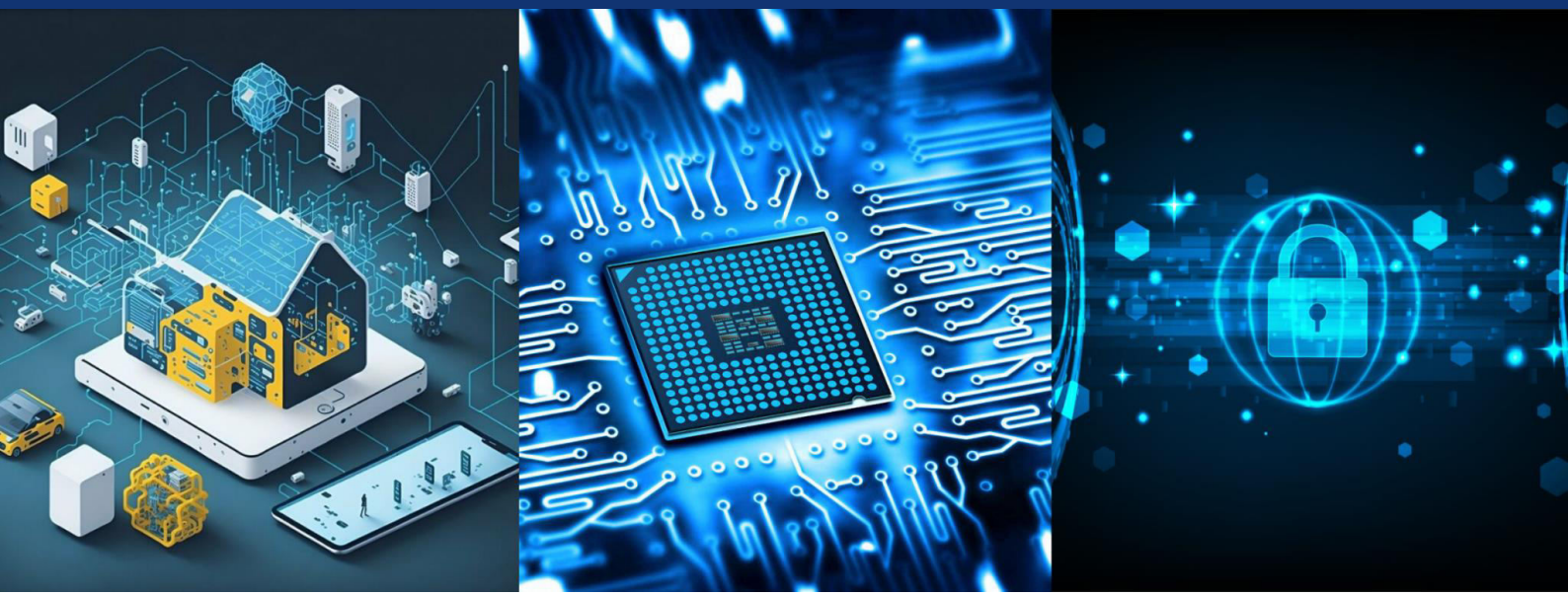
IJIRCCE

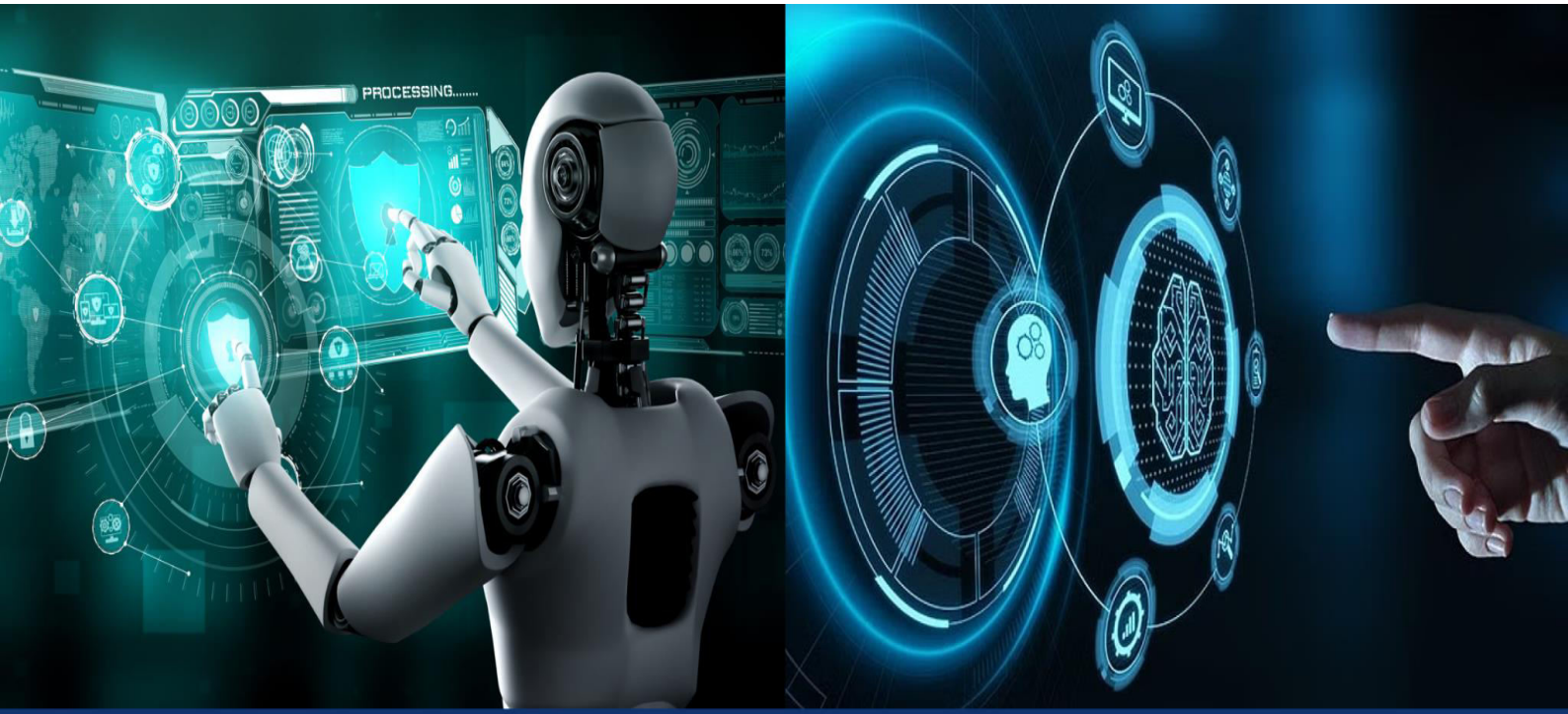
ISSN(O): 2320-9801

ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 9.927

Volume 14, Issue 11, November 2026

© 2026 IJIRCCE | Volume 14, Issue 11, November 2026 DOI: 10.15680/IJIRCCE.2026.1411011

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

AquaVision: A Web-Based Deep Learning Framework for Fish Detection, Species Classification, Instance Segmentation and AI-Assisted Health Assessment

Anushree S Bavikatti [1], Harshitha K S[1], Kiranmayi H K[1], Mangalagouri V G[1], Chaithra G S[2]

UG Students, Dept. of CSE, Jain Institute of Technology, Davangere, Karnataka, India[1]

Assistant Professor, Dept. of CSE, Jain Institute of Technology, Davangere, Karnataka, India[2]

ABSTRACT: Monitoring aquarium fish by eye is difficult. Several fish can share one scene, many species look alike, and early signs of disease are easy to miss. In this paper we describe AquaVision, a web application that takes a fish image or a short video and, in one pass, detects each fish, predicts its species, segments it, and gives an AI-assisted visual health assessment. Detection, species classification and segmentation come from a single Mask R-CNN model with a ResNet50-FPN backbone that covers 16 aquarium species. Each detected fish is then cropped and sent to a separate MobileNetV3 classifier with five health classes: Healthy, Ich, Fin Rot, Fungal and Dropsy. If the health model is less than 40% confident, the result is shown as Uncertain and no disease-specific advice is given. The system also has a knowledge base of species and care information, a FastAPI backend, a React interface, frame sampling for video, and optional PDF reports. At a 0.40 confidence threshold, the detection model reached a precision of 0.452, recall of 0.416, F1 score of 0.433 and mean IoU of 0.765 on the validation set. The health classifier reached 69.11% test accuracy and a macro-F1 of 0.6040. These numbers are modest, and we discuss where the models fall short. We see AquaVision as a working prototype showing that these tasks can be combined in one usable tool, not as a veterinary diagnosis system.

KEYWORDS: Fish Detection, Fish Species Classification, Instance Segmentation, Mask R-CNN, MobileNetV3, Fish Health Assessment, Deep Learning, Aquaculture Computer Vision, FastAPI, React.

I. INTRODUCTION

People who keep aquarium fish, or run small fish farms, need to check on their fish regularly, and that is harder than it sounds. Some species look very similar, and early signs of disease such as white spots, torn fins, fungus or swelling are easy to overlook without experience. Manual checking also depends on who is looking. Computer vision can help by picking up visual patterns from photos and video frames.

