Detection and Classification of Mosquito Larvae Based on Deep Learning Approach

Pauzi Ibrahim Nainggolan*, Syahril Efendi, Mohammad Andri Budiman, Maya Silvi Lydia, Romi Fadillah Rahmat, Dhani Syahputra Bukit, Umi Salmah, Sri Malem Indirawati and Riza Sulaiman, *Member, IAENG*

Abstract— **This paper addresses the challenge of suboptimal biological control of the Aedes aegypti mosquito, which serves as a vector for Dengue, Chikungunya, and Zika viruses. These arthropods pose a significant threat to approximately one-third of the global population annually, capable of causing severe pain, hemorrhagic fever, and brain defects in unborn children with a single bite. The research introduces a technologically effective solution employing deep neural networks (DNNs) to conduct surveys during the immature larval stage. Our approach enables automatic identification of the biological vector in the larval stage, achieving a higher accuracy of 81.7% in region-of-interest segmentation. Moreover, it classifies larvae as Aedes positive or negative with an accuracy of 97%, significantly reducing response time from days to seconds without human intervention. The proposed solution is cost-effective, minimizing the need for trained entomologists, laboratories, and expensive equipment. Utilizing microscope-based image acquisition hardware, a computer with CPU hardware, and a petri dish, sample capture and analysis become straightforward. The advantages of this proposal are particularly valuable in underdeveloped countries and remote regions, where economic constraints may limit access to preventive health services and biological vector control.**

Index Term— **Aedes Agepty, computer vision, panoptic segmentation, deep Learning.**

Manuscript received October 8, 2023; revised November 5, 2024.

This work was supported by Ministry of Education of Republic Indonesia under Research Grant 2023, in accordance with the Letter of Assignment Agreement, Number: 12/UN5.2.3.1/PPM/KP-DRTPM/B/2023.

Pauzi Ibrahim Nainggolan is an Assistant Professor of Department of Computer Science, Universitas Sumatera Utara, Medan, 20155, Indonesia (e-mail: nainggolan@usu.ac.id)

Syahril Efendi is a Professor of Faculty of Department of Computer Science, Universitas Sumatera Utara, Medan, 20155, Indonesia (e-mail: syahril1@usu.ac.id)

Mohammad Andri Budiman is an Associate Professor of Department of Computer Science, Universitas Sumatera Utara, Medan, 20155, Indonesia (e-mail: mandrib@usu.ac.id)

Maya Silvi Lydia is an Associate Professor of Department of Computer Science, Universitas Sumatera Utara, Medan, 20155, Indonesia (e-mail: maya.silvi@usu.ac.id)

Romi Fadillah Rahmat is an Associate Professor of Department of Information Technology, Universitas Sumatera Utara, Medan, 20155, Indonesia (e-mail: romi.fadillah@usu.ac.id)

Dhani Syahputra Bukit is an Assistant Professor of Faculty of Public Health, Universitas Sumatera Utara, Medan, 20155, Indonesia (e-mail: dhanibukit@usu.ac.id)

Umi Salmah is an Associate Professor of Faculty of Public Health, Universitas Sumatera Utara, Medan, 20155, Indonesia (e-mail: umisalmah@usu.ac.id)

Sri Malem Indirawati is an Associate Professor of Faculty of Public Health, Universitas Sumatera Utara, Medan, 20155, Indonesia (e-mail: srimalem@usu.ac.id)

Riza Sulaiman is a Professor of Institute of Visual Informatics, Universiti Kebangsaan Malaysia, Bangi, 43600 Malaysia (e-mail: riza@ukm.edu.my)

I. Introduction

Dengue hemorrhagic fever (DHF) is widespread in over 100 countries, mainly in tropical areas. The initial cases in Indonesia emerged in 1968 in Jakarta and Surabaya. In 2015, there were 126,675 cases with 1,229 deaths reported across 34 provinces [1]. Data from the Ministry of Health's P2P Directorate General indicates a Dengue Incidence Rate in Indonesia of 51.5 per 100,000 population in 2019, dropping to 40 in 2020. North Sumatra Province recorded 7,584 DHF cases with 37 deaths in 2019[2][3].

Despite efforts like spraying and larvicides, DHF cases persist. The primary prevention strategy focuses on community empowerment, notably through the Jumantik One House One Movement initiative [4]. This program, coordinated by the Dengue Hemorrhagic Fever Operational Working Group (Pokjanal DBD) and Larvae Monitor (Jumantik), aims to enhance the Eradication of 3M Plus Mosquito Nests (PSN) within communities [5].

The process of determining the Aedes Breeding Index (ABJ) through house surveys is time-consuming and relies on the accuracy of Jumantik, the larva monitor, to identify mosquito types. Integrating technology is crucial to improve the efficiency and accuracy of ABJ data, providing real-time updates with geographical information for predicting potential Dengue Hemorrhagic Fever (DHF) outbreaks. The use of advanced technology, particularly artificial intelligence, can transform larva monitoring, significantly reducing the time needed by Jumantik to determine ABJ. This research marks a crucial step in transitioning from manual larva monitoring to digitization, addressing information constraints in implementing DHF control programs [6].

Numerous studies have explored the application of Computer Vision in the identification and classification of Aedes mosquito larvae [7][8][9][10]. The Histonet research by Azman and Sarlan (2020) adopts a deep learning approach, leveraging predictions based on object sizes and the total number of input images that may be messy or overlapping. This research encompasses mosquito larvae data and medical datasets, demonstrating advantages over segmentation methods. Notably, Histonet research highlights that directly learning and predicting object size distributions, without the need for explicit pixel-accurate instance segmentation, significantly improves performance. This streamlined approach results in an 85% reduction in model parameters, enabling a more efficient architecture that can be trained with a greater number of annotations [11]. However, it is essential to note that the Histonet research primarily focuses on identifying the number of objects without classifying them. To address this limitation, it becomes imperative to

incorporate instance segmentation, which allows for the identification of the class of each object found.

