ISSN: 2633-4828 Vol. 5 No.4, December, 2023

International Journal of Applied Engineering & Technology

# CLASSIFICATION OF UNDERWATER MINES WITH CONVOLUTIONAL NEURAL NETWORKS

**Manonmani Srinivasan, Shanta Rengaswamy and Ganesh N Naik**
Rashtreeya Vidyalaya College of Engineering, Bengaluru, Karnataka, India Visvesvaraya Technological University, Belagavi, Karnataka, India
manonmanis@rvce.edu.in, shantharangaswamy@rvce.edu.in and ganeshnnaik.cs22@rvce.edu.in

## ABSTRACT
*Underwater mine classification utilizing Convolutional Neural Networks (CNNs) represents a cutting-edge approach to deep learning for maritime security. CNNs, known for their effectiveness in image processing tasks, are adapted to analyze sonar or imaging data for the identification and categorization of underwater mines. This methodology involves the CNN's training process on diverse datasets containing images or features associated with different mine types. The network learns to automatically extract hierarchical and spatial features, enabling it to discern subtle patterns indicative of various mine classes. The application of CNNs in underwater mine classification aims to improve the precision and effectiveness of this process, contributing to improved maritime safety and security by providing automated, real-time mine detection capabilities.*

*Keywords: CLAHE, Convolutional Neural Networks, RVUMR-14 (RV College of Engineering Underwater Mine Research), YOLOv8.*

## INTRODUCTION
Lately, Convolutional Neural Networks (CNNs) have emerged as a powerful and versatile class of deep learning models, revolutionizing the field of computer vision and making significant contributions to various other fields like natural language processing and medical image analysis represent specific areas of expertise. Developed and inspired by the visual processing capabilities of the human brain, CNNs have demonstrated unparalleled success in tasks such as image classification, object detection, and image segmentation.

Traditional neural networks, while effective for many tasks, face challenges in handling high-dimensional data like images. CNNs address this limitation by introducing convolutional layers that enable the networks to learn hierarchical representations of patterns in the input data automatically. This architectural innovation has proven particularly adept at capturing spatial hierarchies and local dependencies, making CNNs well-suited for tasks involving grid-like data such as images and, more recently, even sequential data like time series and text.

CNNs have revolutionized object detection by offering a durable and effective framework for automated visual recognition. Unlike traditional methods that rely on handcrafted features. CNNs autonomously acquire hierarchical representations of features starting from unprocessed pixel data, enabling them to discern complex patterns and objects in images. The application of CNNs in object detection involves the utilization of convolutional layers to localize and identify objects inside an image. These networks can effectively capture spatial hierarchies and relationships between pixels, allowing for precise localization and classification of objects. One of the major advantages lies in their ability to handle varying scales and orientations of objects, making them highly suitable for real-world scenarios. State-of-the-art architectures like Faster R-CNN, YOLO (You Only Look Once), and SSD (Single Shot MultiBox Detector) leverage CNNs to achieve remarkable accuracy and real-time performance, making CNN-based object detection a pivotal technology with a spectrum of applications from autonomous vehicles and surveillance systems to healthcare and industrial automation.

Our research leverages the cutting-edge YOLO v8 (You Only Look Once) model, representing the pinnacle of object detection capabilities in the field of computer vision. YOLO v8 stands out as a state-of-the-art model, embodying the latest advancements in advanced learning techniques and convolutional neural networks. Renowned for its efficiency and accuracy, YOLO v8 adopts a unified approach, enabling simultaneous object localization and classification within a single pass through the network. This model excels in real-time applications, making it a paramount choice for scenarios where rapid and precise object detection is critical. By

employing YOLO v8, our research aims to capitalize on its state-of-the-art features to enhance the accuracy and efficiency of object detection in our specific domain, contributing to the forefront of contemporary computer vision research and applications.

## RELATED WORKS

In the paper [1], the study focuses on detecting marine mines using images captured by cameras mounted on various platforms such as drones, submarines, ships, and boats. Due to a limited dataset, images were sourced from the Internet and augmented to create two datasets—one for floating mines and another for underwater mines. The detection models employed include YOLOv5 and SSD for floating mines and YOLOv5 and EfficientDet for underwater mines. Additionally, the performance of these models was evaluated on an IoT device, specifically the Raspberry Pi 4 with an attached camera.

In paper [2], the research explores the application of YOLO, SSD, and EfficientDet utilizing deep learning models to identify sea mines, addressing safety concerns in navigation amid geopolitical tensions and armed conflicts. The study utilized augmented and synthetically generated datasets for both floating and underwater mines, achieving high accuracy in object recognition. The feasibility of real-time deployment on portable devices, such as Raspberry Pi, was also demonstrated, with a frame processing time of 2 seconds, potentially improved with high-performance cameras.

In paper [3], implemented in the PYTHON platform, the proposed method demonstrates improved performance on the DOTA dataset, surpassing other deep learning networks with notable metrics such as Mean Average Precision (mAP = 85.5%), F1-score (83.78%), Detection rate (97%), and Precision-Recall curve.

In the paper [4], The survey systematically reviews various object detection methods, categorizing them into one-stage and two-stage detectors. Additionally, it outlines traditional and emerging applications of object detection. Representative branches within the field are also explored. The survey concludes by discussing the architectural aspects of integrating these detection methods for building effective and efficient systems. Furthermore, it highlights development trends to align with the latest algorithms and suggests areas for future research.

In paper [5], The deeper network architecture of YOLO-v5 + R-FCN enhances feature extraction capabilities, leading to more accurate feature recognition and improved detection performance for densely arranged target images in remote sensing data. This study offers valuable perspectives for the application of remote sensing technology in China, promoting the use of satellites for target detection tasks in relevant fields and in paper [6] improved YOLOv3 was used for implementation.

