**Journal of Applied Informatics and Computing (JAIC)**
Vol.8, No.2, December 2024, pp. 463~471
e-ISSN: 2548-6861


# Real-time Detection Transformer (RT-DETR) of Ornamental Fish Diseases with YOLOv9 using CNN (Convolutional Neural Network) Algorithm

**Dwi Nurul Huda [1]*, Mochammad Rizki Romdoni [2]**, Liza Safitri [3]**, Ade Winarni [4]**, Abdur Rahman [5]****
* Sistem Informasi, Sekolah Tinggi Teknologi Indonesia Tanjung Pinang
** Teknik Informatika, Sekolah Tinggi Teknologi Indonesia Tanjung Pinang
dwi.nurulhuda@gmail.com [1], rizki@sttindonesia.ac.id [2], savetree3300@gmail.com [3], adewina16@gmail.com [4], abdur.rahmanux@gmail.com [5]

**Article Info**

***Article history:***

Received 2024-09-24
Revised 2024-10-14
Accepted 2024-10-21

***Keyword:***

*CNN,*
*Disease Detection,*
*Real-time Detection Transformer (RT-DETR),*
*Soft Voting Ensemble Learning,*
*YOLOv9.*

**ABSTRACT**

The lack of specialized tools to check the condition of ornamental fish has hindered effective management. This research proposes a novel software architecture that uses the YOLOv9 model combined with RT-DETR to enable accurate and timely identification of ornamental fish conditions including fish diseases, empowering farmers and hobbyists with a valuable resource. This integration is done using Soft Voting Ensemble Learning technique. To achieve this goal, an Android mobile application successfully classified healthy fish and accurately identified common diseases such as bacteria, fungal, parasitic, and whitetail. Based on the test results, the integration accuracy of the YOLOv9 and RT-DETR models produced a high result of 0.8947 while the stand-alone YOLOv9 showed 0.8889 and the stand-alone RT-DETR of 0.8904. Recommendations are given for the combination of YOLOv9 and RT-DETR in condition detection and diagnosis of ornamental fish diseases.



## I. INTRODUCTION

The ornamental fish business has surged since the COVID-19 pandemic by up to 70%[1]. In 2022 alone, Indonesia has become the world's second-largest ornamental fish exporting country, with export value reaching USD 36.4 million[2]. According to data from the Ministry of Maritime Affairs and Fisheries in the first quarter of 2023, Indonesia's ornamental fish exports rose 16.2% compared to the same period in the previous year[3]. This indicates that ornamental fish have a high potential value in business, so it is important to maintain their health so that their survival is maintained. Ornamental fish is a type of fish that has a marine or freshwater habitat whose function is to beautify the house/room. Cultivating ornamental fish or becoming an ornamental fish enthusiast is a hobby for some people, not only adding economic value but also reducing stress. Ornamental fish can be enjoyed when their color, size, and growth are healthy. Unhealthy ornamental fish will make the appearance no longer attractive such as changing the color of the fish, the movement is not as agile as usual, to the transmission of disease to other ornamental fish until it results in death. Unhealthy ornamental fish can be caused by several factors such as disease. One of the problems in fish farming is when fish contract disease. Some ornamental diseases that are often encountered are caused by bacteria or fungi. Based on the results of interviews collected, it was found that the lack of knowledge of cultivators and ornamental fish enthusiasts about sick ornamental fish and in diagnosing diseases resulted in late countermeasures.

Currently, the way to identify unhealthy ornamental fish is still done conventionally, such as looking at abnormal fish behavior. However, since the light from the air into the water is refracted, it is difficult for the human eye to observe the health of the fish, causing failure to administer medicine on time or missing the best treatment period, resulting in huge economic losses. This is a technical bottleneck in ornamental fish farming. Therefore, to help ornamental fish farmers and enthusiasts, it is important to develop an automatic fish disease identification and analysis method.

Most of the previous research related to YOLO is to make changes to the network structure, which aims to adjust to the area of research, for example, such as proposing a disease detection method in fish farms using YOLO v5 which has improved the network structure, namely replacing CSPNet to C3 to make it simpler and replacing the conventional kernel size from 3x3 to Convolutional Kernel Group (CKG) The result of this improvisation is speed and accuracy of 99.75% compared to the old YOLO model[4].

Integrating Yolo v8 with the Real-Time Detection Transformer (RT-DETR) [5] obtained promising results for detecting fish in turbid waters with a mAP (mean Average Precision) value of 0.971 compared to the standard Yolo v8 mAP value of 0.912. RT-DETR offers possible capabilities that YOLO models do not have[6]. RT-DETR also demonstrates effective ship detection accuracy and robustness in complex marine environments[7]. Fish behavior anomaly detection using deep sort algorithm with Yolo v8[8] which can extract information related to abnormal fish behavior such as swimming direction and fish sleeping time with mAP value of 73.49 compared to Yolov8 before modification of 71.36. CNN model was successfully used to detect fresh fish[9].

We propose a novel approach to check the healthy or sick condition of ornamental fish and diagnose the disease using a combination of YOLOv9 and RT-DETR models, both of which use CNN to extract important features from images. We combine advanced object detection models, including RT-DETR and YOLOv9, to identify and classify fish diseases accurately. To further improve performance, we used Soft Voting Ensemble Learning, which combines RT-DETR and YOLOv9 predictions for more robust results. We evaluated the overall performance of these models using four key metrics: accuracy, precision, recall, and F1-score.

## II. Methods

Research in fish disease detection has increasingly utilized deep learning algorithms. [10] employed a CNN architecture customized with ResNet-50, achieving an accuracy of 87.46%. YOLOv9 modeling combined with RT-DETR on the detection of ornamental fish conditions and diseases was built to help farmers and ornamental fish enthusiasts check the health conditions of ornamental fish with improved accuracy better than using only one model independently. The combination of models uses soft voting ensemble learning techniques with testing parameters using accuracy, recall, precision, and F1-Score.