Most earlier fish-analysis work handles one task at a time: classifying species, detecting fish, or recognising disease. Someone who just wants to know what fish they have and whether it looks sick needs all of these together. So we built AquaVision as a single web application. The user uploads one image or a short video, and the system finds every fish, predicts its species, outlines each fish with a mask, checks its visible health, and adds species information and care advice.

We describe the health output as an AI-assisted visual assessment, not a diagnosis. Lighting, reflections, motion blur, fish hidden behind objects, cluttered backgrounds and similar-looking diseases all affect image-based models, and our own results show that the health model makes mistakes. For this reason AquaVision shows its confidence scores, and it marks a health prediction as Uncertain when the confidence is too low.

The main contributions of this work are as follows:

- A web system that combines fish detection, species classification, instance segmentation, health classification, a knowledge base and PDF reporting.

© 2026 IJIRCCE | Volume 14, Issue 11, November 2026 DOI: 10.15680/IJIRCCE.2026.1411011

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

- A Mask R-CNN pipeline that handles several fish in one image, with 16 supported species and a mask for each fish.
- A MobileNetV3 health classifier with five classes and a rule that rejects low-confidence predictions.
- A simple way to handle short videos: sample frames, reuse the image pipeline, and show one representative frame, without claiming to track individual fish.
- A separate knowledge layer, so species and care information can be edited without retraining the models.

II. RELATED WORK

Deep learning is now widely used for fish recognition. Deka et al. tested deep CNNs for classifying freshwater fish species. Ben Tamou et al. worked on species identification in difficult underwater images, using targeted augmentation and a hierarchical classification scheme . In a later paper they used incremental learning with a knowledge-distillation loss, so that a CNN can be updated when new data arrives.

For detecting and segmenting individual objects we use Mask R-CNN. It adds a mask branch to Faster R-CNN, so one network handles boxes, classes and masks. That suits multi-fish scenes, because every fish gets its own box and mask instead of the whole image receiving a single label.

The health step needs a second model that can run on many small fish crops. MobileNetV3 was designed with a hardware-aware architecture search and is small and fast, which fits this role.

Fish-disease detection has also received attention. Li et al. proposed YOLO-FD, a multi-task detector that adds semantic segmentation to a YOLO-based network. Another study combined several deep models to segment and classify fish diseases in aquaculture. These studies show that automatic health assessment is possible, but cluttered backgrounds and the varied look of diseases still make it difficult.

For preparing masks, Segment Anything and the lighter MobileSAM show that prompt-based segmentation can speed up annotation. We used MobileSAM to generate pseudo-masks for a balanced subset of our data before training Mask R-CNN.

