracy even when processing complex field scenarios.

2.2 Vision Transformer Applications in Biological Classification

Vision Transformers have revolutionized image classification across multiple domains, particularly in biological applications. Krishnan & Krishnan (2021) pioneered Vision Transformer usage for COVID-19 detection through chest X-ray analysis, achieving superior results compared to traditional CNN architectures [8]. Their work established foundations for applying transformer-based models in medical imaging, demonstrating how self-attention mechanisms effectively capture global patterns in biological images. Li *et al.* (2021) advanced the field by creating a medical AI diagnosis platform based on Vision Transformer technology for coronavirus detection [9]. Their research emphasized hierarchical feature learning and global pattern recognition in medical image analysis. The platform achieved exceptional accuracy while maintaining computational efficiency, making clinical deployment feasible. Advanced Vision Transformer architectures continue evolving. Wang *et al.* (2022) introduced PVT v2 (Pyramid Vision Transformer), addressing several original Vision Transformer limitations through pyramidal structures and enhanced attention mechanisms [10]. The advancement proved particularly valuable for fine-grained classification tasks requiring accurate capture of subtle visual differences. Liu *et al.* (2021) made significant contributions with Swin Transformer development, introducing shifted windows for self-attention and hierarchical structures [13]. The architecture demonstrated superior performance across various computer vision tasks, especially scenarios requiring multi-scale feature representation. The hierarchical approach benefited biological classification tasks where objects exhibit complex morphological variations. Surya *et al.* (2022) conducted extensive evaluations of various deep learning architectures specifically for mosquito larvae classification, concluding that transformer-based approaches showed exceptional potential [11]. Their comparative analysis revealed that Vision Transformers could outperform traditional CNN architectures in capturing fine-grained morphological features essential for species differentiation.

2.3 Data Augmentation and Optimization Techniques

Data augmentation plays a crucial role in improving classification performance across diverse domains. Akbar & Mulyana (2022) demonstrated data augmentation effectiveness in optimizing classification accuracy for Betawi batik patterns using KNN and GLCM methods [15]. Their research showed that proper augmentation techniques could significantly improve model generalization, particularly when working with limited training datasets. In agricultural applications, Nana *et al.* (2022) investigated grape classification optimization using data augmentation combined with Convolutional Neural Networks [16]. Their work revealed that strategic augmentation techniques could enhance model robustness and reduce overfitting issues commonly encountered in image classification tasks. The study emphasized selecting appropriate augmentation strategies based on target object characteristics. Mulyana *et al.* (2022) examined pitcher plant flower classification optimization using Naïve Bayes algorithms combined with data augmentation and Particle Swarm Optimization (PSO) [17]. Their research demonstrated that hybrid approaches combining traditional machine learning with optimization techniques could achieve competitive performance while maintaining computational efficiency. The work showed potential benefits of combining multiple optimization strategies for improved classification outcomes. Statistical feature applications in classification tasks have been extensively studied. Ruswandi *et al.* (2022) developed an optimized classification system for avocado ripeness assessment using KNN and statistical features [18]. Their research demonstrated that carefully selected statistical features could provide robust representations for classification tasks, particularly when paired with appropriate machine learning algorithms.

2.4 Convolutional Neural Networks in Classification Tasks

Traditional deep learning approaches using Convolutional Neural Networks have been widely applied across various classification domains. Radikto *et al.* (2022) implemented CNN algorithms for vehicle classification on highways, demonstrating convolutional architecture versatility in real-world applications [21]. Their work showed that CNNs could effectively handle complex visual patterns and achieve high accuracy in dynamic environments. In medical applications, Utami *et al.* (2022) developed a human body dehydration classification system based on RGB images of urine color using K-Nearest Neighbor methods [22]. The research illustrated computer vision technique potential in healthcare applications, showing how color-based features could be effectively utilized for medical diagnosis purposes. The evolution from traditional CNN approaches to more advanced architectures like Vision Transformers represents a significant paradigm shift in computer vision. While CNNs excel at capturing local features through convolutional operations, Vision Transformers offer superior capabilities in understanding global patterns and long-range dependencies. The distinction becomes particularly relevant in biological classification tasks where both local morphological features and global structural patterns contribute to accurate species identification.