The challenge within instance segmentation lies in the need to accurately detect and segment each instance of all objects present in an image. This process combines aspects of classic computer vision tasks, where object detection classifies individual objects and locates each using bounding boxes, and semantic segmentation categorizes each pixel into a fixed set of classes, irrespective of object instances. Many implementations employ intricate methods to achieve optimal results. However, Kaiming He et al.'s findings suggest that a simple, flexible, and fast system can outperform the results obtained by previous, more complex instance segmentation approaches [12].

Panoptic segmentation has recently emerged as a method to integrate the tasks of instance segmentation (for individual item classes) and semantic segmentation (for item classes in general). In implementing this approach, separate networks are employed for various semantic segmentation instances, avoiding the need for joint computations. In a study conducted by Kirillov, Girshick, et al., the aim was to unify these tasks at the architectural level by designing a single network capable of handling both tasks concurrently. This involved enhancing the popular instance segmentation method, R-CNN Mask, with semantic segmentation branches, utilizing a shared Feature Pyramid Network (FPN) backbone. The research extensively examined a version of Mask R-CNN coupled with a minimally extended FPN, referred to as Panoptic FPN, demonstrating its robustness and accuracy as a foundational framework for both tasks [13].

The modern challenge of dealing with "information overload" and "data deluge" is being tackled through interdisciplinary research, spanning areas such as visualization, statistical data analysis, machine learning, data mining, and perceptual and cognitive sciences. The primary objective is to extract valuable information and generate reliable insights, thereby creating new knowledge from previously unexplored data. However, there's skepticism about the simplicity and effectiveness of this sub-specialty in handling the ever-expanding volumes of data. Keim et al. highlighted that approaches solely focused on analytics or visuals may not adequately reveal significant information from rapidly growing complex datasets and effectively communicate it to humans. The necessity for more comprehensive methods integrating analytical and visual approaches is emphasized to address the challenges posed by increasingly massive datasets [14].

In the pursuit of generating knowledge and uncovering hidden opportunities within vast and intricate datasets, James (Jim) Joseph Thomas (26 March 1946 – 6 August 2010) played a pivotal role. He not only created and sold but also established the field of visual analytics [15][16]. Visual analytics is defined as the discipline leverages visualization and interaction techniques to seamlessly incorporate human expert judgment into the data analysis process[17].

Visual perception engages various brain regions, and even the most basic perceptual tasks rely on a cascading sequence of integrated processes. This sequence begins with the initial sensory registration of a stimulus and progresses through higher cortical regions, culminating in decision-making and, at times, appropriate actions. Additionally, systems related to caring, expecting, and rewarding can also come into play. The coordination of these interdependent components of the brain, along with processing modules, collaborates harmoniously to accomplish specific behavioral objectives.

To accommodate perceptual learning, a successful predictive model must integrate several crucial functions. It must encode stimuli, ascertain the process of making task-relevant decisions, and incorporate learning and testing paradigms during training. Each of these functions can be instantiated in distinct modules. For instance, the representation module is responsible for sensory encoding and the resultant representation. The decision module dictates the decision-making process. The learning module establishes the learning rules applied. The overall model may also impact top-down attention, feedback, and reward effects. Naturally, any behavioral model must account for internal noise as well.

Quantitative models or processes of this kind aim to produce precise and testable predictions for observed phenomena in a given experiment. Assessing the model's predictions against similar outcomes helps determine the effectiveness of the proposed representation, decision, and learning principles. This three-way interaction involving modeling, theory, and experimentation is crucial for advancing our understanding of perceptual learning.

The main goal of this paper is to achieve optimal accuracy. Various factors, such as lighting conditions, shooting distance, and scale variations, can influence accuracy values. Thus, the paper introduces a deep learning-based detection and classification technique using the One-stage Top-Down method for Panoptic Segmentation. Adopting the FPSNet segmentation architecture with modifications enhances the classification accuracy. Data collection involved two methods: direct collection and laboratory-based data collection.

II. Proposed Method

This research is crucial in establishing a model for the classification of mosquito larvae species. The classification model is constructed by extracting morphological features from the segmentation of each organ in various mosquito larvae species. The practical implementation of this model serves as a tool to strategically plan, execute, and enhance the effectiveness of response actions. It also aids in determining the best strategies for mitigating the impact of infectious diseases transmitted from animals (zoonotic diseases). The development of an adequate algorithm is particularly important to accurately identify species, enabling the tracking of mosquito breeding patterns in different regions [18][19][20].

This research uses a comparison of 2 different algorithm models to determine Aedes mosquito larvae. Detection of larval objects is carried out based on annotated body parts. The detection model compared in this research is FPSnet with Yolov5. Annotation is carried out as if we understand an object by dividing the body parts of mosquito larvae into 3 parts, where each part is annotated.

To develop a model capable of recognizing Aedes data, we used the other two genus as part of the validation process to detect Aedes larvae.