### A. *Flowchart Development*

In the development of ornamental fish condition detection applications along with disease diagnosis, there are several stages. Based on Figure 1, the process can be explained as follows. The first step involves collecting datasets of ornamental fish, both sick and healthy, with a focus on two species: Koi Fish (Cyprinus rubrofuscus) and Gold Fish (Carassius auratus). In the data preprocessing stage, images are filtered to retain only those of good quality, while poor-quality images are discarded. Following this, the data annotation process labels the fish as either healthy or diseased.

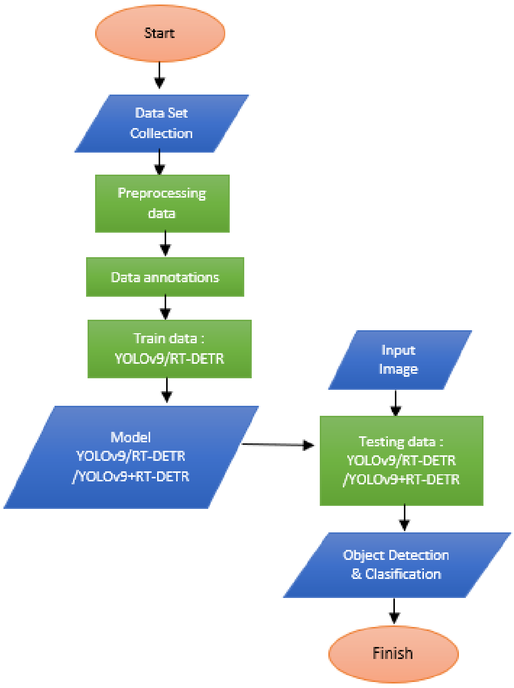


Figure 1. Flowchart Development YOLOv9 and RT-DETR

The training stage then uses the annotated data to train and validate YOLOv9 and RT-DETR models, with assistance from Kaggle (https://www.kaggle.com/). The Soft Voting Ensemble Learning technique is applied when using YOLOv9 + RT-DETR, yielding models that will be utilized in the testing phase. During testing, images are entered and evaluated using YOLOv9, RT-DETR, or the combined YOLOv9 + RT-DETR models based on the previously trained data. The final output provides detection of the ornamental fish's health condition.

### B. *YOLOv9*

YOLOv9 is one of the models used in object detection. YOLOv9 has several updates from the previous version, YOLOv8, such as the use of modules that are wider than YOLOv9 than the previous version, thus allowing a more complex architecture. In addition, based on the comparison of YOLOv9 Modeling combined with RT-DETR on the detection of ornamental fish conditions and diseases, it was built with the aim of helping farmers and ornamental fish enthusiasts to check the health conditions of ornamental fish with a better increase in accuracy than using only one model independently. Model combination using soft voting ensemble learning technique with testing parameters using accuracy, recall, precision, and F1-Score.

Performance assessment [11] using 3 parameters namely precision, recall, and mAP found that the application of YOLOv9 has a better value than the previous version (YOLOv8) even though it requires more memory. YOLOv9 has the advantage of completing tasks at high speed[12].

### C. RT-DETR

Similar to YOLO, RT-DETR is a model commonly used in object detection. In blurry image detection, RT-DETR is one of the solutions that can be used. RT-DETR has higher accuracy results compared to YOLOv9 [13].

### D. Soft Voting Ensemble Learning

Ensemble Learning is a technique of combining the results of multiple detection models, with the aim of improving overall accuracy. This approach is done by soft voting. Soft voting is the probability of the confidence value for each class. Classes that have low confidence values will not be displayed and replaced with higher confidence values to determine the final prediction value. Theoretically and empirically, ensemble learning methods have better performance results compared to single learners such as when handling cases that have high complexity such as classification problems[14]. In ensemble learning the models will be executed independently and then the results will be combined, allowing the advantages of each model to be utilized simultaneously. The Ensemble Learning method can increase the power of both models, with the aim of increasing the effectiveness and efficiency of the application, this model also shows improvement compared to the standalone model[15].

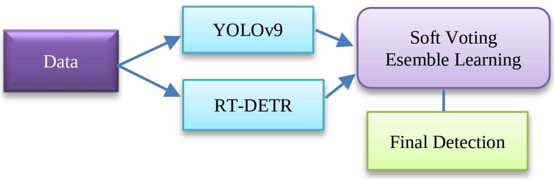


Figure 2. Illustration of Soft Voting Ensemble Learning

Based on Figure 2, the process is as follows. First, an input image—either captured from a camera or selected from a cellphone gallery—is prepared for processing using the YOLOv9 and RT-DETR models. The image is then processed by both models, a step that takes longer than using a single model. Each model generates predictions, identifying bounding boxes in the image to classify whether the fish is healthy or affected by disease. Once both models have produced results, a soft voting process is applied, selecting the prediction with the higher confidence value and discarding any with lower confidence. Finally, the results of the soft voting process are displayed, showing the final prediction for the image based on the combined use of both models.

Our proposed methods involves simultaneously running the YOLOv9 and RT-DETR models on input images. The predictions from both models are then compared, and the one with the highest confidence score is selected as the final output. This strategy aims to leverage the complementary strengths of the individual models, potentially improving overall accuracy and robustness.

### E. Performance Testing

Performance metrics are used to validate the effectiveness of the pre-trained model assessment. Some parameters to assess performance include accuracy, precision, recall, and F1-Score. Validation of the training data model can be assessed with performance metrics from the confusion matrix. The confusion matrix table will represent the results based on TP (true positive), FP (false positive), TN (true negative), and FN (false negative) [16].