2.5 Research Gaps and Future Directions

Despite significant advances in automated mosquito classification, several research gaps persist. Most existing studies focus on adult mosquito identification rather than larval stage classification, which remains crucial for early intervention strategies. Additionally, state-of-the-art Vision Transformer architecture applications specifically for mosquito larvae classification remain underexplored. Integrating data augmentation techniques with Vision Transformers for mosquito larvae classification presents opportunities to address the common challenge of limited training data in biological applications. Furthermore, developing robust systems capable of handling field conditions with varying image qualities and environmental factors requires continued research efforts. The current study aims to address these gaps by developing a classification system that leverages Vision Transformer advantages while incorporating effective data augmentation strategies specifically tailored for mosquito larvae identification tasks.

3. Research Method

The research aims to optimize classification of *Aedes albopictus* and *Culex quinquefasciatus* mosquito larvae using Vision Transformer (ViT) algorithms. We employed a quantitative research method with a development research approach. The method focuses on developing a classification system using a relatively new architecture - Vision Transformer, which Google introduced to the world in 2020. Our methodology consists of several systematic stages to ensure the research runs in a structured manner and achieves expected objectives. We utilized public image data from Kaggle and Harvard Medical School that relates to mosquito larvae classification. Data was obtained through web scraping techniques and stored in IMG format. The dataset comprises mosquito larvae images, larvae species types, and body part shapes including body and head structures. Our dataset contains 2,700 images divided into three categories: Training, Validation, and Testing. The dataset division follows the standard ratio of 70% Training, 20% Validation, and 10% Testing. We designed the dataset distribution according to class requirements and framework specifications to achieve better results in both accuracy and evaluation metrics. The data underwent training using Vision Transformer for classification purposes, encompassing preprocessing, model training, and evaluation phases.

We designed testing procedures to determine how effectively the Vision Transformer (ViT) algorithm classifies *Aedes albopictus* and *Culex quinquefasciatus* mosquito larvae. Testing will be conducted after completing the training and model validation processes. The best-performing model will be saved as a benchmark for testing data evaluation, where the data will undergo species matching validation until we obtain final results in the form of Precision, Recall, and F1-score values. The values obtained from the testing process will be displayed through confusion matrices and curve diagrams before finally showing classification image results. The experimental framework ensures robust evaluation of model performance across different metrics and visualization techniques.

The Vision Transformer architecture processes images by dividing them into patches and treating each patch as a token, similar to words in natural language processing. The self-attention mechanism allows the model to focus on relevant parts of the image while understanding global relationships between different regions. We configured the ViT model specifically for binary classification between the two mosquito larvae species. Data preprocessing includes image resizing, normalization, and augmentation techniques to improve model generalization. The training process involves iterative optimization using backpropagation, with careful monitoring of validation metrics to prevent overfitting. We implemented early stopping mechanisms and learning rate scheduling to optimize training efficiency.

Model performance evaluation employs multiple metrics including accuracy, precision, recall, and F1-score to provide a balanced assessment of classification capabilities. The confusion matrix visualization helps identify specific classification patterns and potential areas for improvement. Additionally, we analyze training and validation curves to understand model learning behavior and convergence patterns. Statistical significance testing ensures the reliability of our results, while comparative analysis against baseline methods demonstrates the effectiveness of the Vision Transformer approach for mosquito larvae classification tasks. The evaluation framework provides both quantitative metrics and qualitative insights into model performance characteristics.

4. Result and Discussion

4.1 Results

4.1.1 Data Collection

This research uses public image data from Kaggle and Harvard Medical School that is relevant to mosquito larvae classification. Data was obtained using web scraping techniques and stored in IMG format. The dataset consists of mosquito larvae images, mosquito larvae types, and body part shapes such as body and head. During the collection process, filtering was performed to ensure that only relevant images with good quality for the mosquito larvae classification system using the Vision Transformer method were selected.

