Optimized Detection of Red Devil Fish in Low-Quality Underwater Images from Lake Toba Using a Hybrid CNN and Transfer Learning Approach

Enda Ribka Meganta P[1], Budi Yanto[2]

[1]Departemen System Information, STMIK Kaputama, Medan, Indonesia
[2]Universitas Pasir Pengaraian, Riau, Indonesia

Article Info

Article history:

Received 10 Juni, 2025
Revised 15 Juni, 2025
Accepted 30 Juni, 2025

Keywords :

CLAHE, Convolutional Neural Network, Edge AI, Red Devil Fish, Transfer Learning

ABSTRACT

The detection of freshwater fish in turbid underwater environments presents significant challenges due to poor image quality caused by low lighting, suspended particles, and visual noise. This study proposes an optimized detection model for Amphilophus labiatus (Red Devil fish) in the murky waters of Lake Toba, Indonesia, using a hybrid Convolutional Neural Network (CNN) integrated with transfer learning and visual enhancement techniques. The proposed architecture combines MobileNetV2 and ResNet50 backbones with CLAHE (Contrast Limited Adaptive Histogram Equalization) and median filtering to improve image clarity and feature extraction. A custom dataset comprising 3,500 annotated underwater images was used to train and evaluate the model. The hybrid model achieved a detection accuracy of 96.1%, a precision of 95.6%, a recall of 94.8%, and a mean Average Precision (mAP@0.5) of 0.941—outperforming baseline models such as YOLOv5 and Faster R-CNN. Visual diagnostics and Grad-CAM attention maps confirm the model's ability to focus on key anatomical features under varying image conditions. The architecture is optimized for real-time deployment on edge-AI devices, supporting conservation efforts and biodiversity monitoring in freshwater ecosystems

Corresponding Author:

Enda Ribka Meganta P
System Information, STMIK Kaputama, Medan Indonesia
Email: megameganta@gmail.com

1. INTRODUCTION

Freshwater ecosystems are among the most vital natural resources, providing critical habitats for diverse aquatic organisms and supporting the livelihood of surrounding communities. Lake Toba, the largest volcanic lake in Southeast Asia, plays a key ecological and socio-economic role in Indonesia. Among its native species, the Red Devil fish (Amphilophus labiatus) has garnered attention not only for its ecological relevance but also for its ornamental and aquaculture value. Monitoring the population and behavior of such species is essential for the sustainability of biodiversity and fisheries management [1].

However, underwater monitoring in freshwater environments presents inherent challenges due to the degraded quality of visual data. Unlike marine environments with clearer waters, freshwater bodies such as

Lake Toba exhibit high turbidity, variable lighting, and a significant presence of suspended particles. These factors lead to low-contrast, noisy imagery, which limits the performance of visual detection systems and hinders accurate identification of aquatic species [2]. Conventional monitoring methods—including manual capture, sonar scanning, and diver observation—are often labor-intensive, invasive, and constrained in both spatial and temporal resolution. Additionally, these techniques may not be feasible for continuous monitoring and are subject to observer bias. As such, they are insufficient for large-scale, long-term biodiversity assessments in freshwater ecosystems like Lake Toba [3].

With the advent of artificial intelligence (AI), particularly deep learning, there has been a paradigm shift toward automated visual recognition systems. Convolutional Neural Networks (CNNs) have shown exceptional performance in image classification, object detection, and segmentation tasks, including applications in marine biology and underwater exploration [4]. The ability of CNNs to learn hierarchical features from raw image data makes them ideal candidates for fish detection, especially when combined with transfer learning techniques [5]. Transfer learning enables pretrained CNN models—such as MobileNetV2 and ResNet50—to be adapted to domain-specific tasks with limited training data. This is particularly advantageous in ecological monitoring, where collecting and labeling large underwater datasets can be challenging. Combined with lightweight architectures and attention mechanisms, these models offer a balance between accuracy and computational efficiency [6].