Most of the work above covers a single task. AquaVision joins them in one web tool: per-fish detection and masks, species and health classification, confidence checks and care guidance, for both images and short videos.

III. PROPOSED ALGORITHM

This section follows the AquaVision pipeline from file upload to the final result. Figure 1 shows the overall architecture.

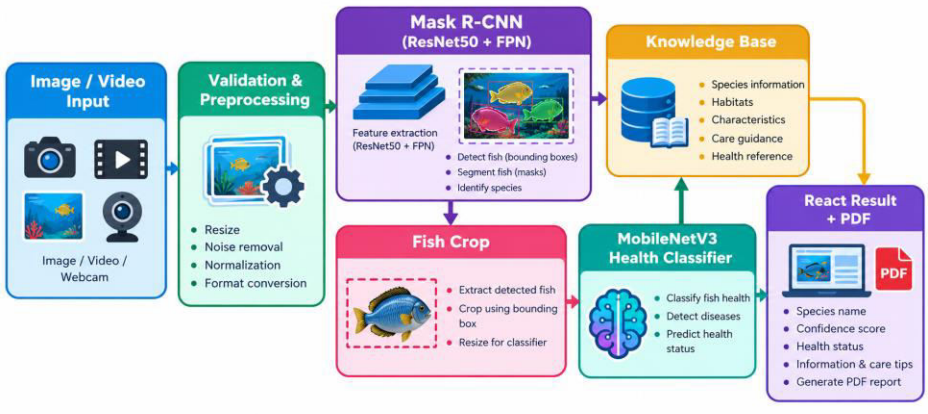


Fig. 1: End-to-end architecture of the proposed AquaVision system.

© 2026 IJIRCCE | Volume 14, Issue 11, November 2026 DOI: 10.15680/IJIRCCE.2026.1411011

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

3.1 Image and Video Acquisition

The user uploads one image or one recorded video. Images can be JPEG/JPG, PNG, WEBP, BMP or TIFF, with a 10 MB limit. Videos can be MP4, MOV, AVI, MKV, WEBM, MPEG/MPG or M4V, up to 100 MB and 30 seconds. The file is checked twice, once in the React frontend and again in the FastAPI backend.

3.2 Image Preprocessing

For images, the backend opens the file with PIL, fixes the EXIF orientation, converts it to RGB and saves a normalized JPEG before inference. Mask R-CNN then resizes the image internally, using a minimum side of 384 pixels and a maximum side of 512 pixels, and keeps the aspect ratio.

Let I denote the decoded RGB image. The normalized model input can be represented conceptually as:

I_p = Resize_and_Normalize (RGB(EXIF_Correct(I)))

3.3 Fish Detection, Species Classification and Instance Segmentation

The main model is Mask R-CNN with a ResNet50 backbone and a Feature Pyramid Network (FPN), implemented in PyTorch with TorchVision. For each image it predicts candidate boxes, class scores and segmentation masks. It has 17 output classes: the 16 supported species plus background. At inference we drop detections with a score below 0.40 and binarize the masks at 0.50.

We apply class-agnostic non-maximum suppression (NMS) with an IoU threshold of 0.50 to remove duplicate detections. During testing, a koi image produced a small extra detection inside the main fish, labelled as a Neon Tetra. To fix this we added a contained-box rule: a detection is dropped when at least 85% of it lies inside a larger box and its area is no more than 50% of that box.

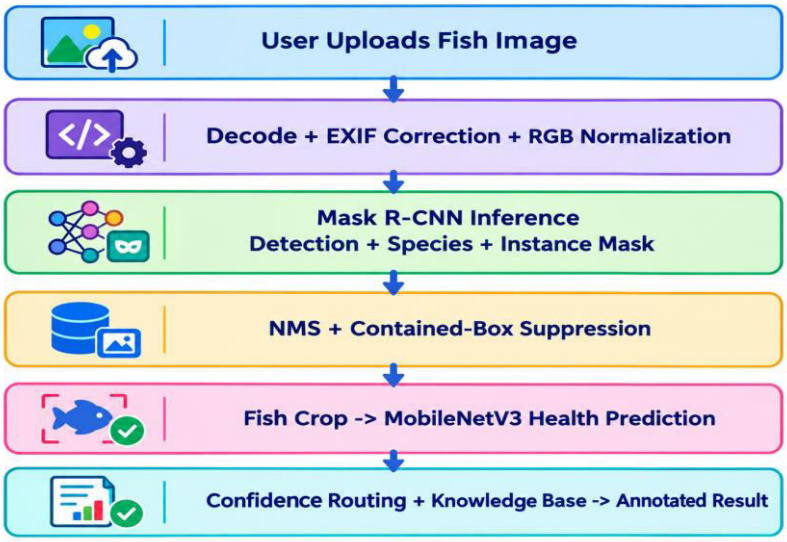


Fig. 2: Image analysis pipeline from upload to the integrated result.

3.4 Supported Fish Species

AquaVision currently supports 16 classes: Guppy, Betta, Goldfish, Koi, Molly, Platy, Swordtail, Angelfish, Neon Tetra, Black Skirt Tetra, Zebra Danio, Tiger Barb, Three Spot Gourami, Corydoras, Bristlenose Pleco and Clown Loach. It cannot yet reject fish outside these classes, so an unsupported species will still be assigned to one of the 16. An explicit unknown-species class is future work.

3.5 Per-Fish Health Assessment

Each accepted detection is cropped from the image and passed to a MobileNetV3 classifier with five classes: Healthy, Ich, Fin Rot, Fungal and Dropsy. Besides the top class, we keep the second-best prediction, all class probabilities, the confidence margin and a confidence label, so the interface can show how sure the model is.