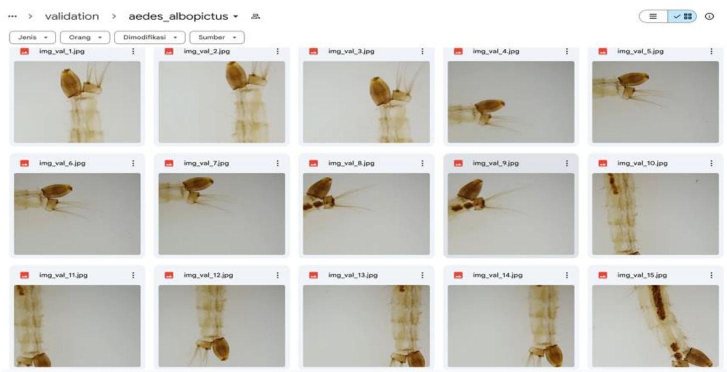

Figure 1. Example of Mosquito Larvae Dataset

4.1.2 Dataset Split

Dataset splitting aims to facilitate the system in performing classification and also avoid overfitting. The collected data amounts to 2,700 images that will be analyzed further. The dataset used totals 2,700 images divided into three classes: Training, Validation, and Testing. The dataset division is calculated according to class requirements and design with the formula 70% Training, 20% Validation, and 10% Testing. This dataset division aims to produce better results in terms of accuracy and evaluation. The data is then trained using Vision Transformer for classification purposes, which includes dataset splitting, model training, and evaluation.

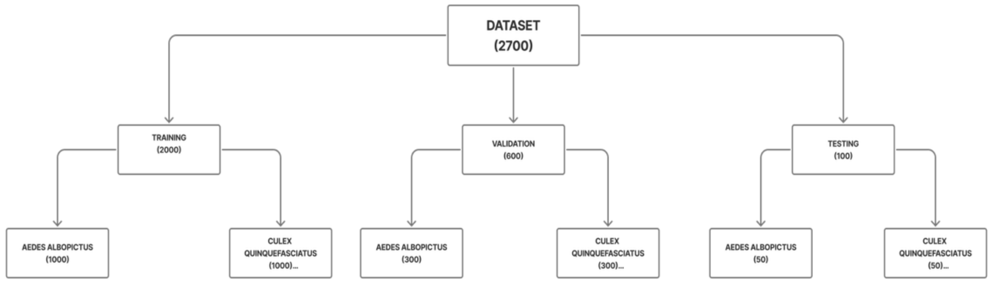

Figure 2. Dataset Split

4.1.3 Vision Transformer (ViT) Architecture Implementation

After going through the dataset split process, the dataset in the form of images will enter the classification process using the Vision Transformer method. In this process, images are resized to the appropriate size (224x224) and broken down into small patches. Then each patch is converted into vectors (embedding) using linear projection. After going through the image preprocessing process, the dataset will be arranged and input into the Vision Transformer architecture. Patch embeddings are added with positional encoding so the model knows the spatial position of each piece. Input is fed into the transformer encoder, similar to text input in NLP.

4.1.4 Model Development and Evaluation

In the model development and evaluation process, the dataset will be managed to obtain a model with the best accuracy that will be saved as the best model class. For development, a pretrained ViT model (for example: vit-base-patch16-224) is used from HuggingFace, then the model is retrained (fine-tuned) using the larvae dataset. After this process, the model will be optimized using an optimizer (Adam) and loss function (CrossEntropy). After going through the model development process, we successfully obtained the best model history results that have been stored in the system which will become a benchmark for evaluation process assessment. In this process, the model will be tested to assess how accurately it distinguishes mosquito larvae images and evaluation results will be assessed using metrics such as: accuracy, precision, recall, and F1-score.

4.1.5 Final Classification Results

In the final classification results process, researchers will display and describe the results of optimization classification of *Aedes albopictus* and *Culex quinquefasciatus* mosquito larvae using the Vision Transformer (ViT) method in the form of confusion matrix, curve graphs, accuracy, precision, recall, and F1-score.