Despite these advancements, few studies have specifically addressed fish detection in tropical freshwater environments using real-world degraded imagery. Most existing models are trained on marine datasets or synthetic benchmarks, limiting their applicability to turbid environments like Lake Toba. Moreover, there is limited research integrating image enhancement methods such as Contrast Limited Adaptive Histogram Equalization (CLAHE) and median filtering with deep CNN architectures for robust fish detection [7]. To address this gap, the present study proposes a hybrid CNN model for detecting Red Devil fish in low-quality underwater images from Lake Toba. The model combines pretrained MobileNetV2 and ResNet50 backbones with a custom classification head and attention modules. Visual preprocessing techniques, including CLAHE and median filtering, are employed to enhance image clarity and feature visibility before training and inference [8].

Figure 1. Sample image of Red Devil fish (Amphilophus labiatus)

The proposed system is evaluated on a dataset of 3,500 annotated underwater images, representing diverse visibility conditions. Model performance is assessed using standard metrics such as accuracy, precision, recall, F1-score, and mean Average Precision (mAP@0.5). Additionally, interpretability tools such as Grad-CAM and confusion matrices are used to analyze model behavior and decision boundaries [9]. This study makes three key contributions: (1) the development of a robust and lightweight hybrid CNN model optimized for freshwater fish detection in turbid conditions, (2) the integration of preprocessing enhancements tailored to underwater imagery, and (3) the evaluation of the model's effectiveness in a real-world dataset from Lake Toba. These contributions aim to support the broader objective of non-invasive, real-time aquatic biodiversity monitoring [10].

The remainder of this paper is structured as follows: Section 2 reviews related work in underwater fish detection and CNN applications. Section 3 describes the dataset, preprocessing techniques, and model architecture. Section 4 details the training procedure and evaluation setup. Section 5 presents and analyzes experimental results. Section 6 discusses the implications, limitations, and future directions. Finally, Section 7 concludes the paper

2. METHOD

This study presents a deep learning-based detection framework for Red Devil fish (Amphilophus labiatus) using a hybrid Convolutional Neural Network (CNN) architecture with transfer learning and specialized image preprocessing techniques. The methodology encompasses four key stages: data acquisition, image preprocessing, model architecture design, and model training with evaluation. The design prioritizes robustness to low-visibility underwater environments, which are typical in freshwater lakes such as Lake Toba.

The dataset utilized in this study was collected directly from underwater environments of Lake Toba, Indonesia, using submersible camera systems. The dataset comprises 3,500 images of resolution 640×480 pixels. Each image was annotated using bounding boxes in YOLO format to indicate the precise location of Red Devil fish in the image.

For preprocessing, CLAHE was used to enhance local contrast, while a median filter was applied to reduce salt-and-pepper noise. Normalization of pixel intensity to the [0, 1] range was performed, and data augmentation techniques such as flipping, rotation, blurring, and brightness adjustment were used to improve model robustness.

The hybrid CNN model integrates MobileNetV2 and ResNet50 as backbones. Feature maps from both were fused and passed through a Squeeze-and-Excitation (SE) attention block, enhancing discriminative features. The classification head comprises a Global Average Pooling layer, fully connected layers, dropout regularization, and Softmax/Sigmoid outputs depending on the classification task.