This research [7] aimed at enhancing facial recognition accuracy, authors suggested employing adaptation techniques to adjust the surrounding light intensity during image capture. In this study, authors utilized the Contrast Adaptive Limited Histogram (CLAHE) method as a means of adaptation.

This paper [8] provides a comprehensive review of mine detection and classification techniques employed in various systems. The author examines both current and previous-generation methods, beginning with classical image processing, progressing to machine learning, and culminating in deep learning approaches.

In this paper [9], Employing traffic signs as a case study, Authors developed image degradation models utilizing the YOLO network. These models incorporate traditional image processing techniques to replicate real-world shooting challenges. Following the establishment of diverse degradation models, They assessed their impact on object detection effectiveness. Leveraging the YOLO network.

This paper [10] presents a comprehensive review of deep learning-based object detection frameworks. The authors commence with a succinct overview of the evolution of deep learning and its prominent tool, the convolutional neural network. Subsequently, we delve into various generic object detection architectures, highlighting modifications and effective strategies aimed at enhancing detection performance.

This paper [11] aims to streamline the process of learning a CNN architecture by exploring the correlation between Fully Connected (FC) layers and various characteristics of datasets. CNN architectures, as well as datasets, are often classified based on their depth, shallowness, width, etc.

In this paper [12], Introducing a novel hierarchical image classification model, Condition-CNN, we aim to overcome certain limitations observed in the branching convolutional neural network, particularly in training time and fine-grained accuracy.

This paper [13] compares the Depth Confidence Network algorithm with the Convolutional Neural Network algorithm in experiments related to image positioning and recognition. Through several iterative experiments, it was observed that in the context of large package images, the Convolutional Neural Network algorithm achieved an average image recognition accuracy of 96.09%, whereas the Depth Confidence Network algorithm achieved 94.42%.

The investigation detailed in this paper [14] focuses on employing automated classification techniques to discern the presence and identify the types of vessels operating within coastal areas through analysis of underwater acoustic signals. We assessed multiple configurations of deep convolutional neural network architectures along with preprocessing filter layers. This evaluation utilized a novel dataset derived from a portion of the vast open-source Ocean Networks Canada hydrophone data.

In this review paper [15] authors have concentrated on underwater crack detection system using various methods such as Manual Visual Inspection, Intelligent Monitoring Techniques, Digital Image Detection Methods and also discussed Dam Crack Detection Based on Underwater Robots.

In this paper [16], authors implemented four models such as Random forest, SVM, feed-forward neural network, and CNN to differentiate surface and underwater vessels in the ocean using low-frequency acoustic presser data. In this CNN is giving better accuracy with least error rate.

In this paper [17] authors have compared two algorithms namely SVM and CNN for machine learning and deep learning algorithms for image classification. For smaller datasets traditional machine learning algorithms are better than deep learning algorithms. For larger datasets deep learning algorithms are giving better accuracy.

In this paper[18] author has explained and summarized all YOLO versions from V1 to V8 and also Scaled, PP, DAMO, NAS versions from years 2015 to 2023 and also major changes in each model and training tricks also discussed.

In this proposed model [19] author has used the YOLOv5 model to identify vision loss in the initial stage and also to identify these diseases in the early stage. In this, they have used the Diabetic Retinopathy dataset to train and test the model and they got superior results with respect to mAP, F1-score, and IOU.

**PROPOSED METHODOLOGY**

In this proposed work shown in Figure 1, we are utilizing the YOLO v8 model for the detection of different mine classes within the RVUMR-14 dataset. The dataset comprises variations with and without an underwater background, and the underwater set undergoes preprocessing using the CLAHE algorithm. We aim to evaluate the efficiency of three datasets: one with no background alterations, one with an original underwater background, and one with a CLAHE-processed underwater background. Through comprehensive evaluation metrics such as F1 score, Precision, Recall, and Mean Average Precision during training and validation, we aim to identify the dataset that, when combined with the YOLO v8 model, delivers optimal results for effective mine detection in diverse underwater conditions.