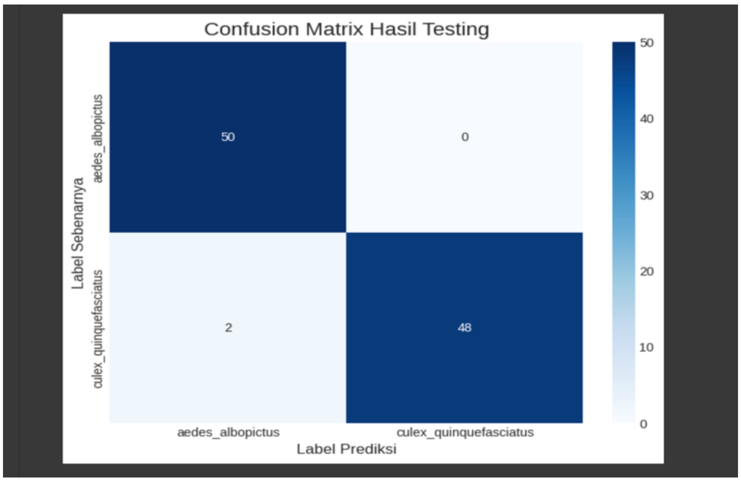

Figure 3. Confusion Matrix

With the following explanation:
1) 50 Aedes images were correctly recognized as Aedes.
2) 48 Culex images were correctly recognized as Culex.
3) 2 Culex images were incorrectly recognized as Aedes.
4) No Aedes images were incorrectly recognized.

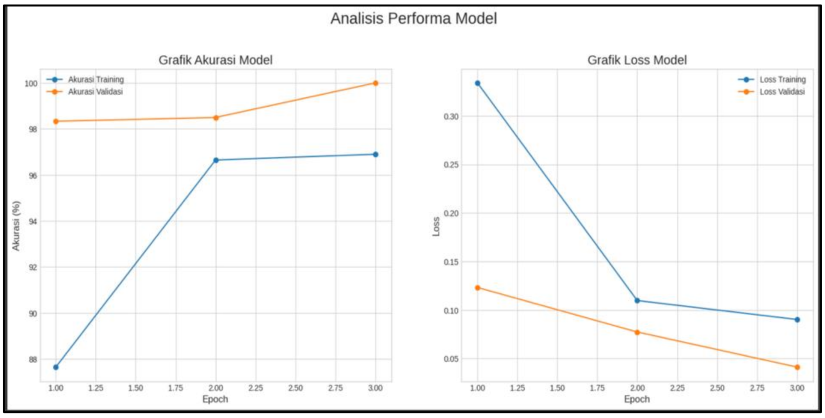

Figure 4. Curve Graph

With the following explanation:

1) Model Accuracy Graph (Left)
   Training accuracy increased significantly from around 87.5% to 96.5% approaching 97%, while validation accuracy remained high and stable, rising from 98.2% to 100% in the third epoch. This improvement shows that the model can learn very well without experiencing overfitting, because validation accuracy does not decrease even though training accuracy continues to increase.
2) Model Loss Graph (Right)
   Training loss decreased sharply from around 0.34 to 0.09, indicating that the model is getting better at minimizing prediction errors during the training process. The consistent decrease in validation loss from 0.12 to 0.04 also indicates that the model is increasingly precise in predicting data it has never seen before, and has good generalization capabilities.

```
Laporan Klasifikasi:
                         precision    recall  f1-score   support

      aedes_albopictus      0.9615    1.0000    0.9804        50
culex_quinquefasciatus      1.0000    0.9600    0.9796        50

              accuracy                          0.9800       100
             macro avg      0.9808    0.9800    0.9800       100
          weighted avg      0.9808    0.9800    0.9800       100
```

Figure 5. Accuracy, Precision, Recall, and F1-Score Results

Precision shows the level of model accuracy in predicting a class, where 96.15% of all predictions for the Aedes class proved correct, and 100% of predictions for the Culex class were also correct. Recall measures how well the model recognizes all data from each class, where the Aedes class is recognized perfectly (100%), while the Culex class is slightly lower (96%) because there are two Culex data that were misclassified as Aedes. Meanwhile, F1-score represents the balance between precision and recall in one metric, which in these results shows high and balanced classification performance in both classes.