Accuracy will measure the accuracy of a model in classifying correctly. The accuracy calculation formula [17]:

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \quad .......(1)$$

Precision will measure the accuracy of model predictions based on measuring the percentage of correct predictions of all predictions made [18].

$$\text{Precision} = \frac{TP}{TP+FP} \quad .......(2)$$

Recall is known as True Positive Rate (TPR). Recall will express the rate of correctly classified positive samples. Recall is calculated as the ratio between the classification of positive samples and all samples assigned to the positive class [19].

$$\text{Recall} = \frac{TP}{TP+FN} \quad .......(3)$$

F1-Score, also known as F-Score and F-Measure, is a metric used to evaluate the performance of machine learning models. F1-Score Formula [20] :

$$\text{F1-Score} = \frac{(\beta^2+1)*precision*recall}{\beta^2*(precision+recall)}\text{, where } \boldsymbol{\beta}=1 \quad .......(4)$$

## III. Results And Discussion

This research produces a system architecture of the program to be built, data set processing, data annotation, and training data to produce model performance metric testing to see the comparison between models that stand independently with models combined using soft voting ensemble learning techniques.

### A. System Architecture

Based on Figure 3, the process is as follows. The aquarium containing ornamental fish serves as the detection object. A cell phone captures images through either the camera or gallery. By implementing TensorFlow Lite, the cell phone camera can detect objects, displaying bounding boxes to indicate healthy fish or signs of disease. Images taken from the camera or gallery are then sent via an API endpoint to Cloud Run, where the image detection process takes place.

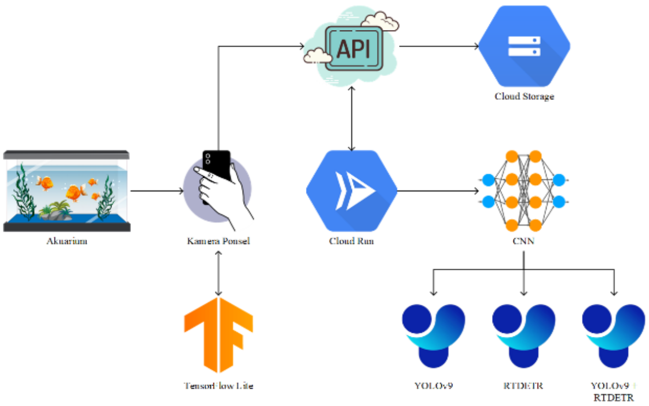


Figure 3. System Architecture

Cloud Run processes the image using a classification engine based on Convolutional Neural Networks to identify fish diseases. If YOLOv9 is selected for detection, the image is processed with the YOLOv9 training model. If RT-DETR is chosen, the RT-DETR model handles the detection. For a combined YOLOv9 + RT-DETR detection, the image is processed using Ensemble Learning with both models. The processed image, now marked with bounding boxes to indicate detection results, is finally stored in Cloud Storage for access and review.

## B. *Data Collection*

The collection of ornamental fish datasets was carried out using the help of kaggle (https://www.kaggle.com/), roboflow (https://roboflow.com/) as well as from image captures made directly by researchers. The initial data results that were successfully collected amounted to 32479 images.

Figure 4. Collecting data

## C. *Data Preprocessing*

Based on the results of image sorting/elimination, the data collected is 1035 data which is divided into 80% training data, 10% valid data, and 10% testing data. Our dataset consists of 640x640 images depicting a variety of fish diseases, including bacterial, fungal, parasitic, and white tail diseases. Bacterial diseases are caused by bacteria, while fungal diseases are caused by fungi. Parasitic diseases are caused by external or internal parasites, and white tail diseases are characterized by damage to the tail fin, often resulting from bacterial or viral infections. These diseases can be influenced by factors such as water quality and fish health.

Data division is divided into 3 parts, namely 828 training data, 107 valid data, and 100 testing data. The distribution of ornamental fish is divided into 5 classes, namely:

TABLE I
ORNAMENTAL FISH CLASS DISTRIBUTION

| No | Class | Data Training | Data Valid | Data Testing |
|---|---|---|---|---|
| 1 | Healthy Fish | 467 | 59 | 58 |
| 2 | Bacterial Diseases | 147 | 19 | 18 |
| 3 | Fungal Diseases | 72 | 10 | 8 |
| 4 | Parasitic Diseases | 44 | 6 | 5 |
| 5 | White Tail Diseases | 98 | 13 | 11 |
| | **Total** | 828 | 107 | 100 |

To enhance model robustness, we employed data augmentation techniques during dataset creation. By setting the brightness value to -25%, we simulated low-light conditions, enabling the model to recognize objects in challenging lighting environments. Additionally, a blur setting of 2.5 pixels was applied to introduce image blur, improving the model's ability to handle out-of-focus images. Finally, noise was added up to 1.5% of the image size to train the model to be more resilient to noise-corrupted images.

## D. *Data Annotation*

Labeling is done using roboflow (https://roboflow.com/). This process will perform labeling of healthy ornamental fish and labeling of ornamental fish diseases.