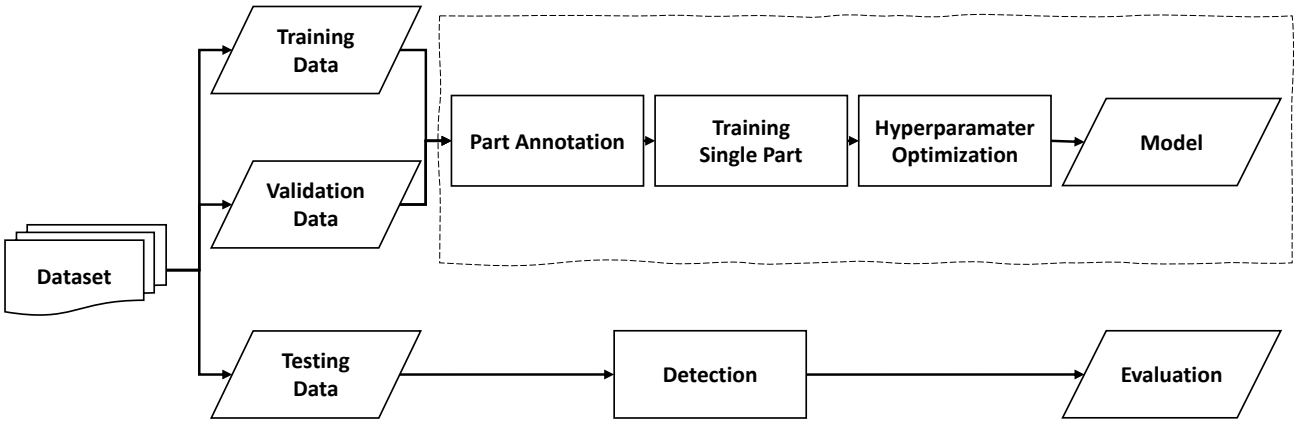

Fig 1. Proposed Method

The research flow scheme comprises five stages, each with specific goals that must be achieved to reach milestones and facilitate the continuous progression of the research, as illustrated in Figure 1. The stages of the research flow are outlined as follows:

A. Data Acquisition

The acquisition of mosquito larvae image data was a collaborative effort with the Centre for Environmental Health Engineering and Disease Control (BTKLPP) Medan. The goal was to collect targeted image data, totaling 600 images, with a breakdown of 150 images for each species. For each species, the images were obtained from 50 larvae, captured in three instances. The size of the acquired images is standardized at 640x480 pixels.

B. Data Cleaning & Preprocessing

The image data cleaning stage involves three processes. Initially, the first process fills in missing values in the images to match the acquired image size. Subsequently, the second process eliminates blurry or out-of-focus image data with inconsistent values. The final process focuses on removing redundant data and outliers.

During this preprocessing stage, three datasets are prepared: RGB data, binary data, and grayscale data. Binary data is derived by transforming the image data from three channels into one-channel binary data, accomplished by determining a threshold value. Similarly, grayscale datasets undergo modification through threshold setting.

The RGB image dataset undergoes labeling by creating segmentation boxes for three organs in the larva: the head, siphon, and stomach. The resulting binary dataset is subjected to an edge detection process to generate input values for larva classification. Grayscale data is utilized for specifying the class of mosquito larvae data.

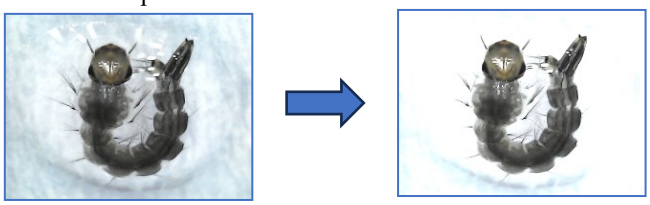

Fig 2. Preprocessing Data

The second stage of preprocessing applied to the dataset involved enhancing the contrast and brightness of the images. These enhancements aimed to optimize the image size calculation process by obtaining the best values. This image improvement process is essential for accurately determining starting and ending points. The adjustments included a 20% increase in brightness and a 40% increase in contrast, as depicted in Figure 2.

C. Training Model

The segmentation results from each area are characterized through feature extraction to identify the dominant criteria for the segmentation of each mosquito larvae species. Morphological characteristics, obtained using shape descriptors, serve as the features extracted for each species and segmentation. The resulting feature map from the shape descriptor is utilized as input in developing a classification model.

To enhance the Faster Panoptic Segmentation Network (FPSNet), an advanced deep learning model for instance segmentation, this study utilizes an improved version. The FPSNet integrates a Convolutional Neural Network (CNN) backbone for feature extraction and a head net (the network after RoI Align) for bounding box classification, regression, and mask segmentation. Three CNN backbones, ResNet-50-FPN, ResNet-101-FPN, and ResNeXt-152-FPN, are selected for comparison based on accuracy and time efficiency (network-depth-feature). The proposed model architecture is depicted in Figure 3 and built based on the work by De Geus et al.[18]. ResNet and ResNeXt are state-of-the-art CNN models and widely employed backbones in FPSNet, with ResNeXt having advantages over ResNet through a repeated building block that aggregates a set of transformations with the same topology. Generally, a deeper model is anticipated to achieve higher accuracy but requires more computational time. A Feature Pyramid Network (FPN) is incorporated into all backbone models to extract additional feature maps from different layers, enhancing instance segmentation[13].

1) Annotation

The image annotation for this study is conducted using a graphical image box annotation tool within the Python environment. To uphold quality and consistency, the annotation process adheres to the following criteria:

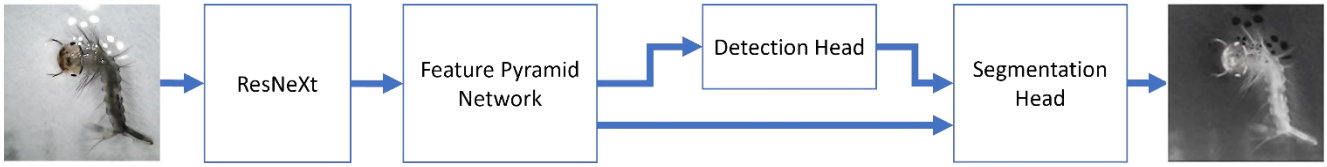

Fig 3. Proposed Modified FPSNet Architecture to detect larvae image

1. Only the parts of the larva inside the image should be annotated, excluding any portions outside of the room.
2. Each annotated object must have an explicit boundary, disregarding implicit borders or edges obscured by shadows.
3. Objects should be definitively classified into one predefined material category; otherwise, they will be disregarded.
4. In cases of significant overlap where major parts of objects are obscured, the annotation will treat them as a single object rather than multiple non-overlapping ones.

Every object, or visible part of an object, meeting the specified criteria is annotated using a box to delineate its boundary. The annotation outcomes are then stored in a JSON file following the Microsoft COCO format [21]. To ensure accuracy, each annotated image undergoes cross-validation by two of the authors, reaching a consensus.