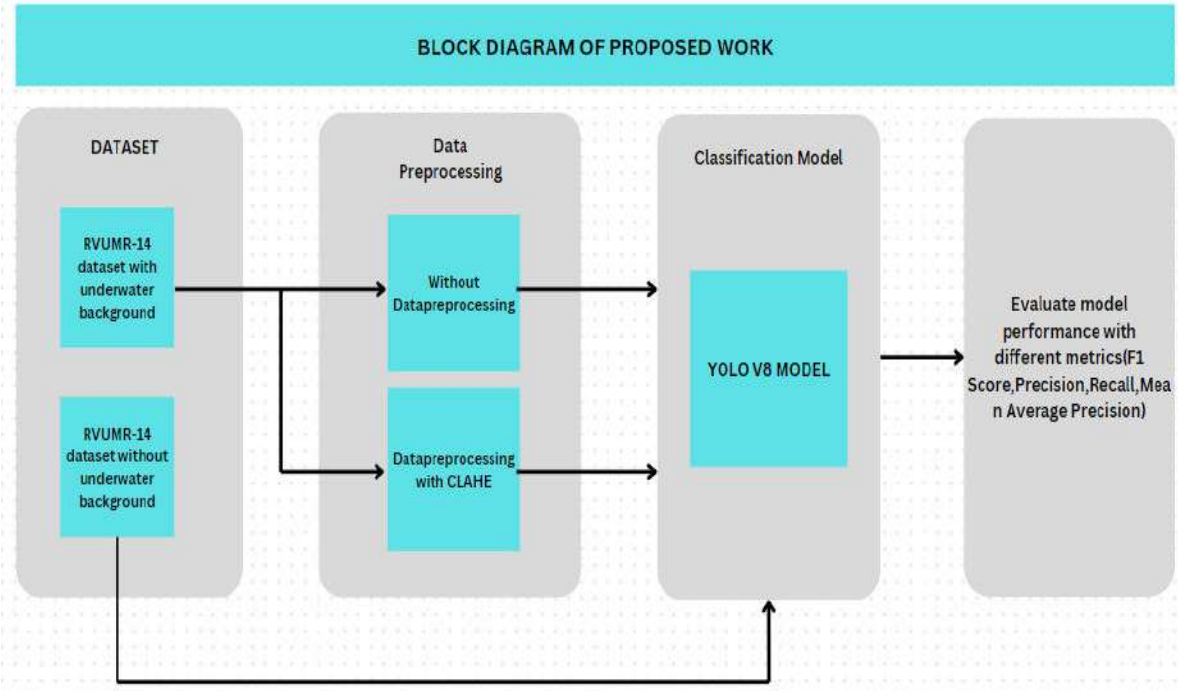


**Figure 1:** Block Diagram of Proposed Work

## YOLO ARCHITECTURE

YOLO, or "You Only Look Once," is a family of real-time object detection algorithms commonly employed in computer vision and deep learning. The YOLO algorithm divides an image into a grid and performs object detection on each grid cell simultaneously. The main advantage of YOLO is its ability to detect objects in real-time with high accuracy.

Below are several critical aspects about YOLO:

1. YOLOv1 (You Only Look Once version 1): The original YOLO algorithm was introduced by Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi in a paper published in 2016. YOLOv1 divided the input image into a grid and predicted both bounding boxes and class probabilities for every grid cell in a singular forward pass.
2. YOLOv2 (YOLO9000): YOLOv2, also known as YOLO9000, was introduced to address the limitation of detecting a limited set of object classes. YOLO9000 was capable of detecting over 9000 object categories by using a hierarchical approach.
3. YOLOv3: YOLOv3 [6], released in 2018, further improved the accuracy of object detection. It introduced several architectural changes, including the use of three different scales for detection, making it more robust in handling different object sizes.
4. YOLOv4: YOLOv4, introduced in 2020, focused on improving speed and accuracy. It incorporated features such as the CSPDarknet53 backbone, PANet, and other optimizations to achieve state-of-the-art performance.
5. YOLOv5 further improved the model's performance and added new features such as hyperparameter optimization, integrated experiment tracking and automatic export to popular export formats.
6. YOLOv6 was open-sourced by Meituan in 2022 and is in use in many of the company's autonomous delivery robots.
7. YOLOv7 added additional tasks such as pose estimation on the COCO key points dataset.

8. YOLOv8 is the latest version of YOLO by Ultralytics. As a cutting-edge, state-of-the-art (SOTA) model, YOLOv8 builds on the success of previous versions, introducing new features and improvements for enhanced performance, flexibility, and efficiency. YOLOv8 supports a comprehensive array of vision AI assignments, encompassing detection, segmentation, and pose estimation, tracking, and classification. This versatility allows users to leverage YOLOv8's capabilities across diverse applications and domains.

YOLO typically consists of a CNN architecture. The network predicts bounding boxes, objectness scores, and class probabilities directly from the raw pixels of the input image.