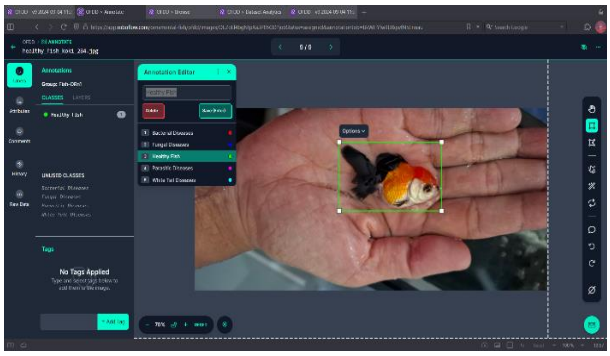

Figure 5. Annotation Data

TABEL II
ORNAMENTAL FISH CLASSIFICATION

| No | Class | Feature |
|---|---|---|
| 1 | Healthy Fish | a. Uniform color of fish skin and fins.<br>b. No wounds or spots on the fish's body.<br>c. Normal fish posture and intact fins. |
| 2 | Bacterial Diseases | a. Open wounds on the fish skin.<br>b. Red spots and swelling on the fish's body.<br>c. There is a change in color on the fish's body. |
| 3 | Fungal Diseases | a. Fungal growth on fish scales<br>b. There is a clear white color around the infected area. |
| 4 | Parasitic Diseases | a. There is irritation on the fish scales.<br>b. There are small black spots. |
| 5 | White Tail Diseases | a. Damage to the tail of the fish.<br>b. Whitening of the tip of the tail or fins. |

### *E. Training Data YOLOv9*

Based on Figure 6, the metrics can be explained as follows. The metrics train/box_loss, train/cls_loss, train/dfl_loss, val/box_loss, val/cls_loss, and val/dfl_loss serve as essential indicators in the training process by measuring model performance progress.

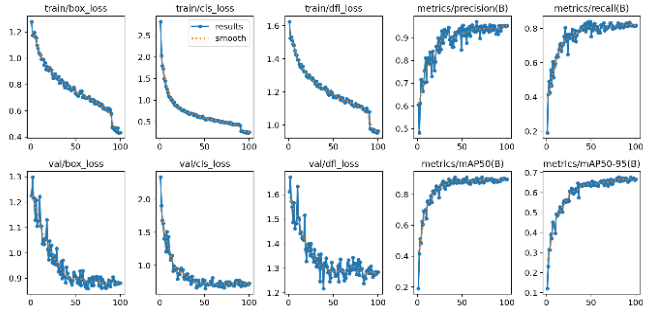


Figure 6. YOLOv9 Training Result

A downward trend in these curves suggests that the model is improving and becoming more accurate. The metrics/precision(B) metric assesses the accuracy of the model's positive predictions relative to the total number of positive predictions. A steadily increasing curve approaching a maximum of 1 indicates that the model is producing more precise positive predictions as training continues. The metrics/recall(B) measures the model's capability to detect all correct instances, with an increasing curve that indicates enhanced object detection accuracy. Metrics/mAP50(B) represents the mean Average Precision at an IoU threshold of 0.5, and its steady increase signals that the model is gaining accuracy in detecting objects across a wider range. Lastly, metrics/mAP50-95(B) calculates the mAP across multiple IoU thresholds (from 0.5 to 0.95). This curve approaches 0.7, showing improved model performance across varying IoU thresholds, signifying robust detection accuracy.

### *F. Training Data RT-DETR*

Based on Figure 7, the metrics can be interpreted as follows. The metrics train/giou_loss, train/cls_loss, train/l1_loss, val/giou_loss, val/cls_loss, and val/l1_loss serve as essential indicators throughout the training process, measuring the model's progress.

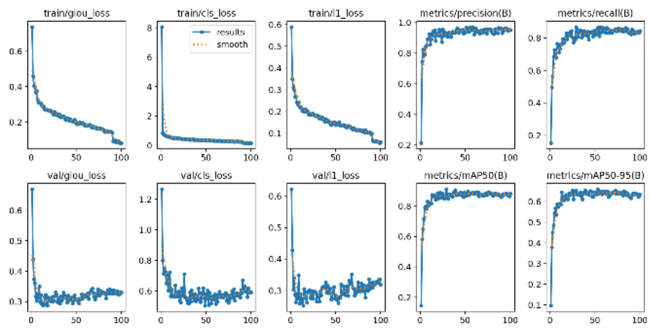


Figure 7. RT-DETR Training Result

A significant downward trend in these curves suggests improvement in the model's accuracy, as it continually learns to make better predictions and reduce errors. While minor fluctuations are observed at certain points in the validation data, this is typical, as validation data tends to be more diverse than training data. The metrics/precision(B) metric reflects the accuracy of the model's positive predictions relative to the total positive predictions, with a steady increase in the curve nearing 1, indicating enhanced precision over time. The metrics/recall(B) measures the model's ability to detect all relevant examples, with an increase toward 0.9, showing that the model is detecting nearly all objects in the dataset. The metrics/mAP50(B) metric, which assesses the mean Average Precision at a 0.5 IoU threshold, displays high stability, indicating consistent accuracy in predictions. Finally, the metrics/mAP50-95(B), which averages the mAP across multiple IoU thresholds (from 0.5 to 0.95), increases gradually to about 0.6, demonstrating reliable model performance across varying overlap levels and confirming strong overall accuracy.

### *G. Result of Testing (Detection and Classification)*

The following are the test results of an android-based application called FishDeas with detection using Yolov9, RT-DETR and YOLOv9 combined RT-DETR.

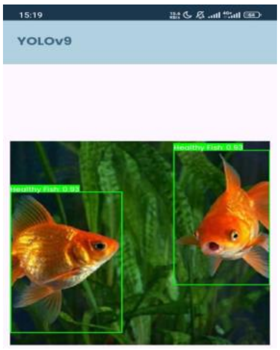


Figure 8.a Healthy Fishes


Gambar 8.b Disease Fish


Figure 9.a Healthy Fishes


Figure 9.b Disease Fish

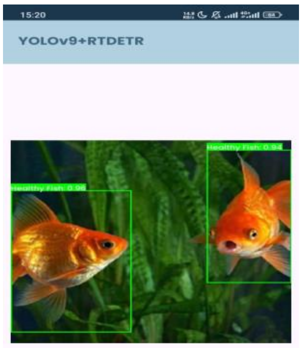


Figure 10.a Healthy Fishes


Figure 10.b Disease Fish

FishDeas successfully detects healthy fish and disease fish as shown in Figure 8.a, 8.b, 9.a, 9.b and 10.a, 10.b. From each of these images, there is a confidence value as in table III.