Figure 4 illustrates examples of annotated images, where each labeled instance is highlighted by a colored mask and a bounding box displaying its category name in the top-left corner.

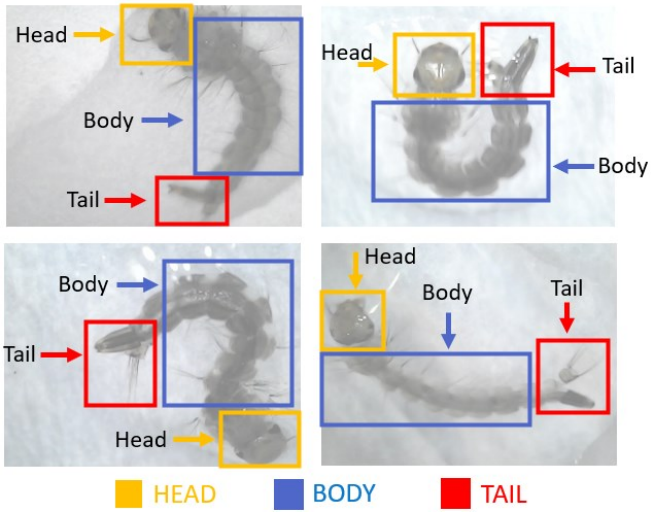

Fig 4. Image Annotation

2) Detection

The FPSNet architecture necessitates a backbone that can generate a singular feature map and execute object detection. This object detection network is vital for producing bounding boxes utilized in forming attention masks for object instances, while the single feature map is crucial for making dense panoptic segmentation predictions.

The output of the ResNet-50-based Feature Pyramid Network comprises a set of feature maps from various levels of the feature extractor. However, for predicting dense panoptic segmentation, a single feature map is required. Similar challenges were encountered by the authors in [13] when attempting semantic segmentation on a multi-scale feature map. To address this, they resolved the issue by upsampling and combining the multiple layers of feature maps. We adopt the same implementation strategy in our approach.

D. Classification

The classification process in this research employs two classifiers to achieve the highest accuracy. Specifically, ResNet and VGG are chosen as the classifiers due to their current popularity. ResNet tackles the challenge by introducing two types of 'shortcut connections': Identity shortcut and Projection shortcut. VGG16 encompasses a total of 138 million parameters. It is crucial to highlight that all convolutional layers are of size 3x3, and max-pooling layers are of size 2x2 with a stride of two.

E. Metrics

The classification model is initially tested using test data. Subsequently, the classification model undergoes testing with the k-fold cross-validation technique. Beforehand, the dataset is divided into training data, validation data, and test data, with a dataset distribution ratio of 80%, 10% for validation data, and 10% for test data.

In the field of computer vision, several assessment indicators or metrics are commonly used to evaluate the performance of computer vision systems. Here are some of the commonly used evaluation metrics in object detection:

Precision quantifies the extent of true positive expectations out of all positive predictions made by the model. Precision is calculated utilizing the taking-after equation

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

Recall quantifies the extent of true positive forecasts out of all real positive instances within the dataset. The recall is calculated utilizing the taking-after-the equation.

$$Recall = \frac{TP}{TP+FP} \quad (2)$$

Mean Average Precision (mAP): mAP expands the average precision (AP) by averaging AP values over numerous protest classes or targets. It is commonly utilized in multi-class question discovery errands, where it gives a general degree of location execution over diverse target categories.

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \quad (3)$$

Panoptic Quality (PQ), the common evaluation metric described in [18], is what we employ. It has two components: the Detection Quality (DQ) solely evaluates instance classes, while the Segmentation Quality (SQ) evaluates the quality of all categories. Equation 5 presents the mathematical representations of PQ, SQ, and DQ, where p and g stand for predictions and ground truth, and TP, FP, and FN stand for true positives, false positives, and false negatives,

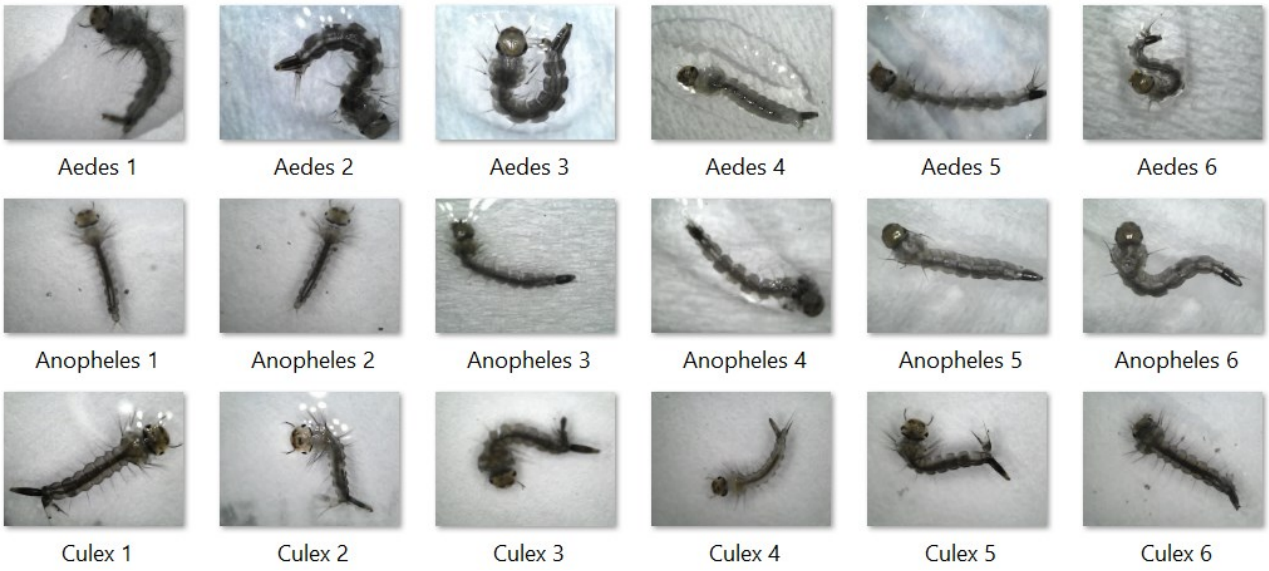

Fig 5. Dataset of Mosquitos larva

respectively. It is simple to see that DQ may be thought of as a type of detection accuracy and that SQ is the common mean IoU measure normalized for matching instances. If the pixel IoU of the prediction and the ground truth are more than 0.5, the matching threshold is set to 0.5, and the prediction is regarded as successful.