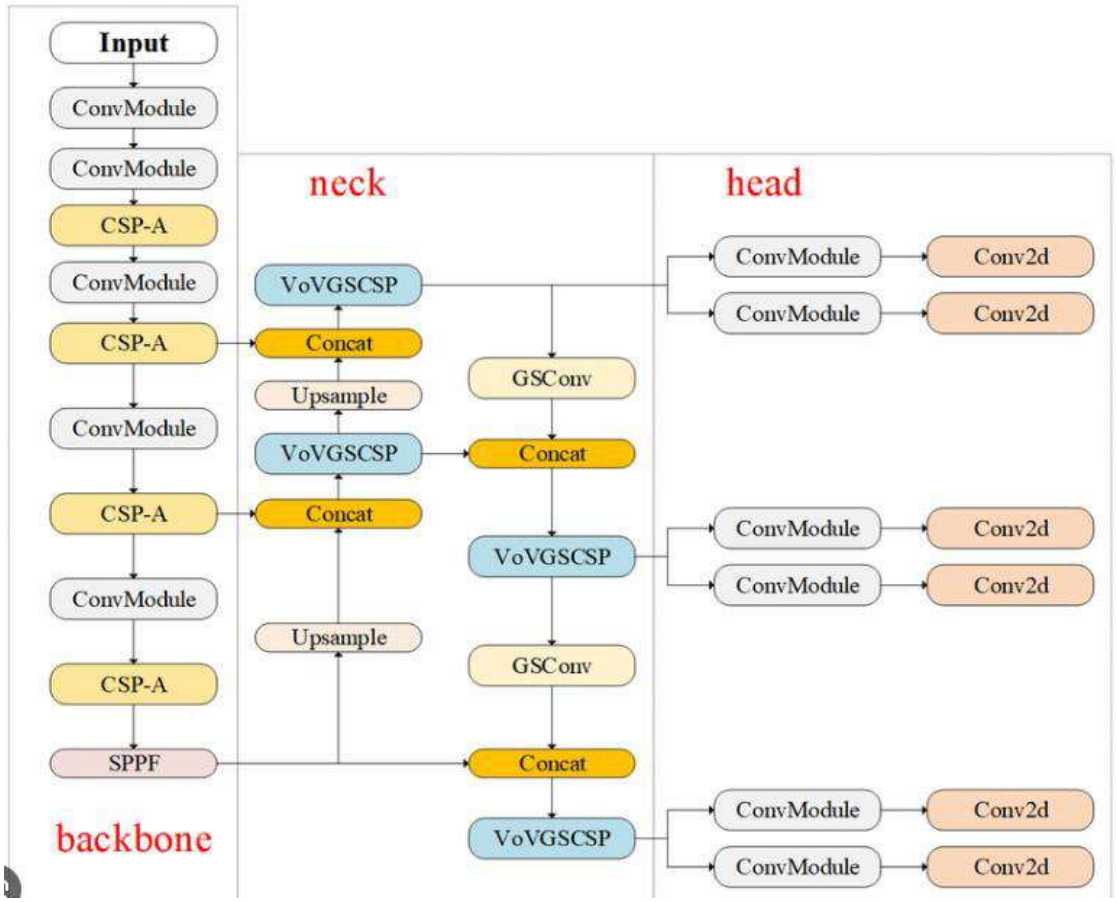


**Fig. 2:** Architecture of YOLO v8

The overall architecture of YOLOv8 is shown in Figure 2. It is divided into three main parts:

1. **Backbone:** The backbone is the primary body of the network, responsible for extracting features from the input image. YOLOv8 uses a modified version of the Darknet53 architecture, known as CSPDarknet53, as its backbone.
2. **Neck:** The neck is responsible for connecting the backbone to the head and for processing the features extracted by the backbone. YOLOv8 uses a novel C2f module instead of the traditional YOLO neck architecture. The C2f module is designed to improve the flow of information between different levels of the network.
3. **Head:** The head is responsible for anticipating the bounding boxes and class probabilities for each object in the image. YOLOv8 uses a modified version of the YOLOv5 head architecture, which uses a spatial attention mechanism to improve detection accuracy.

The combination of these three components makes YOLOv8 a powerful and versatile object detector capable of being utilized for a wide variety of tasks, including real-time object detection, video surveillance, and autonomous vehicles.

## OVERVIEW OF OUR PROPOSED APPROACH

### 5.1 Dataset

Creating a custom dataset for underwater mine detection involves a meticulous and multifaceted process that aims to train a robust and accurate model. The initial step entails the careful collection of diverse images, encompassing different classes such as underwater scenes featuring mines and those devoid of any mine presence. The present study introduces RVUMR-14, a tailor-made dataset crafted specifically to tackle the classification challenges posed by underwater mines. Addressing the scarcity of publicly available underwater mine images, this dataset offers a comprehensive collection of annotated photographs for both training and evaluation purposes. Encompassing images of 14 distinct types of underwater mines commonly encountered in marine environments, RVUMR-14 aims to fill a critical gap in the available resources for mine classification research. To address the issue of dataset size, diverse augmentation techniques were implemented to expand the number of images per classification category. This diversity is paramount to ensuring that the model can adeptly discern between various underwater scenarios and effectively identify the presence or absence of mines. Following this, meticulous class labeling becomes imperative, as each image must be annotated with the corresponding class label, serving as a foundational element for supervised learning. The sample underwater image dataset is shown in Figure 3.

**Fig. 3** Sample images of RVUMR-14 Naval Mine dataset

To further enhance the dataset's variability and resilience, a series of data augmentation methods have been systematically applied. These augmentation methods, including but not limited to inverting and flipping, play a central role in artificially expanding the dataset. By subjecting the images to transformations such as rotation, scaling, cropping, and adjustments in brightness and contrast, the model becomes adept at recognizing mines

under a myriad of conditions. The judicious use of data augmentation is instrumental in preventing overfitting and ensuring the model's ability to generalise effectively to novel, unseen data. and ensuring the model's ability to generalise effectively to novel, unseen data.

Once the augmentation process is complete, the dataset is judiciously partitioned into distinct sets for training, validation, and testing. This segregation is essential for evaluating the model's performance on unseen data, fine-tuning hyperparameters during the validation phase, and ultimately assessing the model's generalization capabilities. Additionally, the inclusion of underwater images without mines in the dataset serves to reinforce the model's capability to differentiate between favorable and unfavorable instances, contributing to its overall efficacy in real-world applications. Through this comprehensive approach to dataset creation, the resulting model is poised to exhibit a high level of accuracy, adaptability, and reliability in underwater mine detection scenarios.

### 5.2 Training Model with No Background RVUMR-14 Dataset

To assess the model's accuracy and compare its performance across different scenarios, we are training it specifically on images featuring isolated mines without background context. The sample image dataset is shown in Figure 4. This focused subset serves as a benchmark, allowing a direct comparison with the performance of the model regarding the complete dataset. By evaluating the models ability to discern mines without contextual information, we gain insights into its robustness and adaptability in diverse underwater environments. This approach provides a nuanced understanding of model's sensitivity to background variations, ultimately contributing to the refinement of its capabilities for effective underwater mine detection.

**Fig. 4:** RVUMR-14 Naval Mine dataset without background

### 5.3 Removing Noise from RVUMR-14

To boost the accuracy of our underwater mine detection model, we're focusing on noise reduction in the dataset. Employing the Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm, we effectively address unwanted variations in illumination and other distortions. CLAHE optimizes image quality by enhancing local contrast, emphasizing mine features, and suppressing irrelevant noise. This strategic preprocessing step aims to fortify the model against challenges in diverse underwater conditions, fostering improved precision and adaptability in mine detection scenarios.