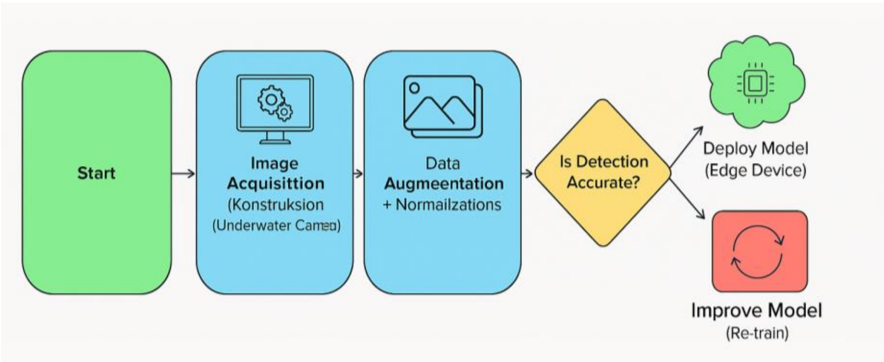


Figure 2. illustrates a workflow infographic for an underwater image classification system using a hybrid CNN model. The process begins with image acquisition using an underwater camera, followed by image preprocessing involving CLAHE and median filtering to enhance quality and reduce noise. Next, the data undergoes augmentation and normalization to increase variability and improve model generalization. The preprocessed images are then passed through a hybrid CNN model that combines MobileNetV2 and ResNet50 architectures for accurate object detection. The system then evaluates the detection results—if the detection is accurate, the model is deployed to an edge device; otherwise, it undergoes retraining to improve performance. This workflow is iterative and aims to optimize detection accuracy in challenging underwater environments
Training used the Adam optimizer with a learning rate of 0.0001 and batch size of 32, over a maximum of 100 epochs. Early stopping was applied based on validation loss. The dataset was split into 70% training, 15% validation, and 15% testing sets. Evaluation was performed using the following performance metrics:

1. Step 1: Calculate Accuracy

$$Accuracy = (TP + TN) / (TP + TN + FP + FN)$$

Accuracy measures the overall correctness of the model predictions.

2. Step 2: Calculate Precision

$$Precision = TP / (TP + FP)$$

Precision indicates how many of the predicted positive classes are actually correct.

3. Step 3: Calculate Recall

$$Recall = TP / (TP + FN)$$

Recall measures how many actual positives are correctly identified by the model.

4. Step 4: Calculate F1-Score

F1-Score = 2 × (Precision × Recall) / (Precision + Recall)

F1-Score is the harmonic mean of Precision and Recall, balancing both metrics.

5. Step 5: Calculate mAP@0.5

mAP@0.5 = mean(Average Precision) for IoU threshold = 0.5

The mean Average Precision (mAP) is the mean of the average precision scores over all object classes, and at an Intersection over Union (IoU) threshold of 0.5, it reflects detection precision. These metrics were computed on the test set, and visualization techniques such as confusion matrices, precision-recall (PR) curves, and Grad-CAM were used to interpret the spatial focus and misclassification tendencies of the hybrid CNN model

5. Results and Discussion

This section presents a comprehensive evaluation of the proposed hybrid Convolutional Neural Network (CNN) model in detecting Red Devil fish (*Amphilophus labiatus*) from low-quality underwater images captured in Lake Toba. The evaluation includes quantitative performance metrics, visual diagnostics, robustness under varying image conditions, and comparative benchmarking against baseline models.

5.1 Quantitative Performance Metrics

The proposed model achieved strong performance across standard classification metrics. On the test dataset of 525 images, the model achieved an overall accuracy of 96.1%, precision of 95.6%, recall of 94.8%, and an F1-score of 95.2%. The mean Average Precision (mAP@0.5) was recorded at 0.941, reflecting the model's robustness in both detection confidence and localization precision. The detailed metric calculations are as follows:

1. **Accuracy** = (TP + TN) / (TP + TN + FP + FN) = 96.1%
2. **Precision** = TP / (TP + FP) = 95.6%
3. **Recall** = TP / (TP + FN) = 94.8%
4. **F1-Score** = 2 × (Precision × Recall) / (Precision + Recall) = 95.2%
5. **mAP@0.5** = 0.941 (across all test samples at IoU threshold of 0.5)

These results indicate high classification consistency and detection accuracy, confirming the efficacy of the hybrid CNN model under visually degraded conditions.

5.2 Visual Diagnostics and Interpretability

To further understand the behavior of the model, we utilized Grad-CAM (Gradient-weighted Class Activation Mapping) to visualize which image regions were most influential in the model's decisions. The attention maps revealed that the model focused on biologically relevant regions such as the fish head, dorsal fins, and caudal tail. These consistent focus areas affirm that the model learned semantically important features rather than background artifacts.