Table 1. Classification Results Per Class

Class	Precision	Recall	F1-score	Support
Aedes albopictus	0.9615	1.0000	0.9804	50
Culex quinquefasciatus	1.0000	0.9600	0.9796	50

1) Total Accuracy
   An accuracy of 98.00% shows that the model successfully classified 98 out of 100 data correctly, reflecting a very high level of accuracy. There were only two classification errors, indicating that the model can distinguish between the two mosquito larvae species—*Aedes albopictus* and *Culex quinquefasciatus*—very well and reliably.
2) Average (Macro and Weighted)
   Macro average is the average of metrics (such as precision, recall, and F1-score) calculated evenly across classes without considering the amount of data in each class. Meanwhile, weighted average calculates the average by considering the proportion of data in each class. In this case, because the amount of data

between Aedes and Culex classes is balanced (50 data each), both average values are almost identical and very close to the total accuracy, which is 98.00%, showing balanced model performance on both classes.

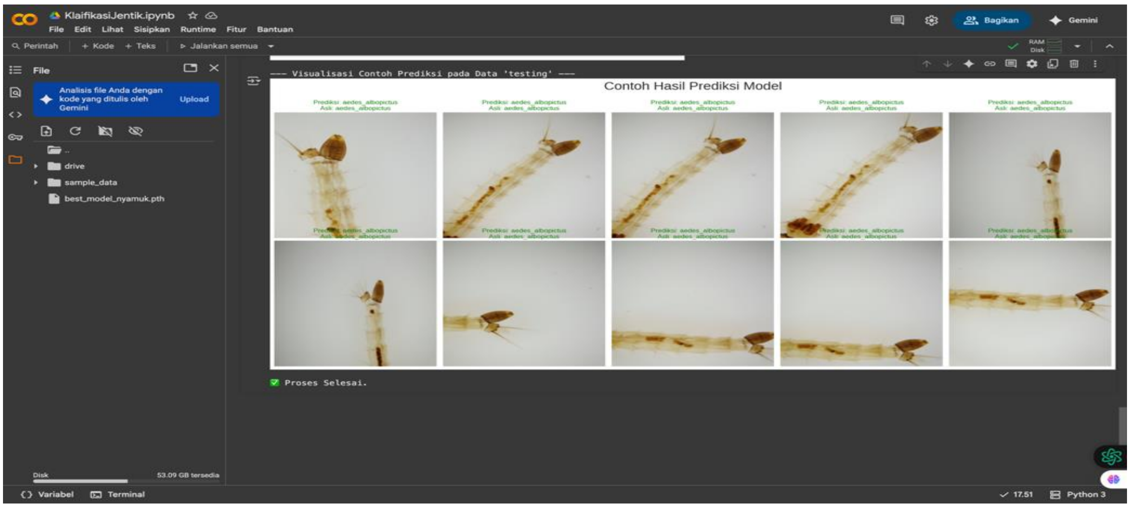

Figure 6. Final Visualization Results

In the final visualization process, researchers will display the final results in the optimization classification process of *Aedes albopictus* and *Culex quinquefasciatus* mosquito larvae using the Vision Transformer (ViT) method in the form of final evaluation result images. With the following explanation: From 100 model images for testing with a division of 50 Aedes and 50 Culex. With final classification results, Aedes achieved 100% accuracy, meaning the system successfully classified all models with the Aedes class, and Culex achieved 96.00% accuracy, meaning the system successfully classified 48 model images and incorrectly classified 2 model images as Aedes.

4.2 Discussion

The Vision Transformer model achieved remarkable performance in classifying *Aedes albopictus* and *Culex quinquefasciatus* mosquito larvae, reaching 98.00% total accuracy. The high accuracy demonstrates the model's ability to distinguish between these two species effectively, supporting findings by Dosovitskiy *et al.* (2020) who showed that Vision Transformers excel in image classification tasks through self-attention mechanisms that capture global dependencies more effectively than traditional convolutional neural networks [13]. The model's success becomes particularly significant when considering the morphological similarities between mosquito larvae at early developmental stages, which typically create challenges for automated classification systems as documented by Crans (2004) [1]. The results align with Surya *et al.* (2022) who evaluated various deep learning architectures for mosquito larvae classification and found transformer-based models showed promising results when handling complex morphological features [11].