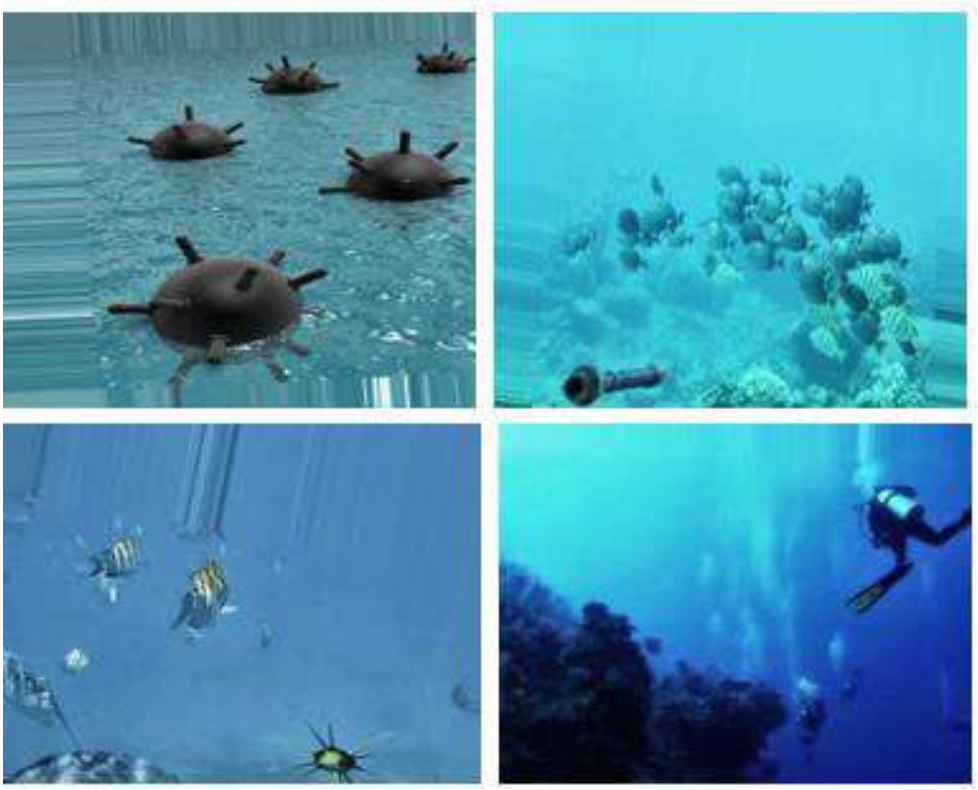

**Fig. 5** RVUMR-14 Naval Mine dataset without noise

### 5.4 CLAHE Algorithm

The Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm stands out as a sophisticated image processing technique renowned for its ability to dynamically enhance local contrast, making it a crucial tool in various applications, particularly in the realms of medical imaging and computer vision. In contrast to traditional histogram equalisation methods, CLAHE operates by dividing an image into smaller, non-overlapping tiles, effectively localizing the enhancement process. This adaptive approach proves invaluable in scenarios where lighting conditions are uneven or when there's a need to highlight specific details within distinct regions of an image.

The brilliance of CLAHE rests upon its capability to independently apply histogram equalisation to each tile, catering to the unique characteristics of different areas within the image. By doing so, CLAHE avoids the common pitfall of over-amplifying contrast in brighter regions, a drawback associated with traditional histogram equalisation. Through the use of a specified threshold, CLAHE judiciously limits contrast amplification, ensuring that the enhancement process is controlled and tailored to the specific requirements of each local region.

This algorithm finds widespread utility in medical imaging, where revealing subtle structures in X-rays or MRIs is paramount, and in computer vision applications, where nuanced image analysis demands optimal contrast. CLAHE's adaptability to local features not only enhances visual quality but also proves instrumental in minimizing the impact of noise, making it an indispensable tool for refining image quality and facilitating subsequent image-processing tasks. In essence, CLAHE's nuanced and adaptive contrast enhancement capabilities position it as a cornerstone in the arsenal of image processing techniques, contributing significantly to the advancement of accurate and robust visual analyses. The step-by-step procedure of the CLAHE algorithm is shown in figure 5.

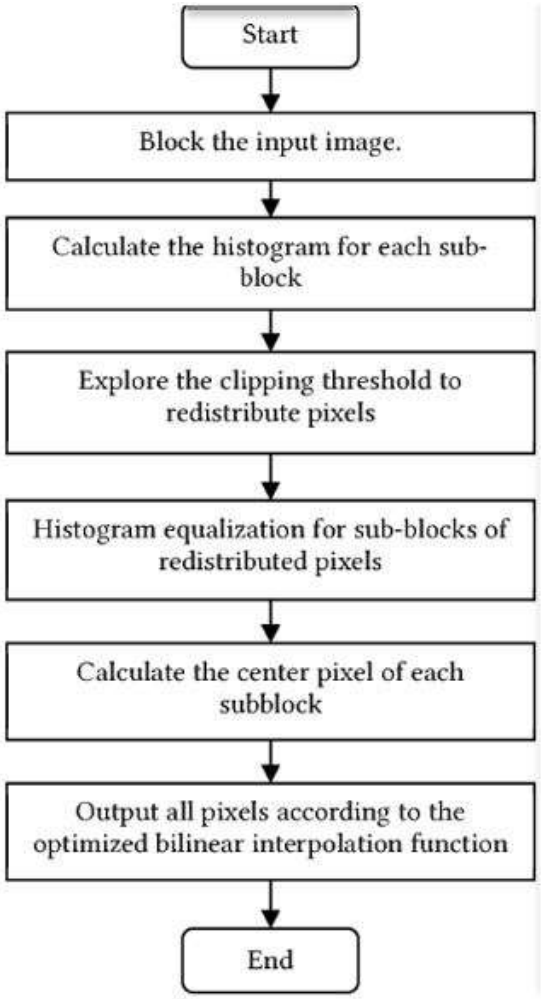


**Fig.5** Flowchart of CLAHE algorithm

The comparison features the original image on the left and the output image after applying Contrast Limited Adaptive Histogram Equalization (CLAHE) on the right is shown in Figure 6. The original image provides an unprocessed view, capturing inherent details and lighting conditions. In contrast, the CLAHE-enhanced image on the right demonstrates the algorithm's impact, selectively amplifying contrast in different regions.