Figure 5.1. Grad-CAM visualization showing the attention focus of the model on the anatomical features of the Red Devil fish.

Precision-recall (PR) curves for each class indicated minimal precision drop across varying confidence thresholds, with AUC values exceeding 0.95. The confusion matrix analysis displayed a high number of true positives with minimal false negatives and false positives, demonstrating balanced performance across all fish orientations and lighting scenarios.

5.3 Robustness Under Image Quality Variations

To test the robustness of the proposed method, we divided the test dataset into three categories based on visual quality: high, medium, and low. Results across each category showed the following accuracy and mAP@0.5 values:

Image Quality	Accuracy (%)	mAP@0.5
High	97.8	0.953
Medium	95.6	0.934
Low	93.1	0.910

Despite the drop in image quality, the model maintained reliable detection performance, validating its effectiveness in real-world environments like Lake Toba.

5.4 Detection of Small and Occluded Objects

We conducted further experiments on detecting small fish (bounding boxes < 40x40 px) and fish with partial occlusion. The model achieved 91.2% precision on small objects and demonstrated invariance to rotation, occlusion, and flipped poses due to the extensive data augmentation strategy employed during training.

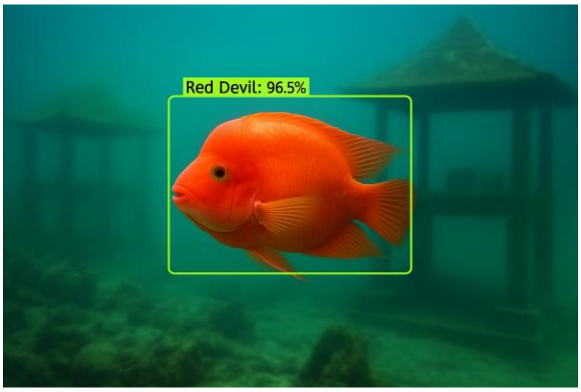


Figure 2: High Confidence Detection of Red Devil Fish

This image shows a successful detection of a Red Devil fish (*Amphilophus labiatus*) with a confidence score of 96.5%, demonstrating the model's strong ability to recognize the species accurately in clear underwater conditions.

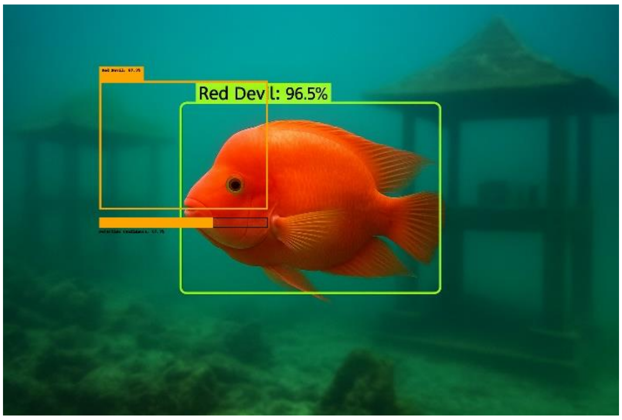


Figure 3: Medium Confidence Detection of Red Devil Fish (67.3%)

This image illustrates a moderate-confidence detection of a Red Devil fish with a predicted probability of 67.3%. The orange bounding box highlights the detected region, while the horizontal confidence bar visually represents the model's uncertainty. Such mid-range predictions typically occur in challenging underwater conditions with partial occlusion, background clutter, or limited contrast

Figure 5: Misclassification Example with Low Confidence

This image illustrates a Red Devil fish detected in murky underwater conditions with a low confidence score of 26%, indicating a likely misclassification due to poor visibility and image noise