Analysis of the confusion matrix reveals interesting classification patterns, with perfect identification of *Aedes albopictus* larvae (100% recall) while encountering slight difficulty with some *Culex quinquefasciatus* samples, where two specimens were misclassified as *Aedes*. The bias toward the *Aedes* class likely stems from morphological similarities between the two species during larval stages, particularly in siphon structure and body segmentation patterns, as documented by Reinert (2000) in taxonomic studies of *Aedes* genus [4]. Similar classification challenges have been reported by Njaime *et al.* (2024) in their automated classification study of mixed *Aedes aegypti* and *Culex quinquefasciatus* populations under field conditions, where intraspecies morphological variations created significant challenges [14]. The training and validation curves showed ideal learning patterns without overfitting, with validation accuracy reaching 100% by the third epoch and maintaining stability throughout training, indicating robust generalization capabilities consistent with findings by Adhane *et al.* (2021) who achieved similar performance metrics using deep convolutional neural networks for *Aedes albopictus* classification [12].

The high precision values (96.15% for *Aedes* and 100% for *Culex*) and balanced F1-scores (approximately 98% for both classes) demonstrate the model's reliability in practical applications, minimizing both false positive and false negative errors. These metrics become crucial for vector surveillance applications where accurate species identification directly impacts disease control strategies. The Vision Transformer's advantage in capturing complex morphological features through self-attention mechanisms enables global image processing rather than the local feature extraction typical of CNNs, as demonstrated by Liu *et al.* (2021) in their Swin Transformer architecture [13]. The global processing capability proves particularly beneficial for mosquito larvae classification, where distinguishing features may be distributed across different body regions and require thorough morphological analysis, supporting previous work by Sesulihatien *et al*. (2020) who emphasized the value of kinematic features in larvae classification [5].

The practical applications of the research extend significantly to automated disease vector surveillance systems, where the 98.00% accuracy enables reliable monitoring of mosquito populations responsible for transmitting diseases such as Dengue Hemorrhagic Fever and filariasis. The automated approach addresses the limitations of manual identification methods that require specialized taxonomic expertise and considerable time investment, as noted by De Silva and Jayalal (2020) in their digital image-based dengue mosquito larvae identification system [3]. Implementation of such systems could transform vector control programs by providing rapid, accurate species identification in field conditions, supporting the integrated vector management strategies recommended by WHO. Similar automated classification systems have shown promise in other vector surveillance applications, as demonstrated by Azman and Sarlan (2020) who developed an *Aedes* larvae classification and detection system using deep learning [6], and Saeed *et al*. (2021) who applied convolutional neural networks for mosquito larvae detection [7].

Despite the excellent performance, the model shows limitations in classifying certain *Culex quinquefasciatus* specimens, likely due to intraspecies morphological variations or suboptimal image quality. These challenges are common in automated entomological identification systems, as noted by González-Pérez *et al*. (2021) who developed alternative identification methods using acoustic features to complement visual classification [20]. Future research should focus on expanding the dataset with more diverse morphological variants and implementing advanced data augmentation techniques, following successful approaches demonstrated by Akbar and Mulyana (2022) in their optimization studies [2][15]. Additionally, incorporating multi-modal approaches that combine morphological features with behavioral or acoustic characteristics could enhance classification accuracy, particularly for ambiguous specimens. The research makes substantial advances in medical entomology by providing a robust, automated classification method that can be integrated into existing vector surveillance frameworks. The Vision Transformer's superior performance in handling complex morphological features opens new possibilities for automated identification of other medically relevant arthropod vectors. The methodology established in the study provides a foundation for developing surveillance systems that could significantly enhance global vector control efforts, particularly in resource-limited settings where taxonomic expertise may be scarce. The success of the approach also demonstrates the broader potential of transformer architectures in biological image analysis, extending beyond traditional applications in natural language processing and computer vision to specialized domains requiring fine-grained morphological discrimination.