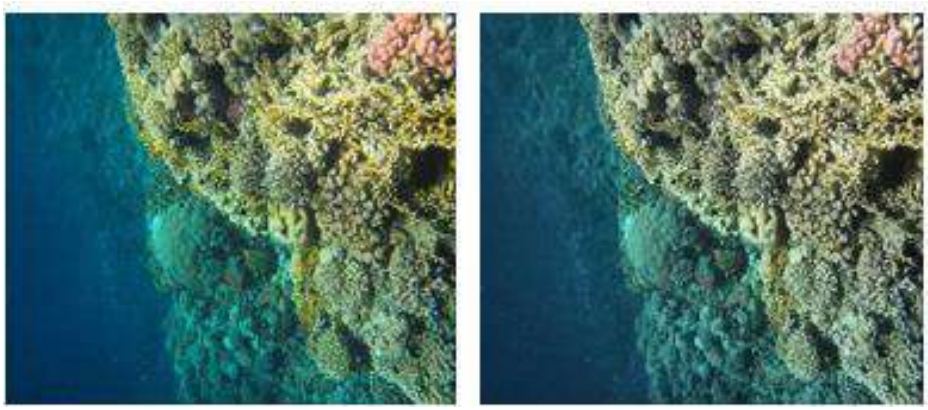

**Fig. 6** Comparison of CLAHE output with the original image

This process unveils finer details and improves overall visual clarity. The side-by-side display succinctly illustrates how CLAHE enhances local contrast, addressing uneven lighting issues and optimizing image quality for improved information extraction.

## EXPERIMENT

### 6.1. Specifications

In the execution of this task, our computational infrastructure was powered by the formidable Nvidia RTX 2060 Super GPU in conjunction with the AMD Ryzen 7 3700x CPU. This hardware synergy was instrumental in providing the computational prowess necessary for the intricate processes involved in our model training. Leveraging the cutting-edge capabilities of the TensorFlow 2x version, our model benefited from a robust and efficient deep learning framework. The utilization of the specified hardware configuration, comprising the Nvidia RTX 2060 Super GPU and the AMD Ryzen 7 3700x CPU, ensured accelerated processing speeds and parallelized computations, thereby significantly expediting the model training process. The model was meticulously trained for 10 epochs, a strategic decision aimed at striking a balance between training time and achieving convergence for optimal performance. This hardware and software orchestration reflects an intentional choice customized for the specific requirements of the task, underscoring the importance of a well-matched computational environment in the pursuit of successful model training and development.

### 6.2 Assessment of Indicators

In evaluating the outcomes and conducting a thorough analysis of our dataset, we rely on a comprehensive set of performance indicators, namely F1 score, Precision, Recall, and Mean Average Precision (mAP). Each of these metrics plays a distinct yet complementary role in assessing the efficiency and resilience of our model.

**F1 Score:**

The F1 score, being a harmonic mean of Precision and Recall, offers a balanced assessment that accounts for both false positives and false negatives making it particularly useful when there is an imbalance between the classes. A higher F1 score indicates a better balance between precision and recall, signifying a more accurate model. The Formula for F1 score is shown in equation 1.

**F1 = 2 * (Precision*Recall)/(Precision+Recall)** Eq(1)

**Precision:**

Precision, also known as a positive predictive value, gauges the accuracy of the model concerning the instances it classifies as positive. It is determined by dividing the number of true positives by the sum of true positives and false positives. Precision is crucial when the cost of false positives is high, as it ensures that the positively identified instances are accurate.

**Recall:**

Recall, or sensitivity, assesses the model's ability to capture all the relevant instances of a particular class.It is computed by dividing the number of true positives by the sum of true positives and false negatives. Recall is essential in scenarios where missing positive instances is more critical than false positives. The Formula for recall is shown in equation 2.

**Recall = True Positives / (True Positives + False Negatives)** Eq(2)

Here's a breakdown of the terms in the formula:

- **True Positives (TP):** The count of instances that are positive and were correctly identified as positive by the model.
- **False Negatives (FN):** The count of instances that are positive but were incorrectly identified as negative by the model.

**Mean Average Precision (mAP):**

Mean Average Precision is commonly used in object detection tasks. It considers precision at different levels of recall and computes the average precision over a range of recall values. mAP provides a comprehensive evaluation of the model's performance across varying thresholds, offering insights into how well the model identifies objects at different levels of confidence.

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i$$

By incorporating these metrics into our evaluation process, we gain a nuanced understanding of the model's i Eq(3) strengths and weaknesses. This multifaceted assessment ensures a comprehensive and robust evaluation, guiding further refinement and optimization for superior performance in real-world applications. The Formula for mAP is shown in equation 3.

**RESULT AND ANALYSIS**

Our investigation involves a comprehensive comparison of outcomes generated by the YOLO (You Only Look Once) model across three distinct datasets. The initial set of results emanates from a meticulously curated dataset intentionally crafted to exclude extraneous background elements. This strategic curation isolates objects of interest, providing an optimal environment to evaluate the YOLO model's proficiency in accurately detecting and classifying objects under pristine, uncluttered conditions.

In the second set of results, our focus shifts to a dataset deliberately imbued with noise. This deliberate introduction of visual disturbances emulates real-world scenarios where images may exhibit imperfections and distractions. By subjecting the YOLO model to these challenging conditions, we aim to assess its robustness and adaptability, critical attributes for ensuring reliable performance in diverse and less controlled environments.