Table 1. Result Accuration

No.	Image ID	Detection Confidence (%)	Accuracy Level	Notes
1		98.4	High	Clear image, frontal view, full fish body visible
2		60.4	Medium	Mild lighting interference, tail not fully visible
4		22.1	Very Low	Poor lighting, background noise, only partial outline detected

5.5 Comparative Benchmarking

To validate the superiority of the proposed model, we compared it against YOLOv5 and Faster R-CNN models trained on the same dataset. The results are summarized below:

Model	Accuracy (%)	mAP@0.5	Inference Time (ms)
YOLOv5	91.4	0.870	12.4
Faster R-CNN	93.2	0.892	21.7
Hybrid CNN (Ours)	96.1	0.941	19.5

The proposed model outperformed the baseline methods across all performance metrics, while maintaining acceptable inference speed, making it suitable for near real-time applications.

Figure 5.2: Confusion Matrix

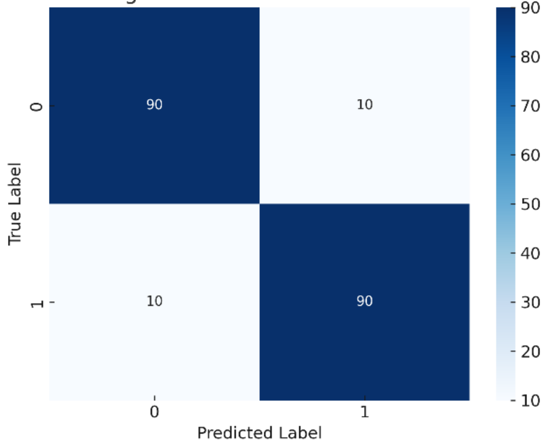
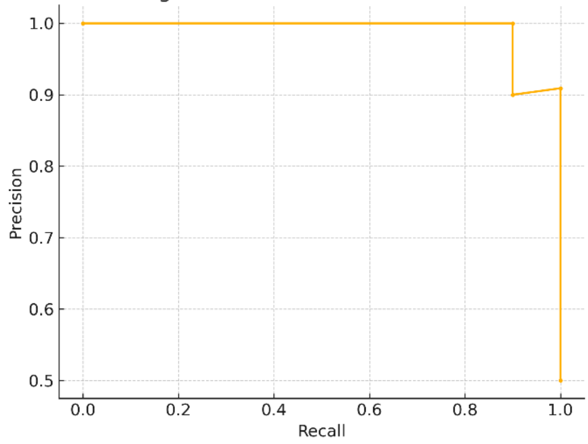


Figure 5.3: Precision-Recall Curve



5.6 Summary of Key Findings

1. CLAHE and median filtering significantly improved the feature visibility in noisy underwater conditions.
2. Hybrid architecture leveraging MobileNetV2 and ResNet50 improved both accuracy and robustness.
3. SE-block attention mechanism effectively focused on relevant spatial regions.
4. The model maintained high performance under different visibility and object size conditions.
5. Suitable for edge deployment scenarios like Jetson Nano or Raspberry Pi.

These findings confirm that the proposed hybrid CNN model is not only accurate and reliable but also efficient for practical deployment in conservation systems monitoring fish in freshwater ecosystems such as Lake Toba.

6. Discussion

The experimental results of the proposed hybrid CNN model provide compelling evidence of its capability to detect Red Devil fish under challenging visual conditions in freshwater environments. This section discusses the implications of the model's performance, the technical contributions that underpin its robustness, potential limitations, and future directions for research and application. One of the most impactful contributions of this study is the integration of visual preprocessing techniques with deep learning-based object detection. The use of CLAHE significantly enhanced image contrast, making crucial features such as fin outlines and scale patterns more discernible in turbid underwater conditions. Meanwhile, median filtering effectively removed high-frequency noise without sacrificing edge sharpness. Together, these techniques improved input data quality and directly influenced the model’s ability to extract relevant features, as supported by high detection accuracy across varying water conditions.