© 2026 IJIRCCE | Volume 14, Issue 11, November 2026 DOI: 10.15680/IJIRCCE.2026.1411011

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

3.6 Confidence-Aware Decision Routing

Not every health prediction should be shown as a result, so we use a fixed confidence rule. If the highest health probability is below 40%, the fish is marked Uncertain and no disease-specific treatment advice is returned. From 40% to 69.99% the prediction is accepted with Moderate confidence, and from 70% upwards with High confidence. The predicted class then decides whether the fish is shown as Healthy or Diseased.

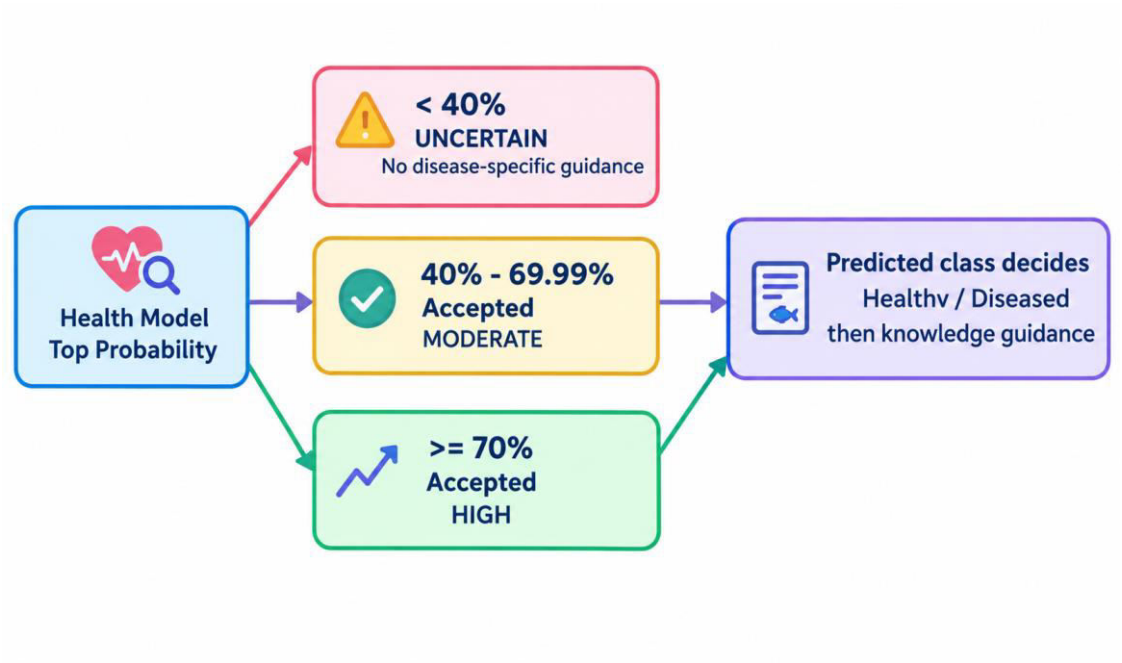


Fig. 3: Confidence-aware routing for the MobileNetV3 health output.

3.7 Knowledge-Based Information and Guidance

Species and care information live in a separate knowledge layer, not inside the neural networks. For each species it stores the scientific name, care level, temperament, diet, recommended temperature, pH range, minimum tank size and a short description. A second table holds a summary and recommended care steps for each health condition. Keeping this apart from the models means the content can be reviewed and updated without retraining. When the health result is Uncertain or Unknown, disease-specific guidance is deliberately not shown.

3.8 Short Video Analysis

AquaVision reads videos with OpenCV [13]. It first extracts the duration, frame rate, resolution and frame count, then samples about one frame per second and runs each sampled RGB frame through the full image pipeline. We do not track fish across frames, so fish numbers only make sense inside a single frame.

To give the user a clear result, we pick one representative frame. Frames are ranked by the number of detected fish first, then by average species confidence and average health confidence as tie-breakers. The chosen frame is saved, annotated and shown with the same per-fish cards as for images. An earlier version listed every sampled frame, but the same fish appeared repeatedly under different numbers, which was confusing and made the report very long. Showing one frame also avoids suggesting that Fish 1 in one frame is the same fish as Fish 1 in another.

© 2026 IJIRCCE | Volume 14, Issue 11, November 2026 DOI: 10.15680/IJIRCCE.2026.1411011

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

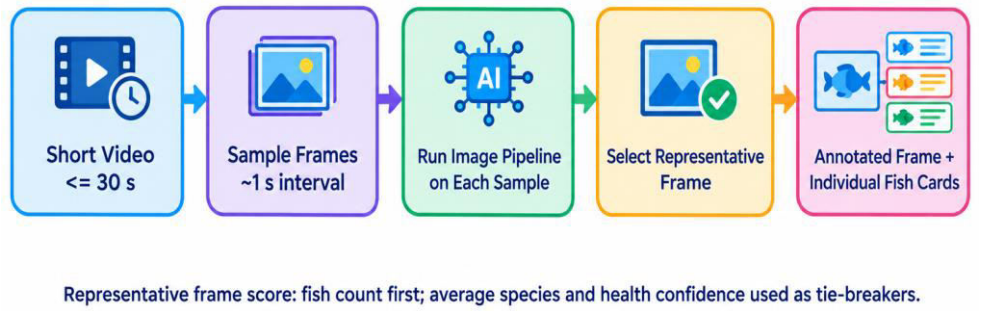


Fig. 4: Short-video analysis and representative-frame selection workflow.

3.9 Pseudocode of the Proposed Algorithm

1. Initialize Mask R-CNN species/detection model, MobileNetV3 health model, knowledge base, and API services.
2. Accept one image or one short video from the user and validate type, size, and duration.
3. For image input, decode, correct orientation, convert to RGB, and preprocess for inference.
4. Run Mask R-CNN to obtain bounding boxes, species labels, scores, and instance masks.
5. Apply score filtering, NMS, and contained-box suppression.
6. For every remaining fish detection, crop the fish region and run MobileNetV3 health classification.
7. If health confidence < 0.40, mark the result Uncertain and suppress disease-specific guidance.
8. Else accept the predicted health class; label 0.40-0.6999 as Moderate and >= 0.70 as High confidence.
9. Retrieve species information and safe care/treatment guidance from the knowledge base.
10. Create an annotated image and return fish-level JSON to the React frontend.
11. For video input, sample frames, repeat Steps 4-9 internally, select the representative frame, and return only that frame's user-facing fish results.
12. Optionally generate a PDF report from the final analysis JSON.