$$PQ = \frac{\sum_{(p,g)\in TP} IOU(p,g)}{|TP|} \times \frac{|TP|}{|TP|+\frac{1}{2}|FP|+\frac{1}{2}|FN|} \quad (4)$$

III. Test, Result and Discussion

A. Image acquisition Datasets

The first task of this research is to collect the dataset used in this research. The total number of larval image data is 1185 images with a dataset class ratio of 284 Aedes, 166 Anopheles, and 735 Culex. With the assistance of entomologists, the image data is categorized into three groups: image data of Aedes larvae, image data of non-Aedes larvae Anopheles and image data of non-Aedes larvae Culex. Figure 5 shows some images included in the sample and their categories.

Fig 6. Data acquisition

As previously stated, we collect the larvae from two source: dataset from the Centre for Environmental Health Engineering and Disease Control and our own data acquisition. The image data acquisition was carried out using a digital microscope connected to an Android handheld as shown in Figure 6. The larvae collection took place at various locations in Tebing Tinggi City. The acquired images have dimensions of 640x480 pixels, a resolution of 96 dpi, and a depth of 24 bits. The image data results are presented in the figure 5. After normalization we used 150 image data on each genus of mosquito larvae.

B. Experimental Results

Our aim is to find the model with the best performance that able to categorize the larvae into three categories; Aedes, Culex, and Anopheles. The process consists of two part: detection and classification. Each part will have its own model to be compare.

1) Training Model

During the detection process, we used two model: YOLOv5 and FPSnet. The result is measured by comparing the precision and recall. Based on the results gathered during the test, we found out that YOLOv5 performed better than FPSnet. There is, however, a performance difference where FPSnet gave a better time performance.

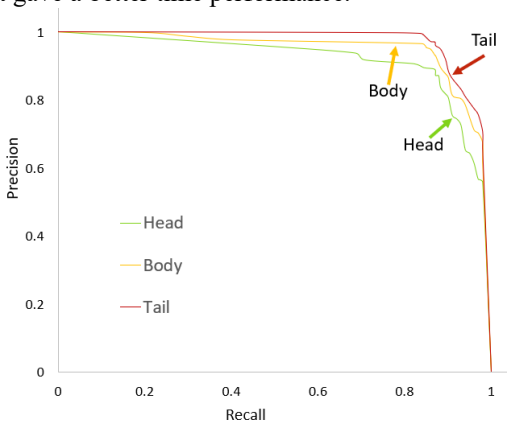

Fig 7. P-R curve on the dataset

TABLE I
THE COMPARISON OF SEGMENTATION METHODS OF MOSQUITO LARVA (%)

Algorithm	Species	P	R	mAP	PQ
YOLOv5	Aedes	90.4	84.5	94.2	74.5
YOLOv5	Culex	89.2	87.7	92.4	78.3
YOLOv5	Anopheles	93.2	86.8	93.9	81.7
FPSnet	Aedes	54.8	64.2	86.5	62.5
FPSnet	Culex	78.1	78.3	83.4	61.1
FPSnet	Anopheles	54.8	63.0	70.8	74.3

But the difference is negligible. The performance of precision-recall can be seen in figure 7 and we separate the curve for each instance we aimed to detect, yellow for head, blue for body, and red for tail. As expected, tail detection performed better than head and body due to its unique characteristics.

Figure 8 show the detection and instance segmentation results using YOLOv5. As shown in the image, our proposed method is able to differentiate the head, body, and tail correctly. The intersections between segmented instance mostly caused when the shape of the larvae is too complex to be segmented by a rectangular region.

2) Tuning the Hyperparameter

In this experiment, hyperparameter tuning was performed by combining various values of the number of epochs (150, 200, 250) and learning rates (0.01, 0.001). These combinations were carefully explored to find the best and most optimal model. The results of the hyperparameter tuning presents the best-tested combinations along with their performance outcomes. After performing hyperparameter tuning, including adjusting the number of epochs and learning rate, the optimal combination was found to be 200 epochs with a learning rate of 0.001. This combination resulted in the highest mAP value among the other combinations, which was 0.893 for all classes. Therefore, the model resulting from this training was chosen to be the system model.

Based on the model training process with 200 epochs and a learning rate of 0.001, the evaluation results of the built model are presented. The generated graphs provide a visual representation of the overall performance of the model. These graphs include several important metrics such as box loss, class loss, and object loss, as well as performance evaluation metrics like precision and recall, as shown in Figure 11.