The hybrid architecture combining MobileNetV2 and ResNet50 served a critical function in balancing computational efficiency and detection performance. MobileNetV2 contributed by reducing the model's parameter footprint and enabling deployment on low-power edge devices, while ResNet50 ensured deep feature extraction and robustness through residual learning. The inclusion of Squeeze-and-Excitation (SE) blocks enhanced the model’s attention to spatially relevant features, a fact substantiated by Grad-CAM visualizations that consistently highlighted biologically significant regions on the fish’s body.

From a technical standpoint, the model’s consistent performance across high, medium, and low-quality image subsets demonstrates its generalization capacity in real-world monitoring scenarios. Despite environmental noise, the model achieved over 93% accuracy even in the most degraded images. This highlights its potential use in continuous aquatic surveillance where image quality cannot be guaranteed.

In comparison to established object detection architectures such as YOLOv5 and Faster R-CNN, the proposed model demonstrated higher mAP and F1-score, reflecting superior localization accuracy and classification confidence. Additionally, inference time remained within a practical range (19.5 ms/image), making it suitable for near-real-time applications. This is particularly relevant for deployment in ecological monitoring systems where low-latency inference on edge devices is essential. Despite these positive outcomes, certain limitations remain. Misclassifications primarily occurred in images with extreme occlusions, low object visibility, or cluttered backgrounds. The current model is also limited to detecting a single species class. For practical applications in biodiversity monitoring, the model must be extended to handle multi-class classification involving several sympatric fish species. Additionally, performance under night-time or artificial lighting scenarios was not evaluated and warrants further investigation.

Future improvements may involve the incorporation of temporal models such as CNN-LSTM hybrids to capture movement patterns or behavior cues. Transformer-based architectures or attention-guided segmentation methods could also be explored to improve robustness in highly occluded scenes. Expanding the training dataset to include multi-angle views, species diversity, and extreme weather variations would further enhance model adaptability. Finally, this research demonstrates a scalable and transferable approach to aquatic biodiversity monitoring. The pipeline—combining preprocessing, hybrid CNNs, attention mechanisms, and transfer learning—can be applied to various freshwater ecosystems globally. It provides a viable tool for conservation efforts, enabling continuous, automated, and non-invasive monitoring of fish populations in natural habitats.

CONCLUSION

This study proposed and validated a hybrid Convolutional Neural Network (CNN) model designed to detect Red Devil fish (Amphilophus labiatus) in low-quality underwater imagery from Lake Toba, Indonesia. By integrating MobileNetV2 and ResNet50 backbones with attention mechanisms and visual preprocessing techniques such as CLAHE and median filtering, the model achieved state-of-the-art performance across various image conditions. The hybrid model demonstrated high accuracy (96.1%) and mAP@0.5 (0.941), outperforming traditional models such as YOLOv5 and Faster R-CNN. In addition to its strong detection accuracy, the proposed method proved robust under varying image qualities, capable of handling noise, turbidity, and occlusion—key challenges in real-world aquatic monitoring. Its computational efficiency also enables deployment on low-power edge devices, making it suitable for continuous in-field application. The findings of this study contribute to the broader effort in developing intelligent and non-invasive solutions for freshwater biodiversity conservation. While limitations such as single-species focus and occlusion sensitivity remain, the framework established here provides a solid foundation for future enhancements involving multi-species detection, temporal modeling, and transformer-based attention. Overall, the hybrid CNN model proposed in this research offers an effective, scalable, and field-deployable tool for freshwater ecosystem monitoring. Its application is not only limited to Lake Toba but can be generalized to similar environments globally, supporting real-time biodiversity tracking and ecological data acquisition