TABLE III
COMPARISON OF CONFIDENCE VALUES

| No | Model | Nilai Confidence Healthy Fish | Nilai Confidence Unhealthy Fish |
|---|---|---|---|
| 1 | YOLOv9 | 0.93 ; 0.93 | 0.90 |
| 2 | RT-DETR | 0.96 ; 0.94 | 0.95 |
| 3 | YOLOv9+RT-DETR (Soft Voting Ensemble Learning) | 0.96 ; 0.94 | 0.95 |

The confidence value is a value that shows the model's confidence in the detected model. Soft Voting Ensemble Learning applied to YOLOv9 + RT-DETR will follow the highest value between YOLOv9 and RT-DETR which stands independently. In Table III, the highest confidence value of healthy fish between the YOLOv9 and RT-DETR models that stand independently is in the RT-DETR model, namely 0.96; 0.94 and the highest confidence value of sick fish is in the RT-DETR model data, namely 0.95, so the YOLOv9 + RT-DETR model will take the RT-DETR value.

## H. Accuracy

Accuracy is a metric used in modeling that serves to see the accuracy of a model used because it is related to classification performance. The following will present the calculation of accuracy in the YOLOv9, RT-DETR, and YOLOv9 + RT-DETR models.

Table IV shows that the use of the YOLOv9 model combined with RT-DETR has the highest accuracy results compared to the use of the YOLOv9 model and RT-DETR standing alone.

TABLE IV
COMPARISON OF ACCURACY VALUES

| No | Model | Calculation | Accuracy |
|---|---|---|---|
| 1 | YOLOv9 | $= \frac{TP+TN}{TP+FP+FN+TN}$ $= \frac{128+0}{128+8+8+0}$ $= \frac{128}{144}$ | 0.8889 |
| 2 | RT-DETR | $= \frac{TP+TN}{TP+FP+FN+TN}$ $= \frac{130+0}{130+9+7+0}$ $= \frac{130}{146}$ | 0.8904 |
| 3 | YOLOv9 + RT-DETR | $= \frac{TP+TN}{TP+FP+FN+TN}$ $= \frac{136+0}{136+11+5+0}$ $= \frac{136}{152}$ | 0.8947 |

## I. Model Performance Evaluation

This section describes the evaluation using performance metrics such as Precision, Recall, and F1-Score of YOLOv9, RT-DETR and YOLOv9 + RT-DETR which can be seen in Table 1.

### 1) Precision

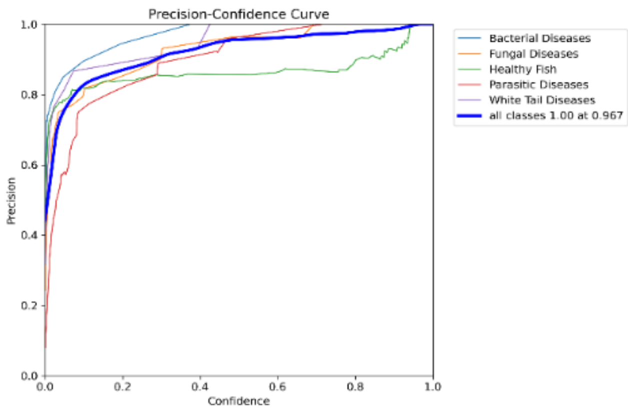


Figure 11. Precision Chart YOLOv9 Model

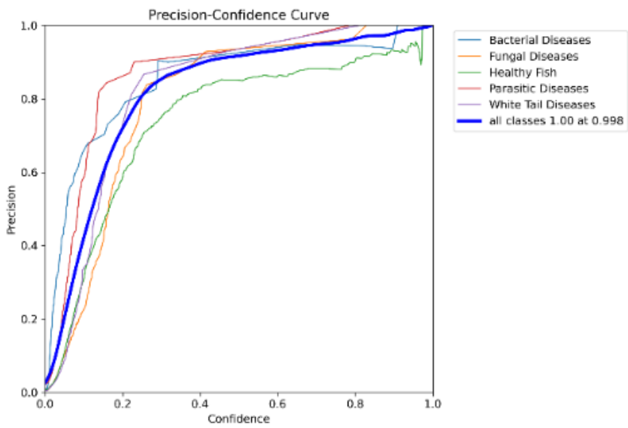


Figure 12. Precission Chart RT-DETR Model

Figure 11 and Figure 12 charts show how the difference in model precision between YOLOv9 and RT-DETR, overall it can be concluded that RT-DETR looks slightly better than YOLOv9 judging from the precision values obtained for all classes, with different confidence values, namely 0.998 in RT-DETR and 0.967 in YOLOv9.