Overall, the graphs indicate that the model has satisfactory performance, with metrics such as box_loss, class_loss, object_loss, precision, recall, and mAP showing good results. However, Both the training loss represented in blue line and validation loss represented in orange line decrease and

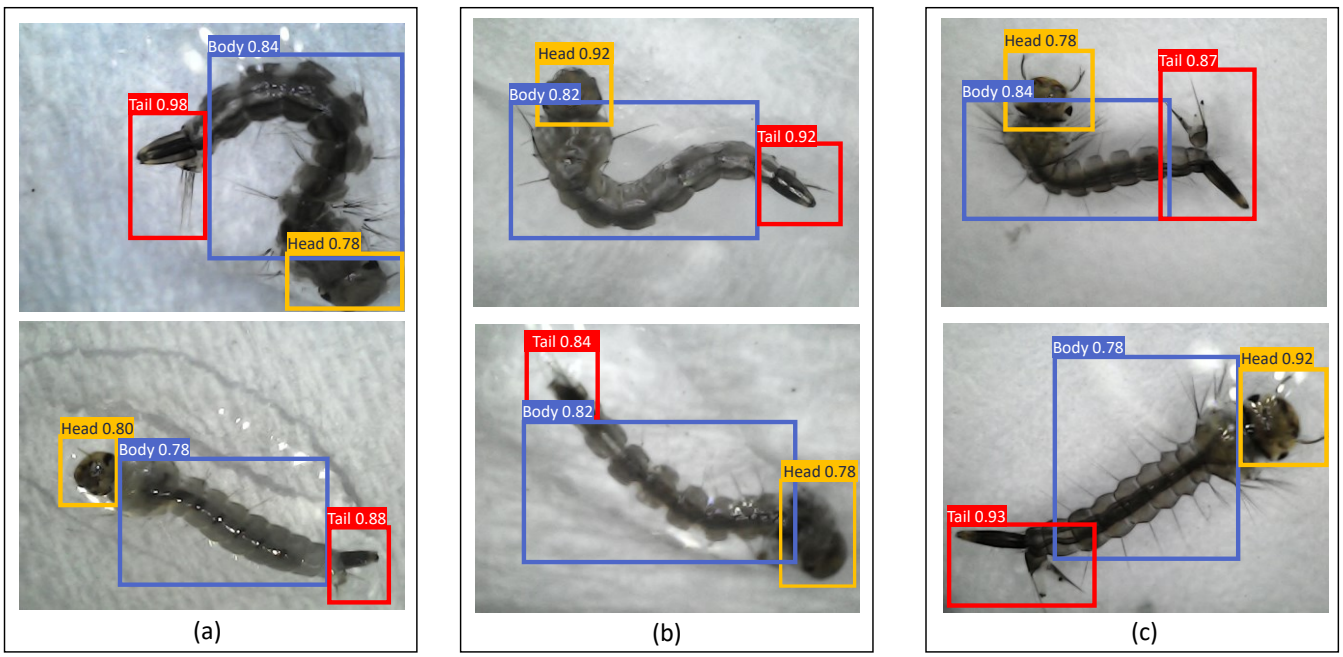

Fig 8. Detection Result (a) Mosquitos Larvae Aedes, (b) Mosquitos Larvae Anopheles, and (c) Mosquitos Larvae Culex

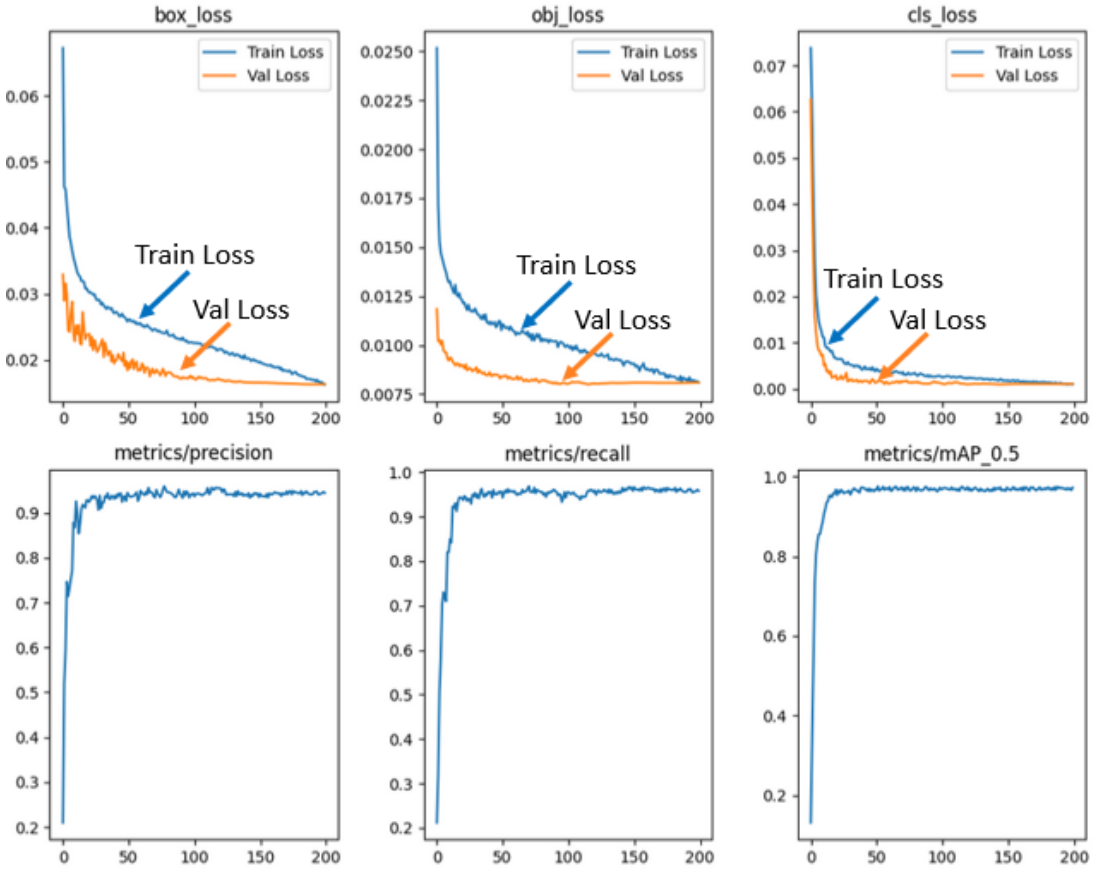

Fig 9. Model Performance Graphs

eventually stabilize at a certain point, indicating an optimal fit as shown in Figure 9.

Although the model shows satisfactory performance overall, these results indicate differences in the model's ability to identify and classify specific larval genera. The Aedes genus has the lowest mAP value among the classes, suggesting that the model might have difficulty recognizing this particular genus.