The third set of results stems from a dataset subjected to the Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm. This algorithm serves as a sophisticated preprocessing technique designed to remove noise and enhance image quality. By incorporating CLAHE, we seek to highlight the impact of noise reduction on the YOLO model's detection precision and overall effectiveness. This multifaceted analysis aims to offer detailed and subtle perspectives into the YOLO model's capabilities under varying conditions, informing strategies for optimization and deployment in practical applications.

**7.1 Results of RVUMR-14 without underwater background**

**Table 1 –** Precision, Loss. Recall and mAP of RVUMR-14 without underwater background Naval Mine Dataset

| epochs | Train / box_loss | train/cls_ loss | train/dfl_ loss | metrics/p recision | metrics/r ecall | metrics/mA P50 |
|---|---|---|---|---|---|---|
| **1** | 1.0548 | 4.0588 | 1.3125 | 0.00745 | 0.95175 | 0.20251 |
| **2** | 1.0034 | 3.0267 | 1.2434 | 0.54248 | 0.4632 | 0.50712 |
| **3** | 0.9881 | 2.5665 | 1.225 | 0.61731 | 0.68535 | 0.69385 |
| **4** | 0.95657 | 2.1835 | 1.2003 | 0.66683 | 0.70328 | 0.75159 |
| **5** | 0.95786 | 1.9228 | 1.1743 | 0.87838 | 0.78074 | 0.85906 |
| **6** | 0.93904 | 1.7409 | 1.1538 | 0.90815 | 0.84213 | 0.92187 |
| **7** | 0.911 | 1.6121 | 1.1294 | 0.91199 | 0.85302 | 0.91483 |
| **8** | 0.87035 | 1.452 | 1.0984 | 0.9424 | 0.8814 | 0.94015 |
| **9** | 0.85627 | 1.3901 | 1.1099 | 0.96662 | 0.88821 | 0.95185 |
| **10** | 0.83002 | 1.3275 | 1.0898 | 0.9572 | 0.90671 | 0.95989 |

Table 1 presents the outcomes of a model trained without the inclusion of a background dataset. The training process spanned 10 epochs, yielding an accuracy rate of 95.7% and a recall rate of 90.6%. The graphical representation of these results is depicted in the accompanying figure 7, displayed as a line graph.

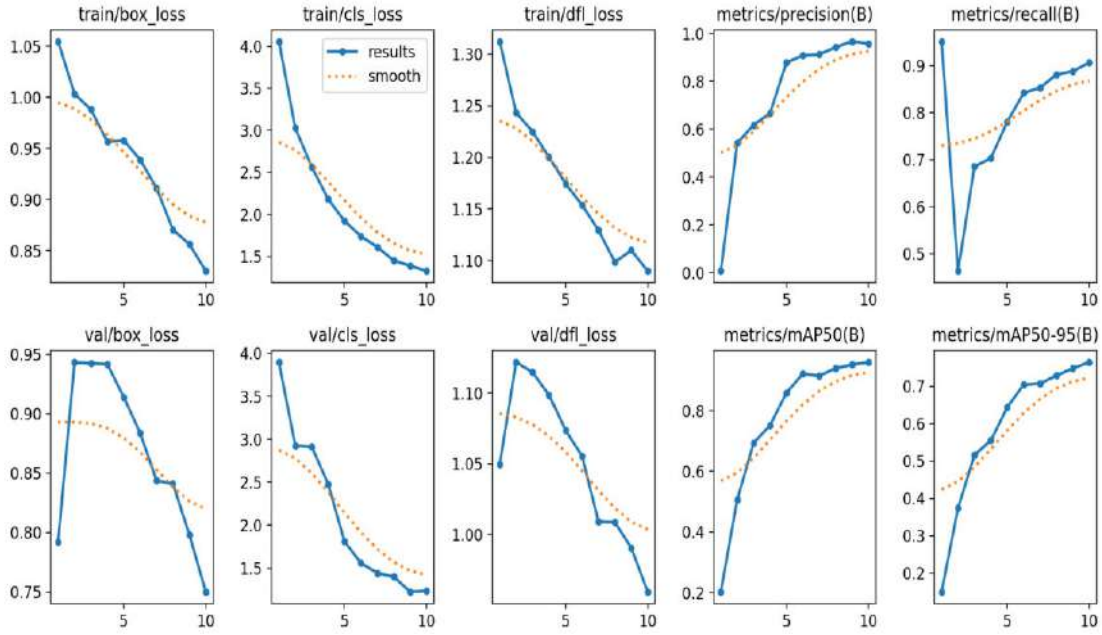


**Fig. 7:** Line graph representation of Precision, Loss. Recall and mAP results of RVUMR-14 without underwater background Naval Mine Dataset