IV. EXPERIMENTAL RESULTS AND SYSTEM EVALUATION

4.1 Detection / Species Model Performance

At a confidence threshold of 0.40, the Stage-2 detection model reached a validation precision of 0.452, recall of 0.416, F1 of 0.433 and mean IoU of 0.765. The IoU shows that the detections it does accept overlap well with the ground truth, but the precision and recall show that it still misses fish and makes wrong predictions, so there is plenty of room to improve.

Metric	Result
Precision	0.452
Recall	0.416
F1 score	0.433
Mean IoU	0.765

Per-class evaluation showed that some species are much weaker than others: Three Spot Gourami (0.062), Black Skirt Tetra (0.158) and Guppy (0.246). We therefore cannot claim equal reliability across all 16 species, and the interface always shows the confidence score.

© 2026 IJIRCCE | Volume 14, Issue 11, November 2026 DOI: 10.15680/IJIRCCE.2026.1411011

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

4.2 Health Classifier Performance

The Stage-2 MobileNetV3 health classifier reached 69.11% test accuracy and a macro-F1 of 0.6040. Performance differs a lot between classes. Ich did best, with an F1 of 0.813. Healthy recall was only 0.400, so more than half of the healthy fish in the test set were assigned to a disease class.

Class	Precision	Recall	F1
Dropsy	0.833	0.357	0.500
Fin Rot	0.417	0.833	0.556
Fungal	0.567	0.680	0.618
Healthy	0.800	0.400	0.533
Ich	0.769	0.862	0.813

These results are the reason for the uncertainty rule, and for calling the health output an AI-assisted visual assessment instead of a diagnosis.

4.3 Functional System Evaluation

We tested the complete prototype by uploading images and videos and checking multi-fish results, confidence display, uncertainty routing, PDF generation and page navigation. Testing also showed where the system fails. An Angelfish image was detected correctly, but a Goldfish test image produced no detections at all, so the model can miss fish. We fixed the koi duplicate detection with the contained-box rule from Section 3.3. We also corrected the video confidence labels so that the frontend shows the same Low, Moderate or High label that the health model produces, and we simplified the video interface to show one annotated representative frame with individual fish cards.

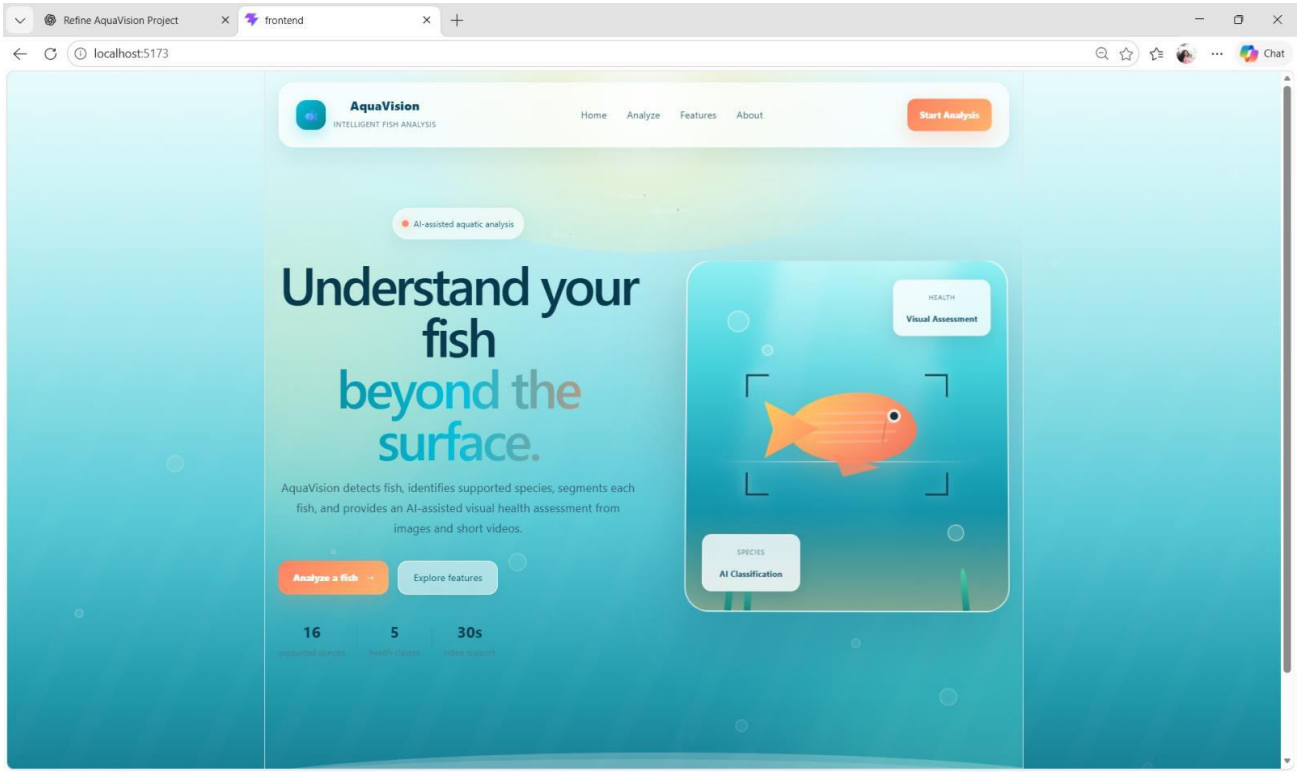


Fig. 5: Home Page of the AquaVision Fish Detection and Classification System

© 2026 IJIRCCE | Volume 14, Issue 11, November 2026 DOI: 10.15680/IJIRCCE.2026.1411011