3) Testing Process

The next step of the experiment is to testing the model by classify the testing data based on its genus categorize. As previously, we argue that by classifying based on the body part of the larvae can increase the performance of classification. For this evaluation, we use Accuracy and F1-Score to assess the performance capabilities of the developed model.

Before proceeding to the classification, we first segmented the images to find the body instances of the larvae and extract the feature of each segmented parts. Using these features, we trained the models by using the labelled data. To measure the performance of the classification, we collect the result and calculate the true positive, true negatives, false positives and false negatives. Table II shows the performance of Yolov5 in classifying each part type of mosquito larva. Overall, the

TABLE II
TRAINING EPOCH 200, LEARNING RATE 0.001

epoch	train/ box_loss	train/ cls_loss	train/ obj_loss	Precision	Recall	mAP 0.5	mAP 0.5:0.95	val/ box_loss	val/ cls_loss	val/ obj_loss
1/200	0.067228	0.025172	0.073781	0.20973	0.21167	0.13159	0.06302	0.032925	0.011846	0.062667
2/200	0.046169	0.017165	0.059651	0.51979	0.3185	0.34591	0.18307	0.028987	0.010210	0.043331
3/200	0.045867	0.015277	0.042308	0.59374	0.50846	0.53533	0.26605	0.03158	0.010269	0.026807
4/200	0.043401	0.014721	0.02857	0.74629	0.58132	0.73147	0.40442	0.029883	0.01003	0.015957
5/200	0.041019	0.014443	0.020506	0.71416	0.70297	0.80619	0.48806	0.024487	0.010222	0.012190
...	…	…	…	…	…	…	…	…	...	...
196/200	0.016749	0.008222	0.00096	0.94118	0.95503	0.97129	0.71198	0.016322	0.008085	0.001052
197/200	0.016445	0.008142	0.001043	0.94257	0.95659	0.96767	0.70832	0.016331	0.008087	0.001053
198/200	0.01665	0.008168	0.001128	0.94393	0.95787	0.97033	0.71089	0.01633	0.008088	0.001054
199/200	0.016322	0.008148	0.00103	0.94641	0.96045	0.96611	0.70666	0.016332	0.008091	0.001056
200/200	0.016301	0.008097	0.001037	0.94407	0.95796	0.97197	0.71268	0.01633	0.008092	0.001055

model performance with YOLOv5 shows good performance results for each genus of mosquito larvae. Model performance tends to be optimal when the model recognizes objects from the characteristics of each part of the flick object.

TABLE III
TEST DATA CLASSIFICATION RESULTS

Class	Presicion	Recall	Accuracy	F-1 Score
Head Aedes	0.824	0.933	0.970	0.875
Trunk Aedes	0.824	0.933	0.970	0.875
Tail Aedes	0.824	0.933	0.970	0.875
Head Anopheles	1.000	0.933	0.993	0.966
Trunk Anopheles	1.000	0.933	0.993	0.966
Tail Anopheles	1.000	0.933	0.993	0.966
Head Culex	0.929	0.867	0.978	0.897
Trunk Culex	0.929	0.867	0.978	0.897
Tail Culex	0.929	0.867	0.978	0.897

C. Discussion

Our detection comparison shows that YOLOv5 gives a better detection and segmentation result. FPSnet gives a better speed in a limited hardware capability but the difference is negligible. Based on our observation, the overall result of the segmentation process able to give an acceptable result to be used in the classification process.

The proposed argument of this research is the classification process should be better given the segmented input from previous process. The discrepancies in results among the three species are notably distinct, particularly between Aedes, Culex, and Anopheles. Based on our analysis, the morphological differences between Aedes and Culex are very subtle. Meanwhile, Anopheles exhibits a significant difference in the tail.

The evaluation of classification results is not significantly different from segmentation results. Morphological features serve as the primary distinguishing characteristics between species. The comparison in classification results reveals that the accuracy of Anopheles is the highest among the other species. A comparison of algorithms reveals distinct trends: YOLOv5 detection outperforms the FPSnet algorithm significantly.

This research on object detection using a larval body part segmentation approach resulted in an accuracy of 97%. For Aedes, 99.3% for Anopheles and 97.8% for Culex. The model built using the YOLOv5 algorithm achieved a precision of 94.4% and a recall of 95.7%, with a resulting mAP 0.5 value of 0.971. The improved detection accuracy is apparent in the Aedes and Culex classes. According to the test results, the proposed algorithm excels in identifying each part of the larva's body. The highest detection is observed in the Tail section due to its distinct features compared to the Head and Body. Challenges arise in Panoptic Quality, where the tails of culex and Aedes exhibit similar features, resulting in the best PQ outcomes in the Anopheles class.

IV. CONCLUSIONS & FUTURE WORKS

This paper addresses the challenge of non-optimal biological control of the Aedes aegypti mosquito, a vector responsible for spreading Dengue, Chikungunya, and Zika viruses. These arthropods pose a threat to a significant portion of the global population annually, with a single mosquito bite capable of causing severe pain, hemorrhagic fever, and even brain defects in unborn children.

During the immature larval stage, Aedes aegypti and Aedes albopictus are easily manageable, vulnerable, and non-infectious. However, there is only a narrow window of time differentiating the innocuous larval stage from the reproductive and hazardous adult stage. Hence, a swift and precise survey and disposal of the immature larval stage become imperative.

This paper introduces a technological solution utilizing deep neural networks (DNNs) for an efficient immature larval stage survey. The proposed method automatically identifies the biological vector in the larval stage, achieving a 97.1% mAP 0.5 in region detection and a 97% accuracy in larva classification as Aedes positive or Aedes negative. This results in a significant reduction in response time from days to few seconds without human intervention. Moreover, the proposed solution proves cost-effective by minimizing the need for trained entomologists, laboratories, and expensive equipment.