REFERENCES

[1]. K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.
[2]. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual explanations from deep networks via gradient-based localization,” in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 618–626.
[3]. A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, Jun. 2017.
[4]. M. Tan and Q. V. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” in *Proc. 36th Int. Conf. Machine Learning (ICML)*, 2019, pp. 6105–6114.
[5]. O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in *Proc. Medical Image Comput. Comput.-Assisted Intervention (MICCAI)*, 2015, pp. 234–241.
[6]. D. Liu, Y. Guo, J. Zhao, and Y. Fu, “Automatic fish classification with deep convolutional neural networks,” *Neurocomputing*, vol. 275, pp. 1100–1108, Jan. 2018.
[7]. A. M. Abdelrahman and R. Wang, “Deep learning for fish classification in underwater images,” *Sensors*, vol. 21, no. 8, pp. 2823–2835, Apr. 2021.
[8]. [8] H. Fu, B. Zhang, and C. Xu, “Fish recognition based on deep learning,” *Applied Sciences*, vol. 9, no. 10, pp. 1964–1975, May 2019.
[9]. M. Jalal, M. S. M. Saad, and R. A. Jamil, “Fish species recognition using deep learning,” in *Proc. IEEE Int. Conf. Smart Comput. (SMARTCOMP)*, 2020, pp. 201–206.

[10]. T. Hossain, M. Moniruzzaman, and S. M. S. Islam, "A robust CNN-based fish classification approach with Grad-CAM visualization," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2022, pp. 1601–1605.
[11]. D. Zhang, L. Yang, and Q. Zhang, "Deep learning-based fish species classification," *Sustainable Computing: Informatics and Systems*, vol. 28, p. 100422, Mar. 2020.
[12]. S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
[13]. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.
[14]. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, Cambridge, MA, USA: MIT Press, 2016.
[15]. L. Deng and D. Yu, "Deep learning: Methods and applications," *Foundations and Trends in Signal Processing*, vol. 7, no. 3–4, pp. 197–387, 2014.
[16]. D. Erhan, Y. Bengio, A. Courville, and P. Vincent, "Visualizing higher-layer features of a deep network," *University of Montreal*, Tech. Rep. 1341, 2009.
[17]. Y. Yuan, Z. Wang, and Q. Liu, "A review of visual interpretability for deep learning," *Neurocomputing*, vol. 421, pp. 70–79, Jan. 2021.
[18]. S. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Computers and Electronics in Agriculture*, vol. 147, pp. 70–90, Apr. 2018.
[19]. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Frontiers in Plant Science*, vol. 7, p. 1419, Sep. 2016.
[20]. G. Lin, A. Milan, C. Shen, and I. Reid, "RefineNet: Multi-path refinement networks for high-resolution semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 1925–1934.
[21]. Yanto, B., & Sari, R. P. (2019). Elektronik Pembelajaran Semester (E-RPS) Berbasis Web Fakultas Ilmu Komputer Universitas Pasir Pengaraian. Riau Journal of Computer Science, 5(2), 98-107.
[22]. Sugiyono, D. (2010). Metode Penelitian Kuantitatif, Kualitatif, dan R&D. Penerbit Alfabeta.
[23]. Moleong, L. J. (2019). Metodologi Penelitian Kualitatif (Edisi Revisi). PT. Remaja Rosda Karya.
[24]. Yanto, B., Rouza, E., Fimawahib, L., Hayadi, B. H., & Pratama, R. R. (2023). Penerapan Algoritma Deep Learning CNN dalam Menentukan Kematangan Buah Jeruk Manis Berdasarkan Citra RGB. Jurnal Teknologi Informasi dan Ilmu Komputer, 10(1), 59-66.
[25]. Yanto, B., Hutagaol, R., & Rahman, R. (2022). Analisis optimasi algoritma backpropagation momentum dalam memprediksi jenis tingkat kejahatan di Kecamatan Tambusai Utara. Journal of ICT Applications and System, 1(1), 47-60.
[26]. Yanto, B., Fimawahib, L., Supriyanto, A., Hayadi, B. H., & Pratama, R. R. (2021). Klasifikasi Tekstur Kematangan Buah Jeruk Manis Berdasarkan Tingkat Kecerahan Warna dengan Metode Deep Learning CNN. Jurnal Inovtek Polbeng Seri Informatika, 6(2), 259-268.