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

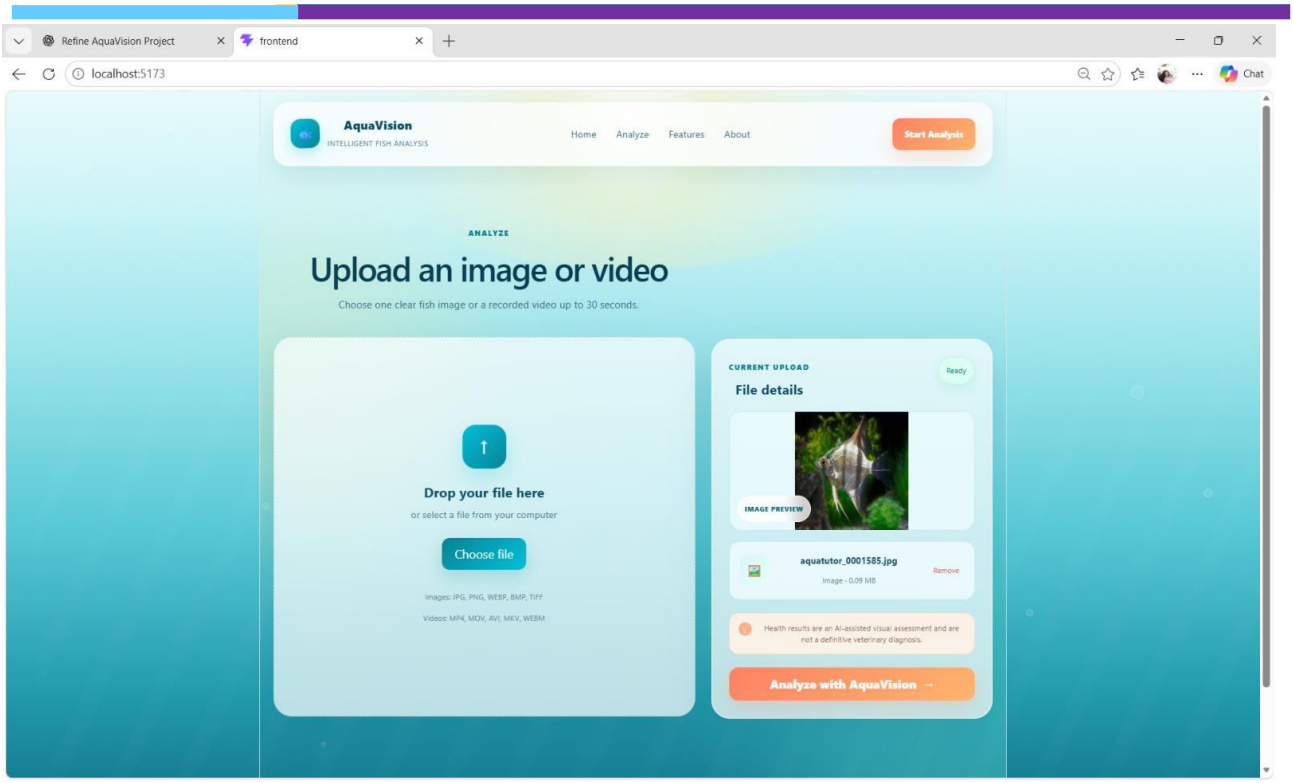


Fig. 6: Image and Video Upload Page of the AquaVision System.

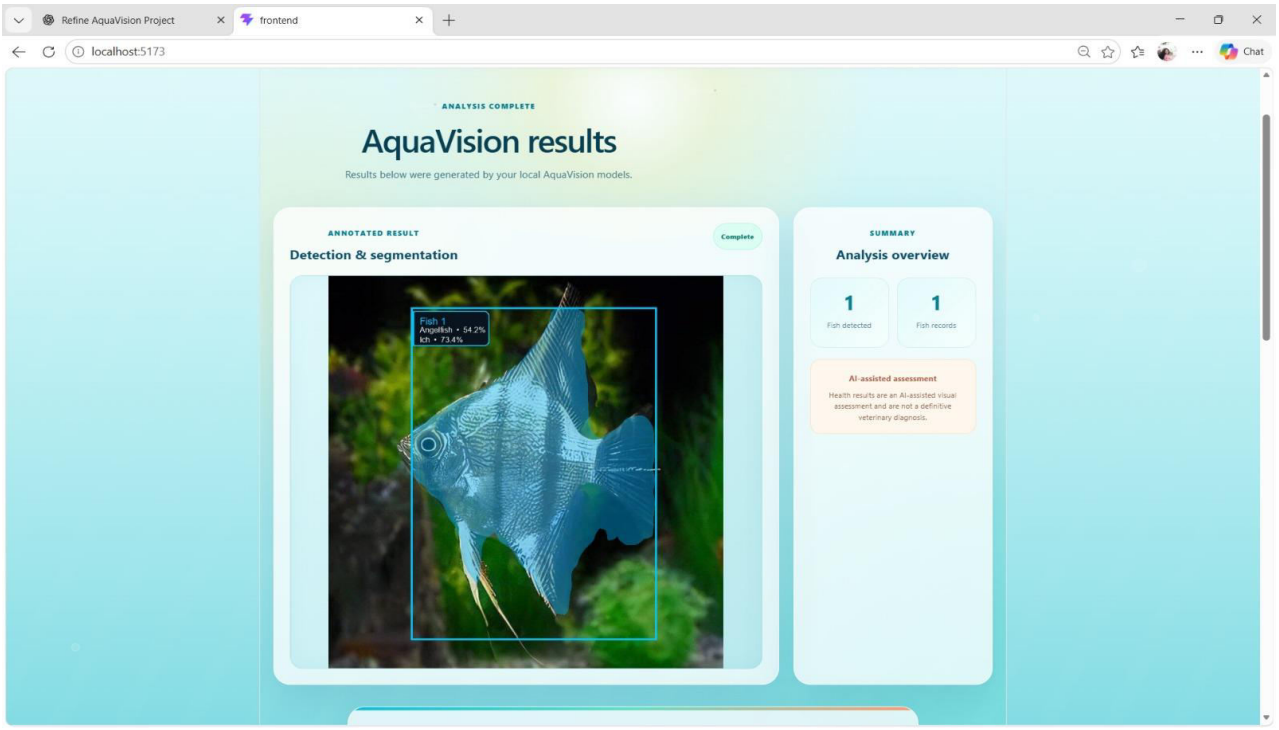


Fig. 7: Individual fish result cards showing species prediction, confidence and AI-assisted health assessment.