*2) Recall*

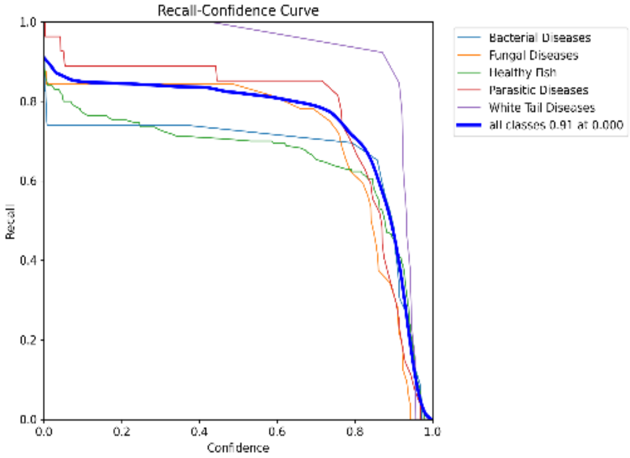


Figure 13. Recall Precision YOLOv9 Model

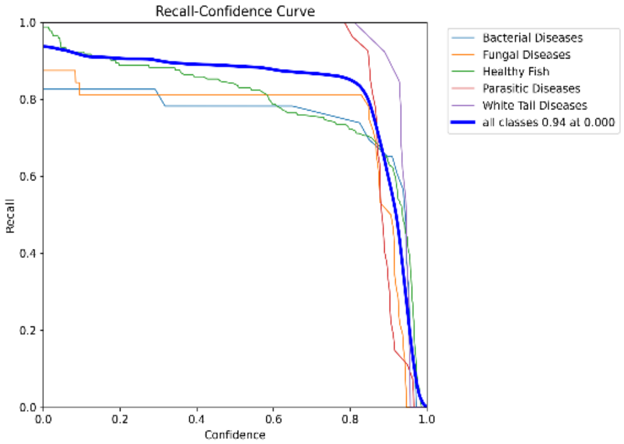


Figure 14. Recall Chart RT-DETR Model

Figure 13 and Figure 14 charts show that RT-DETR has a more stable and higher recall performance than YOLOv9, with the initial recall reaching 0.94 and maintaining it above 0.8 for longer. YOLOv9, on the other hand, starts with a recall of 0.91 and experiences a faster decline as confidence increases. Overall, RT-DETR is superior in maintaining performance at low to medium confidence, while YOLOv9 experiences a faster decline in recall.

*3) F1-Score*

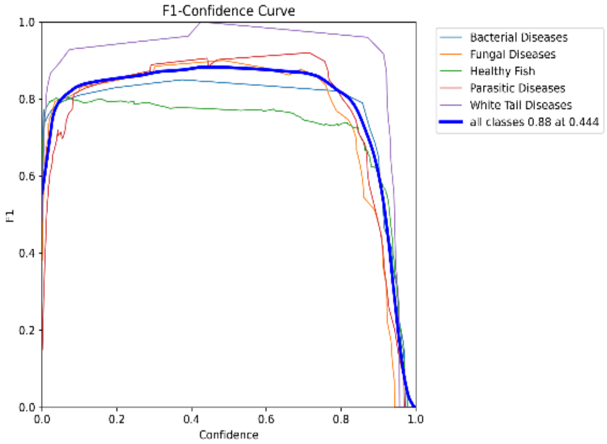


Figure 15. F1-Score Chart YOLOv9 Model

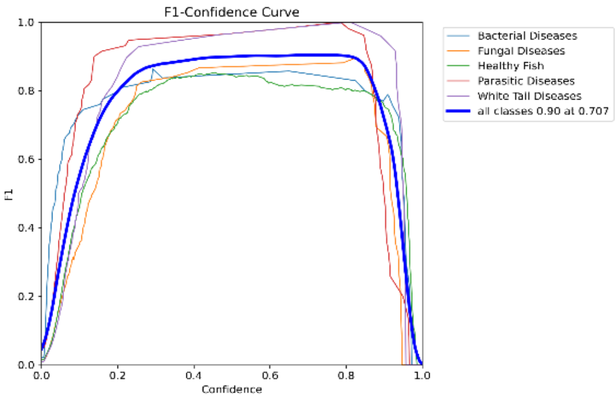


Figure 16. F1-Score Model RT-DETR Chart

Figure 15 and Figure 16 show that YOLOv9 has a slightly lower F1 score performance than RT-DETR. Both models show the same trend of increasing F1 as confidence increases, but RT-DETR tends to have a more stable curve and higher F1 peak, reflecting better generalization ability. The RT-DETR model also shows less fluctuation at various confidence levels, making it more reliable under more variable conditions.

#### 4) *Confusion Matrix*

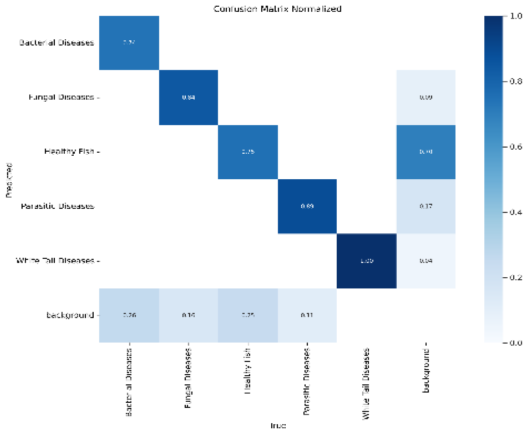


Figure 17. Confusion Matrix Evalution YOLOv9 Model

Figures 17 and 18 show lower accuracy in some classes, such as Bacterial Diseases (0.74) and Healthy Fish (0.75), which use the YOLOv9 model with a higher error rate in predicting the background. RT-DETR, on the other hand, showed improved accuracy in these classes, such as Bacterial Diseases (0.83) and Healthy Fish (0.89), as well as reduced error in background prediction. Overall, RT-DETR performed better with more accurate predictions on most classes than YOLOv9.