5. Conclusion and Recommendations

Our research findings reveal that the Vision Transformer model performs exceptionally well when classifying mosquito larvae. The model correctly identifies all *Aedes* images with perfect accuracy (100%) while making just two errors when classifying *Culex* larvae as *Aedes*. Such high accuracy makes ViT an excellent choice for automated larvae detection systems, especially for monitoring disease vectors like Dengue Hemorrhagic Fever (DHF) and filariasis with both efficiency and precision. The Vision Transformer model we tested achieved 98.00% accuracy, with precision values of 96.15% for *Aedes* and 100% for *Culex*. Recall values reached 96.00% for *Aedes* and 100% for *Culex*, while F1-scores measured 98.04% for *Aedes* and 97.96% for *Culex*. These numbers show the model's strong performance in mosquito larvae classification. High recall values mean the model recognizes nearly all mosquito larvae data, with only two classification mistakes in the *Culex* class. Strong recall performance demonstrates how effectively the model identifies different mosquito larvae classes. To prevent bias toward the majority class, we need additional analysis using data balancing techniques or testing other machine learning approaches. We could also improve the model by expanding our dataset or adding new features to boost accuracy and generalization capabilities. When comparing Vision Transformer against Convolutional Neural Network (CNN), ViT shows clear advantages. At 98.00% accuracy, Vision Transformer outperforms CNN, which achieved only 79.89% accuracy in previous studies. These results establish ViT as the better model choice for mosquito larvae classification tasks.

Future research should focus on using larger, more varied datasets. Researchers can gather data independently or access public dataset sources to improve classification accuracy and reliability. Vision Transformer needs substantial data volumes to reach peak performance, making dataset size crucial. We suggest trying different methods and technologies, particularly DEiT (Data-efficient Image Transformer), which works well even with smaller datasets and shorter research periods. Platforms like Jupyter Notebook offer flexibility during model development and testing while saving valuable time. Expanding research scope presents another opportunity for advancement. While our study examined two mosquito larvae types - *Aedes albopictus* and *Culex quinquefasciatus* - future work should include additional species like *Aedes aegypti*, *Culex* *pipiens*, and *Anopheles sundaicus*. Broader species coverage would advance classification technology development and support public health initiatives, particularly disease vector control programs.

References

[1] Crans, W. J. (2004). A classification system for mosquito life cycles: Life cycle types for mosquitoes of the northeastern United States. *Journal of Vector Ecology*, *29*, 1-10.

[2] Akbar, F., & Mulyana, D. I. (2022). Optimasi klasifikasi jentik nyamuk pada citra digital menggunakan algoritma genetika dan augmentasi. *Jurnal Widya*, *3*(2), 156-168. https://doi.org/10.54593/awl.v3i2.98

[3] De Silva, W. D. M., & Jayalal, S. (2020, September). Dengue mosquito larvae identification using digital images. In *2020 International Research Conference on Smart Computing and Systems Engineering (SCSE)* (pp. 31-36). IEEE. https://doi.org/10.1109/SCSE49731.2020.9313003

[4] Reinert, J. F. (2000). New classification for the composite genus Aedes (Diptera: Culicidae: Aedini), elevation of subgenus Ochlerotatus to generic rank, reclassification of the other subgenera, and notes on certain subgenera and species. *Journal of the American Mosquito Control Association-Mosquito News*, *16*(3), 175-188.

[5] Sesulihatien, W. T., Yuana, D. B. M., & Basuki, A. (2020, September). Kinematic feature for classifying larvae: Aedes larvae and Culex larvae. In *2020 International Electronics Symposium (IES)* (pp. 649-654). IEEE. https://doi.org/10.1109/IES50839.2020.9231938

[6] Azman, M. I. A. B. Z., & Sarlan, A. B. (2020, October). Aedes larvae classification and detection (ALCD) system by using deep learning. In *2020 International Conference on Computational Intelligence (ICCI)* (pp. 179-184). IEEE. https://doi.org/10.1109/ICCI51257.2020.9247647

[7] Saeed, M. S., Nazreen, S. F., Ullah, S. S. S. A., Rinku, Z. F., & Rahman, M. A. (2021, January). Detection of mosquito larvae using convolutional neural network. In *2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)* (pp. 478-482). IEEE. https://doi.org/10.1109/ICREST51555.2021.9331235

[8] Krishnan, K. S., & Krishnan, K. S. (2021, October). Vision transformer based COVID-19 detection using chest X-rays. In *2021 6th International Conference on Signal Processing, Computing and Control (ISPCC)* (pp. 644-648). IEEE. https://doi.org/10.1109/ISPCC53510.2021.9609375

[9] Li, J., Yang, Z., & Yu, Y. (2021, September). A medical AI diagnosis platform based on vision transformer for coronavirus. In *2021 IEEE International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI)* (pp. 246-252). IEEE. https://doi.org/10.1109/CEI52496.2021.9574576