The hardware requirements for implementation include a microscope-based image acquisition setup, a computer with CPU hardware, and a petri dish, making it feasible for deployment in underdeveloped countries and remote regions where economic constraints limit access to preventive health services and biological vector control.

Future work aims to enhance the research contribution further. The proposed method could be extended to video technology for Aedes aegypti larva identification and compared with other segmentation methods for larvae body parts. Additionally, there is room for optimizing hyperparameters and training techniques for the deep learning model, exploring different model architectures and backbones. Lastly, the developed dataset could be enriched in terms of material classification, quantity, variety, and available annotations.

The research in the paper still segregates the tasks of detection and classification, both of which naturally require substantial resources. However, this paper does not incorporate measurements for time consumption and computational capabilities. It is hoped that future research will be able to contribute by quantifying time consumption and assessing computational capabilities effectively.

REFERENCES

[1] M. R. Karyanti *et al.*, "The changing incidence of Dengue Haemorrhagic Fever in Indonesia: A 45-year registry-based analysis," *BMC Infect. Dis.*, vol. 14, no. 1, pp. 1–7, 2014, doi: 10.1186/1471-2334-14-412.

[2] Health Depatment Sumatera Utara, *Health Profile of Sumatra Utara Province 2019*. Dinas Kesehatan Sumatera Utara, 2019.

[3] Health Depatment Sumatera Utara, *Health Profile of Sumatra Utara Province 2022*. Dinas Kesehatan Sumatera Utara, 2022.

[4] Directorate General of Disease Prevention and Control, *Technical Guidelines for the Implementation of Mosquito Breeding Site Eradication*. Kementerian Kesehatan Republik Indonesia, 2016.

[5] Ministry of Health Indonesia, *Guidelines for Prevention and Control of Dengue Fever in Indonesia*, vol. 5. 2017. [Online]. Available: https://drive.google.com/file/d/1IATZEcgGX3x3BcVUcO_l8Yu9B5REKOKE/view

[6] Ministry of Health Indonesia, *Indonesia Health Profile 2021*, Indonesia. 2022.

[7] P. I. Nainggolan *et al.*, "Classification Of Aedes Mosquito Larva Using Convolutional Neural Networks And Extreme Learning Machine," *Proceeding - ELTICOM 2023 7th Int. Conf. Electr. Telecommun. Comput. Eng. Sustain. Resilient Communities with Smart Technol.*, pp. 79–83, 2023, doi: 10.1109/ELTICOM61905.2023.10443125.

[8] P. Kukieattikool *et al.*, "Improvements to the aedes larvae mobile detection system," *Maejo Int. J. Sci. Technol.*, vol. 14, no. 2, pp. 195–208, 2020.

[9] A. Sanchez-Ortiz *et al.*, "Mosquito larva classification method based on convolutional neural networks," *2017 Int. Conf. Electron. Commun. Comput. CONIELECOMP 2017*, 2017, doi: 10.1109/CONIELECOMP.2017.7891835.

[10] R. M. Martins, B. M. Espíndola, P. P. Araujo, C. G. von Wangenheim, C. J. de Carvalho Pinto, and G. Caminha, "Development of a Deep Learning Model for the Classification of Mosquito Larvae Images," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 14197 LNAI, no. October, pp. 129–145, 2023, doi: 10.1007/978-3-031-45392-2_9.

[11] M. I. A. B. Z. Azman and A. B. Sarlan, "Aedes Larvae Classification and Detection (ALCD) System by Using Deep Learning," *2020 Int. Conf. Comput. Intell. ICCI 2020*, no. October, pp. 179–184, 2020, doi: 10.1109/ICCI51257.2020.9247647.

[12] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 386–397, 2020, doi: 10.1109/TPAMI.2018.2844175.

[13] A. Kirillov, R. Girshick, K. He, and P. Dollar, "Panoptic feature pyramid networks," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2019-June, pp. 6392–6401, 2019, doi: 10.1109/CVPR.2019.00656.

[14] D. A. Keim, F. Mansmann, D. Oelke, and H. Ziegler, "Visual analytics: Combining automated discovery with interactive visualizations," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 5255 LNAI, pp. 2–14, 2008, doi: 10.1007/978-3-540-88411-8_2.

[15] J. Kielman, J. Thomas, and R. May, "Introduction: Foundations and frontiers in visual analytics," *Inf. Vis.*, vol. 8, no. 4, pp. 239–246, 2009, doi: 10.1057/ivs.2009.25.

[16] C. Chen, H. Hou, Z. Hu, and S. Liu, "An Illuminated Path: The Impact of the Work of Jim Thomas," *Expand. Front. Vis. Anal. Vis.*, pp. 9–30, 2012, doi: 10.1007/978-1-4471-2804-5_2.

[17] D. Keim, F. Mansmann, J. Schneidewind, J. Thomas, and H. Ziegler, "Visual Analytics: Scope and Challenges Visual Data Mining," *Vis. Data Min.*, vol. 4404, pp. 76–90, 2008.

[18] D. De Geus, P. Meletis, and G. Dubbelman, "Fast panoptic segmentation network," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 1742–1749, 2020, doi: 10.1109/LRA.2020.2969919.

[19] D. S. Prasvita, D. Chahyati, and A. M. Arymurthy, "Automatic Detection of Oil Palm Growth Rate Status with YOLOv5," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 3, pp. 529–537, 2023, doi: 10.14569/IJACSA.2023.0140361.

[20] P. I. Nainggolan, D. I. Purba, R. F. Rahmat, S. Lubis, R. Anugrahwaty, and M. F. S. Putra, "Classification of Normal Lung Image, Bronchitis, And Tuberculosis Using Extreme Learning Machine," in *AIP Conference Proceedings*, 2024, vol. 2987, no. 1, pp. 1–6. doi: 10.1063/5.0200319.

[21] T. Lin, C. L. Zitnick, and P. Doll, "Microsoft COCO : Common Objects in Context," in *In European Conference on Computer Vision*, 2014, pp. 740–755.