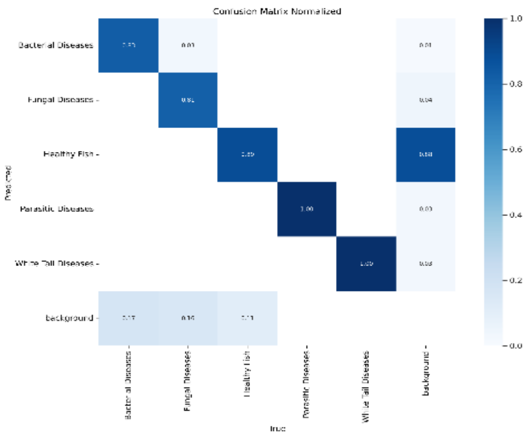


Figure 18. Confusion Matrix Evalution RT-DETR Model

#### 5) *Model Evaluation Result*

Model evaluation results include metrics such as mAP (mean Average Precision) for standalone YOLOv9 and RT-DETR models (Table V and Table VI) as well as recall, precision, and F1-Score metrics that provide a complete picture of the model's ability to detect Ornamental Fish Disease objects in YOLOv9, RT-DETR, and YOLOv9+RT-DETR (Table VII).

TABLE V
RESULT MODEL EVALUATION YOLOV9

| Class Name | Precision | Recall | mAP50 | Accuracy Overall Value (mAP50) |
|---|---|---|---|---|
| Healthy Fish | 0.857 | 0.706 | 0.774 | 88.2% |
| Bacterial Diseases | 1 | 0.73 | 0.844 | |
| Fungal Diseases | 0.96 | 0.844 | 0.898 | |
| Parasitic Diseases | 0.947 | 0.852 | 0.897 | |
| White Tail Diseases | 1 | 0.994 | 0.997 | |

TABLE VI
RESULT MODEL EVALUATION RT-DETR

| Class Name | Precision | Recall | mAP50 | Accuracy Overall Value (mAP50) |
|---|---|---|---|---|
| Healthy Fish | 0.882 | 0.746 | 0.808 | 90.32% |
| Bacterial Diseases | 0.946 | 0.761 | 0.843 | |
| Fungal Diseases | 0.958 | 0.812 | 0.879 | |
| Parasitic Diseases | 0.988 | 1 | 0.994 | |
| White Tail Diseases | 0.985 | 1 | 0.992 | |

Table V and Table VI show the performance calculation of each model The RT-DETR model has the highest value in precision which states that the model rarely classifies.

TABLE VII
RESULT MODEL PERFORMANCE EVALUATION

| No | Model | Akurasi | Precision | Recall | F1-Score |
|---|---|---|---|---|---|
| 1 | YOLOv9 | 0.8889 | 0.9412 | 0.9412 | 0.9412 |
| 2 | RT-DETR | 0.8904 | 0.9353 | 0.9489 | 0.9420 |
| 3 | YOLOv9 + RT-DETR | 0.8947 | 0.9252 | 0.9645 | 0.9444 |

Table VII provides a performance comparison for each model, as follows. The combined model of YOLOv9 and RT-DETR achieves the highest accuracy, indicating its superior effectiveness in detecting and classifying the condition of ornamental fish compared to the standalone YOLOv9 and RT-DETR models. The YOLOv9 model demonstrates the highest precision, suggesting that it seldom misclassifies unhealthy fish as healthy, though it may overlook some healthy fish detections. In terms of recall, the combination model excels, capturing nearly all healthy fish but potentially categorizing some unhealthy fish as healthy. Furthermore, the combination model also achieves the highest F1-Score, reflecting its well-balanced performance in accurately detecting all healthy fish while minimizing the misclassification of unhealthy fish as healthy.

We conducted experiments in a controlled environment using two aquariums, one with turbid water and the other with clear water. Videos, each 28 seconds long, were captured using a Samsung A55 smartphone equipped with an Exynos 1480 processor. The results, as presented in Tables IV and V, demonstrate that the RT-DETR + YOLOv9 model achieved promising performance with a confidence value of 0.88 in turbid water with a latency of 291 milliseconds.

TABLE VIII
TEST RESULT ON TURBID WATER

| Model | Start Time Detected | Confident | FPS | Latency |
|---|---|---|---|---|
| YOLOv9 | 00.01 | 0.79 | 0.79 | 355 |
| RT-DETR | 00.01 | 0.88 | 0.4 | 358 |
| YOLOv9 + RT-DETR | 00.02 | 0.88 | 1.1 | 291 |

In clear water, the RT-DETR + YOLOv9 model achieved a confident value 0.78 with a latency of 360 milliseconds.

TABLE IX
TEST RESULT ON CLEAR WATER

| Model | Start Time Detected | Confident | FPS | Latency |
|---|---|---|---|---|
| YOLOv9 | 00.00 | 0.77 | 0.6 | 355 |
| RT-DETR | 00.02 | 0.79 | 0.3 | 368 |
| YOLOv9 + RT-DETR | 00.02 | 0.78 | 0.3 | 350 |

## IV. CONCLUSION

This research found that the detection of ornamental fish conditions and disease diagnosis is better using the YOLOv9 model combined with RT-DETR using a soft voting ensemble learning technique. The results showed that the accuracy of YOLOv9 and RT-DETR combined had the best value of 0.8947 compared to the standalone YOLOv9 and RT-DETR models. Likewise, for the results of the calculation of performance metrics on the recall and F1-Score metrics, the YOLOv9 model combined with RT-DETR produces better values but not statistically significant even though the precision metric is still less good than the YOLOv9 model. The combination of YOLOv9 and RT-DETR model will be better used in the condition of fish in turbid water because it produces high FPS value and low latency.

## ACKNOWLEDGMENT

We are very grateful for the support provided by LPPM Sekolah Tinggi Teknologi Indonesia Tanjung Pinang and we express our gratitude and highest appreciation to the Ministry of Education and Culture for funding this research to completion. This research is supported by the Fiscal Year 2024 DRPTM Research Grant, Diktiristek, KEMDIKBUDRISTEK.