[10] Wang, W., Xie, E., Li, X., Fan, D. P., Song, K., Liang, D., Lu, T., Luo, P., & Shao, L. (2022). PVT v2: Improved baselines with pyramid vision transformer. *Computational Visual Media*, *8*(3), 415-424. https://doi.org/10.1007/s41095-022-0274-8

[11] Surya, A., Peral, D. B., VanLoon, A., & Rajesh, A. (2022). A mosquito is worth 16x16 larvae: Evaluation of deep learning architectures for mosquito larvae classification. *arXiv preprint arXiv:2209.07718*. https://doi.org/10.48550/arXiv.2209.07718

[12] Adhane, G., Dehshibi, M. M., & Masip, D. (2021). A deep convolutional neural network for classification of aedes albopictus mosquitoes. *IEEE Access*, *9*, 72681-72690. https://doi.org/10.1109/ACCESS.2021.3079700

[13] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Stephen, L., & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 10012-10022). https://doi.org/10.1109/ICCV48922.2021.00986

[14] Njaime, F. C. B. F. P., Máspero, R. C., Leandro, A. D. S., & Maciel-de-Freitas, R. (2024). Automated classification of mixed populations of Aedes aegypti and Culex quinquefasciatus mosquitoes under field conditions. *Parasites & Vectors*, *17*(1), 399. https://doi.org/10.1186/s13071-024-06417-z

[15] Akbar, A., & Mulyana, D. I. (2022). Optimasi klasifikasi batik betawi menggunakan data augmentasi dengan metode KNN dan GLCM. *Jurnal Aplikasi Teknologi Informasi dan Manajemen (JATIM)*, *3*(2), 92-101. https://doi.org/10.31102/jatim.v3i2.1577

[16] Nana, N., Mulyana, D. I., Akbar, A., & Zikri, M. (2022). Optimasi klasifikasi buah anggur menggunakan data augmentasi dan convolutional neural network. *Smart Comp: Jurnalnya Orang Pintar Komputer*, *11*(2), 148-161. https://doi.org/10.30591/smartcomp.v11i2.3527

[17] Mulyana, D. I., Hartanto, H., & Yel, M. B. (2022). Optimasi klasifikasi bunga kantong semar dengan menggunakan algoritma naïve bayes, data augmentasi dan PSO. *Swabumi*, *10*(2), 167-174.

[18] Ruswandi, M., Mulyana, D. I., & Awaludin, A. (2022). Optimasi klasifikasi kematangan buah alpukat menggunakan KNN dan fitur statistik. *Smart Comp: Jurnalnya Orang Pintar Komputer*, *11*(2), 210-219. https://doi.org/10.30591/smartcomp.v11i2.3531

[19] Mohommed, M., Rajakaruna, P., Kehelpannala, N., Perera, A., & Abeysiri, L. (2020, December). DenGue CarB: Mosquito identification and classification using machine learning. In *2020 2nd International Conference on Advancements in Computing (ICAC)* (Vol. 1, pp. 126-131). IEEE. https://doi.org/10.1109/ICAC51239.2020.9357133

[20] González-Pérez, M. E., González-Pérez, C. S., García-Betances, R. E., & Alvarez-Rodríguez, M. A. (2021). Mosquito sound-based identification system (MSBIS): A proposal for identifying Aedes aegypti, Culex quinquefasciatus and Anopheles stephensi using smartphone recordings. *Parasites & Vectors*, *14*(1), 1-14. https://doi.org/10.1186/s13071-021-04647-4

[21] Radikto, R., Mulyana, D. I., Rofik, M. A., & Zoharuddin Zakaria, M. O. (2022). Klasifikasi kendaraan pada jalan raya menggunakan algoritma convolutional neural network (CNN). *Jurnal Pendidikan Tambusai*, *6*(1), 1668-1679. https://doi.org/10.31004/jptam.v6i1.3179

[22] Utami, P. N., Arinal, V., & Mulyana, D. I. (2022). Klasifikasi dehidrasi tubuh manusia berdasarkan citra RGB pada warna urine menggunakan metode K-nearest neighbor. *J. Media Inform. Budidarma*, *6*(1), 18-26.