© 2026 IJIRCCE | Volume 14, Issue 11, November 2026 DOI: 10.15680/IJIRCCE.2026.1411011

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

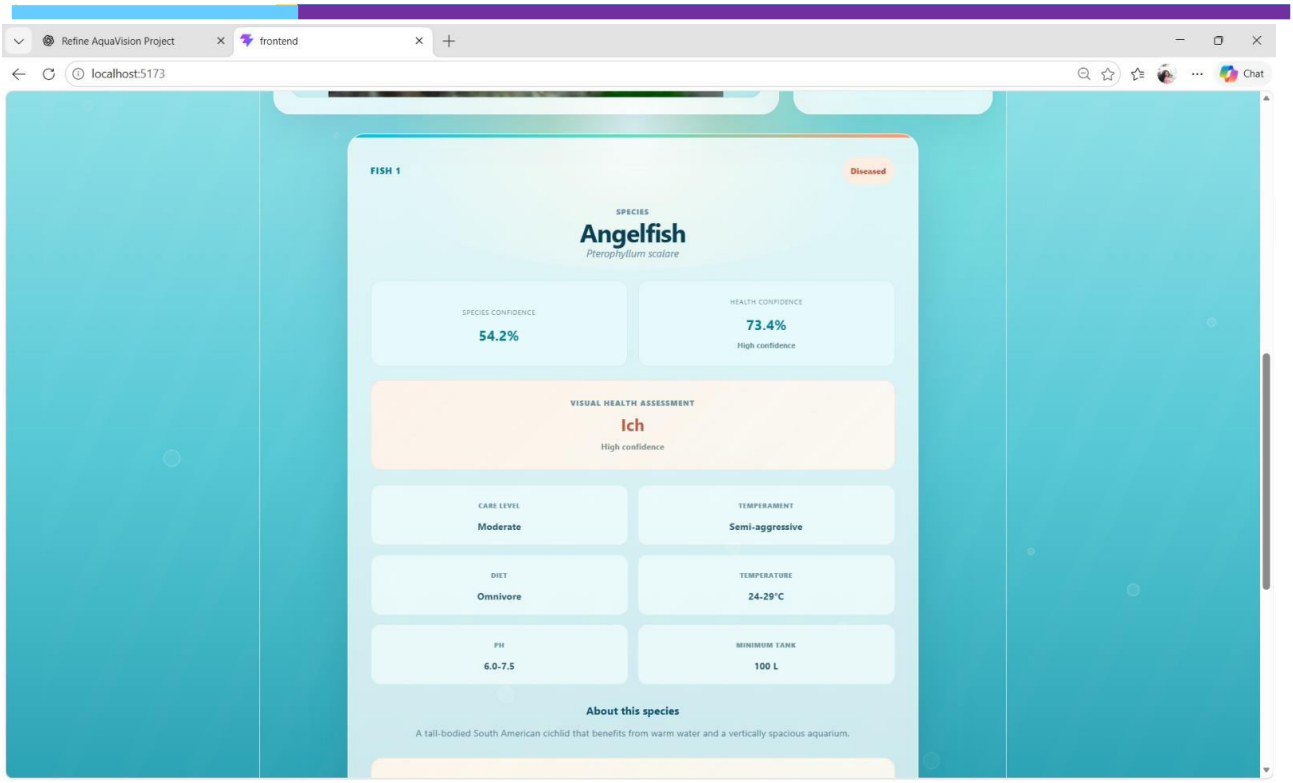


Fig. 8: Fish Species Identification and Health Assessment Results

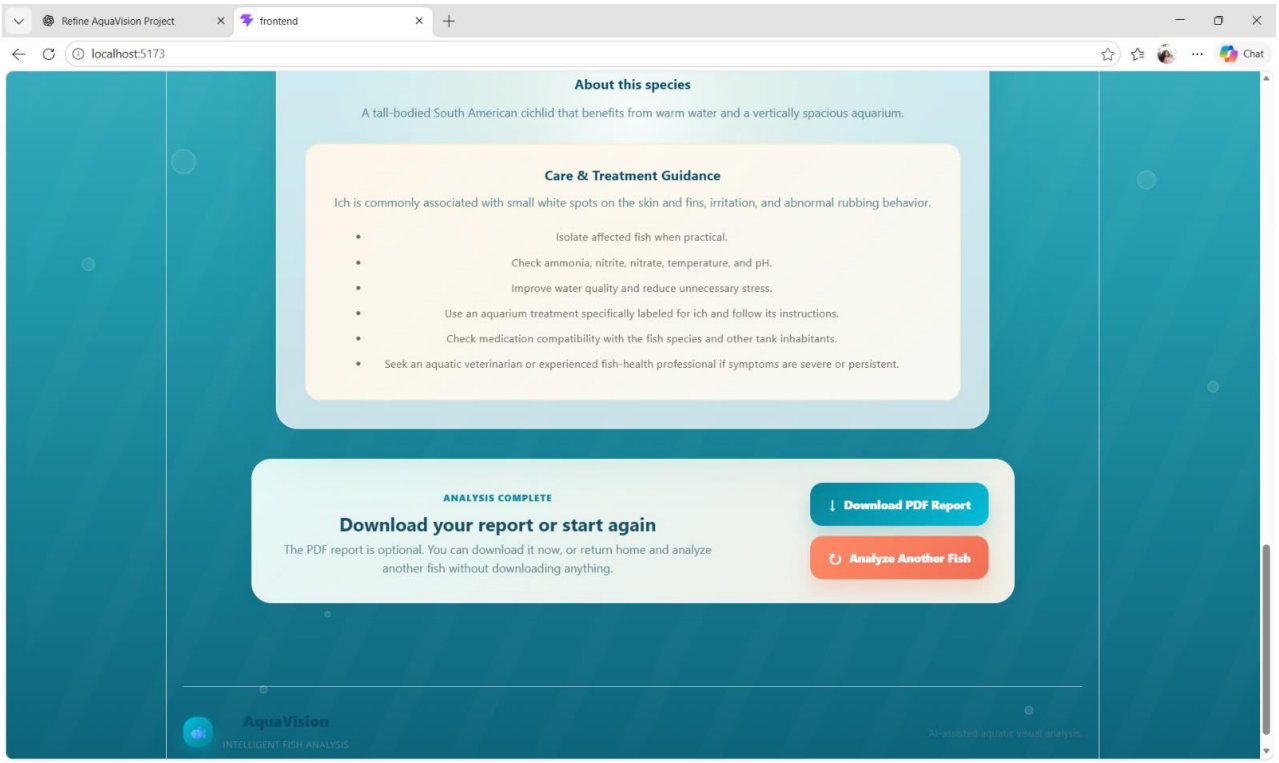


Fig. 9: Fish Care and Treatment Guidance with PDF Report Generation

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

4.4 Example Video Analysis

In one 5.16-second test video, the representative frame was picked near 0.97 seconds and contained four fish. Two were predicted as Goldfish, with species confidences of 82.1% and 81.9%, and two as Platy, with lower confidences of 42.7% and 42.2%. The health predictions were Ich at 59.3% (Moderate), Dropsy at 46.7% (Moderate), Dropsy at 52.2% (Moderate) and Dropsy at 88.4% (High). This example shows that the confidence routing and display work as designed. It does not show that the fish actually had these conditions.

Fish	Species prediction	Health prediction
Fish 1	Goldfish - 82.1%	Ich - 59.3% (Moderate)
Fish 2	Goldfish - 81.9%	Dropsy - 46.7% (Moderate)
Fish 3	Platy - 42.7%	Dropsy - 52.2% (Moderate)
Fish 4	Platy - 42.2%	Dropsy - 88.4% (High)

4.5 Web Application and Report Generation

The frontend is built with React and Vite and has Home, Analyze, Features, About and Results pages. The FastAPI backend has separate endpoints for image analysis, video analysis and report generation. Annotated images are served from a results folder, and ReportLab [14] creates an optional PDF with the same per-fish information shown on the website.

V. CONCLUSION AND FUTURE SCOPE

We presented AquaVision, a web application that detects fish, predicts their species, segments them and gives an AI-assisted visual health assessment from images and short videos. It uses Mask R-CNN with a ResNet50-FPN backbone for detection, species and masks, and MobileNetV3 for five-class health prediction. Health predictions below 40% confidence are shown as Uncertain, and a separate knowledge layer supplies species information and care guidance. The prototype produces annotated results, per-fish cards and downloadable PDF reports.

The results show a working but limited system. The detection model reached a precision of 0.452, recall of 0.416, F1 of 0.433 and mean IoU of 0.765 on the validation set, and the health classifier reached 69.11% test accuracy with a macro-F1 of 0.6040. Some species, such as Three Spot Gourami, are recognised poorly, healthy fish are often labelled as diseased, and the model missed a Goldfish image entirely. These are the main problems to fix, and the system should not be used as a diagnosis.