## REFERENCES

[1] E. Setiawan, R. F. Suwarman, A. Firmansyah, and M. D. Saputra, ‘Pelatihan Budidaya Dan Pemasaran Ikan Cupang Hias Untuk Mengatasi Masalah Ekonomi Di Masa Pandemi’, *J. Abdimas Sang Buana*, vol. 2, no. 1, p. 20, 2021, doi: 10.32897/abdimasusb.v2i1.511.

[2] RRI, ‘Indonesia Negara Kedua Pengekspor Ikan Hias Terbesar’. [Online]. Available: https://www.rri.co.id/bisnis/418692/indonesia-negara-kedua-pengekspor-ikan-hias-terbesar

[3] Administrator, ‘Buah Manis Kebijakan Ekonomi Biru’. [Online]. Available: https://indonesia.go.id/kategori/editorial/7752/buah-manis-kebijakan-ekonomi-biru?lang=1#:~:text=Semester I-2023 nilai ekspor,menggeser posisi Singapura dan Belanda.

[4] Z. Wang, H. Liu, G. Zhang, X. Yang, L. Wen, and W. Zhao, ‘Diseased Fish Detection in the Underwater Environment Using an Improved YOLOV5 Network for Intensive Aquaculture’, *Fishes*, vol. 8, no. 3, 2023, doi: 10.3390/fishes8030169.

[5] E. Li, Q. Wang, J. Zhang, W. Zhang, H. Mo, and Y. Wu, ‘Fish Detection under Occlusion Using Modified You Only Look Once v8 Integrating Real-Time Detection Transformer Features’, *Appl. Sci. Switz.*, vol. 13, no. 23, 2023, doi: 10.3390/app132312645.

[6] Y. Zhao *et al.*, ‘DETRs Beat YOLOs on Real-time Object Detection’, 2023.

[7] Z. Xing, J. Ren, X. Fan, and Y. Zhang, ‘S-DETR: A Transformer Model for Real-Time Detection of Marine Ships’, *J. Mar. Sci. Eng.*, vol. 11, no. 4, 2023, doi: 10.3390/jmse11040696.

[8] Z. Zhao, ‘Abnormal Behavior Fish and Population Detection Method based on Deep Learning’, vol. 4, no. 3, pp. 2–6, 2023.

[9] R. Mahendran and G. P. Seneviratne, ‘Using CNN to Identify the Condition of Edible Fish’, *7th Int. Conf. Inf. Technol. Res. Digit. Resil. Reinvention ICITR 2022 - Proc.*, no. December 2022, 2022, doi: 10.1109/ICITR57877.2022.9993419.

[10] S. Dash, S. Ojha, R. K. Muduli, S. P. Patra, and R. C. Barik, ‘Fish Type and Disease Classification Using Deep Learning Model Based Customized CNN with Resnet 50 Technique’, *J. Adv. Zool.*, Mar. 2024, doi: 10.53555/jaz.v45i3.4194.

[11] H. M. Zayani *et al.*, ‘International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING Unveiling the Potential of YOLOv9 through Comparison with YOLOv8’, *Orig. Res. Pap. Int. J. Intell. Syst. Appl. Eng. IJISAE*, vol. 2024, no. 3, pp. 2845–2854, 2024.

[12] R. Sapkota *et al.*, ‘Comprehensive Performance Evaluation of YOLOv10, YOLOv9 and YOLOv8 on Detecting and Counting Fruitlet in Complex Orchard Environments’, vol. 2024, no. August, 2024, doi: 10.48550/arXiv.2407.12040.

[13] S. Tang and W. Yan, ‘Utilizing RT-DETR Model for Fruit Calorie Estimation from Digital Images’, *Inf. Switz.*, vol. 15, no. 8, 2024, doi: 10.3390/info15080469.

[14] E. K. Sahin, ‘Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest’, *SN Appl. Sci.*, vol. 2, no. 7, pp. 1–17, 2020, doi: 10.1007/s42452-020-3060-1.

[15] T. Aravanis, I. Hatzilygeroudis, and G. Spiliopoulos, ‘Ensemble Modelling for Predicting Fish Mortality †’, *Appl. Sci. Switz.*, vol. 14, no. 15, pp. 1–16, 2024, doi: 10.3390/app14156540.

[16] T. Fawcett, ‘An introduction to ROC analysis’, *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, 2006, doi: 10.1016/j.patrec.2005.10.010.

[17] A. B. Prakosa, Hendry, and R. Tanone, ‘Implementasi Model Deep Learning Convolutional Neural Network (CNN) Pada Citra Penyakit Daun Jagung Untuk Klasifikasi Penyakit Tanaman’, *J. Pendidik. Teknol. Inf. JUKANTI*, vol. 6, no. 1, pp. 107–116, 2023.

[18] O. B. Elhalid and A. H. Işık, ‘Enhancing Mobile Acute Lymphoblastic Cancer Detection: Transfer Learning from YOLOv9 to TensorFlow for Real-Time Applications’, *Int. J. Pioneer. Technol. Eng.*, vol. 3, no. 01, pp. 21–26, 2024, doi: 10.56158/jpte.2024.70.3.01.

[19] S. A. Hicks *et al.*, ‘On evaluation metrics for medical applications of artificial intelligence’, *Sci. Rep.*, vol. 12, no. 1, pp. 1–9, 2022, doi: 10.1038/s41598-022-09954-8.

[20] M. Sokolova, N. Japkowicz, and S. Szpakowicz, ‘Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation’, *AAAI Workshop - Tech. Rep.*, vol. WS-06-06, no. January, pp. 24–29, 2006, doi: 10.1007/11941439_114.