Future work includes splitting the data by source image to avoid leakage, collecting more better and balanced images for the weak species, adding an Unknown class for unsupported species, and improving recall on healthy fish. For video, we would like to add tracking such as ByteTrack or DeepSORT and combine evidence across frames. Longer-term ideas include severity estimation, water-quality sensors, mobile deployment and faster inference on a GPU.

Any real use for treatment decisions would need validation with aquatic-health professionals. Until then, AquaVision should be treated as decision support only.

REFERENCES

[1] J. Deka, S. Laskar, and B. Baklial, “Automated Freshwater Fish Species Classification Using Deep CNN,” Journal of The Institution of Engineers (India): Series B, vol. 104, pp. 603–621, 2023. https://doi.org/10.1007/s40031-023-00883-2

[2] A. Ben Tamou, A. Benzinou, and K. Nasreddine, “Targeted Data Augmentation and Hierarchical Classification with Deep Learning for Fish Species Identification in Underwater Images” Journal of imaging, vol.8, no. 8, Art. 214, 2022.
https://doi.org/10.3390/jimaging8080214

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

[3] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in Proc. IEEE International Conference on computer vision (ICCV), 2017 PP, 2961-2969.
https://doi.org/10.1109/ICCV.2017.322

[4] A. Howard et al., "Searching for MobileNetV3," in Proc. IEEE International conference on computer vision (ICCV), 2019 PP, 1314-1324.
https://doi.org/10.1109/ICCV.2019.00140

[5] X. Li et al., "YOLO-FD: An Accurate Fish Disease Detection Method Based on Multi-Task Learning" Expert systems with Applications, vol. 258, Art 12085, 2024.
https://doi.org/10.1016/j.eswa.2024.125085

[6] C. Zhang et al., "Faster Segment Anything: Towards Lightweight SAM for Mobile Applications," arXiv preprint arXiv:2306.14289, 2023.
https://arxiv.org/abs/2306.14289

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90

[8] T.-Y. Lin et al., "Feature Pyramid Networks for Object Detection," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2117–2125. https://doi.org/10.1109/CVPR.2017.106



INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details