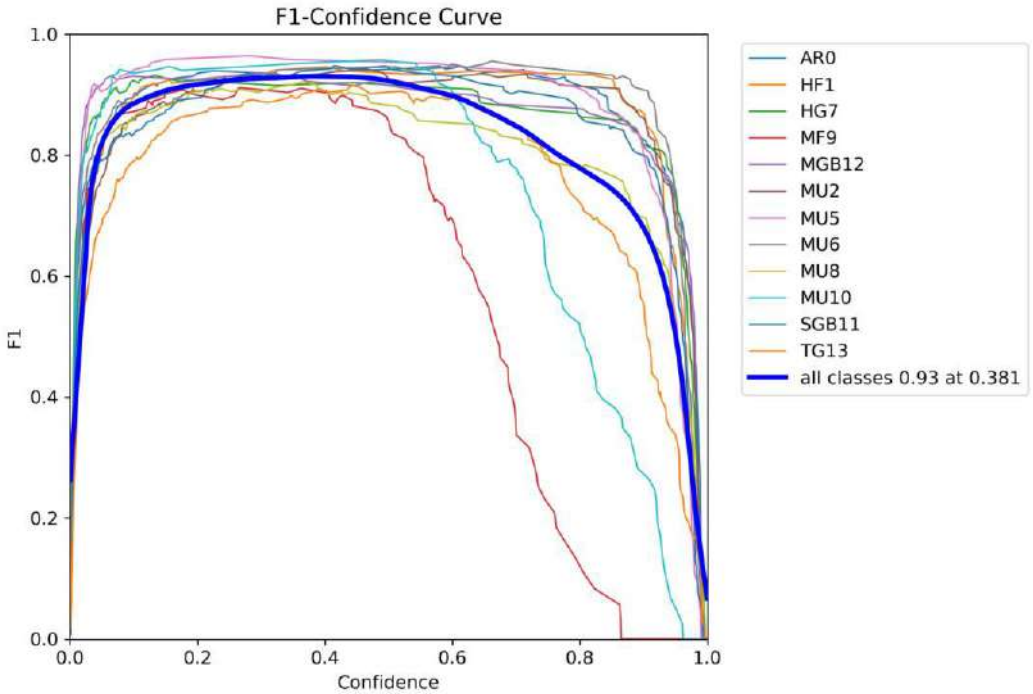


**Fig.8** F1-Confidence curve of RVUMR-14 without underwater background Naval Mine Dataset

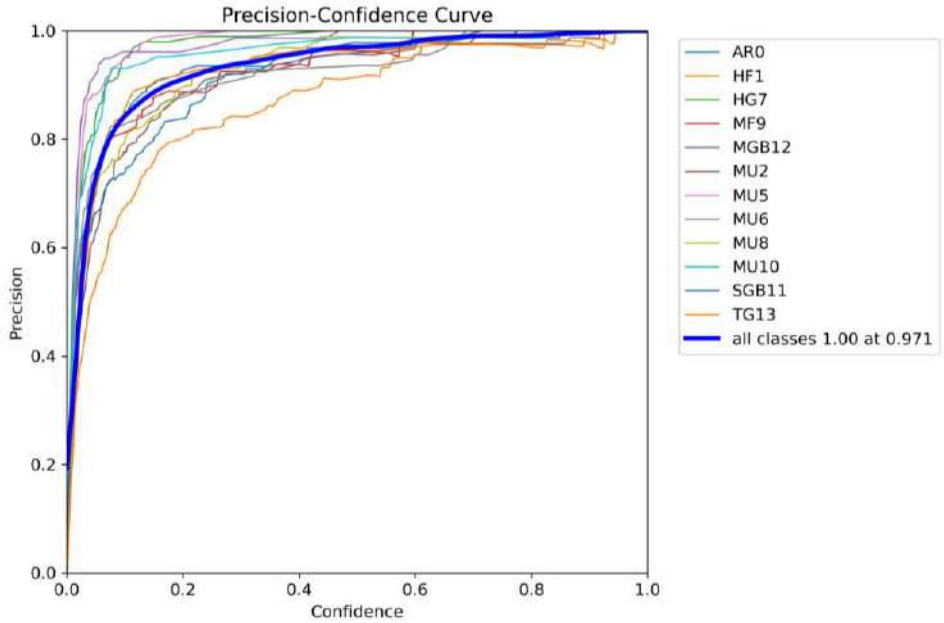


**Fig. 9** Precision-Confidence curve of RVUMR-14 without underwater background Naval Mine Dataset

Figures 8 and 9 illustrate the F1-Confidence curve and the Precision-Confidence curve, respectively. In Figure 8, the F1 score is prominently depicted at 93%, representing the average across all classes of mines. Meanwhile, in Figure 9, the Precision-Confidence curve reaches a notable precision level of 97.1%.

### 7.2 Results of the Original RVUMR-14 with an Underwater Background

**Table 2 -** Precision, Loss. Recall and mAP of RVUMR-14 original Naval Mine Dataset with an underwater background

| epoch | train/box_loss | train/cls_loss | train/dfl_loss | metrics/precision | metrics/recall | metrics/ mAP50 |
|---|---|---|---|---|---|---|
| **1** | 1.202 | 4.4936 | 1.2066 | 0.48494 | 0.15292 | 0.20467 |
| **2** | 1.2231 | 3.4561 | 1.1801 | 0.75167 | 0.477 | 0.69747 |
| **3** | 1.2044 | 2.7414 | 1.1741 | 0.75264 | 0.7411 | 0.82565 |
| **4** | 1.1901 | 2.4294 | 1.1562 | 0.81502 | 0.79527 | 0.86902 |
| **5** | 1.1702 | 2.2558 | 1.1133 | 0.86044 | 0.8777 | 0.94031 |
| **6** | 1.1747 | 2.0492 | 1.1348 | 0.91453 | 0.91914 | 0.96294 |
| **7** | 1.1395 | 1.985 | 1.1045 | 0.96257 | 0.92414 | 0.97811 |
| **8** | 1.0971 | 1.7941 | 1.0703 | 0.97684 | 0.9664 | 0.98583 |
| **9** | 1.0733 | 1.6923 | 1.074 | 0.96736 | 0.96801 | 0.98737 |
| **10** | 1.0387 | 1.548 | 1.0498 | 0.97709 | 0.97702 | 0.99013 |

Table 2 encapsulates the findings derived from the dataset infused with noise. Following a rigorous training regimen spanning 10 epochs, the discerning observations reveal a remarkable performance, with both accuracy and recall registering at an impressive 97.7%. This outcome underscores the model's resilience in handling data imbued with noise, demonstrating its robustness and proficiency in maintaining high accuracy even under challenging conditions. This is shown in figure 10.

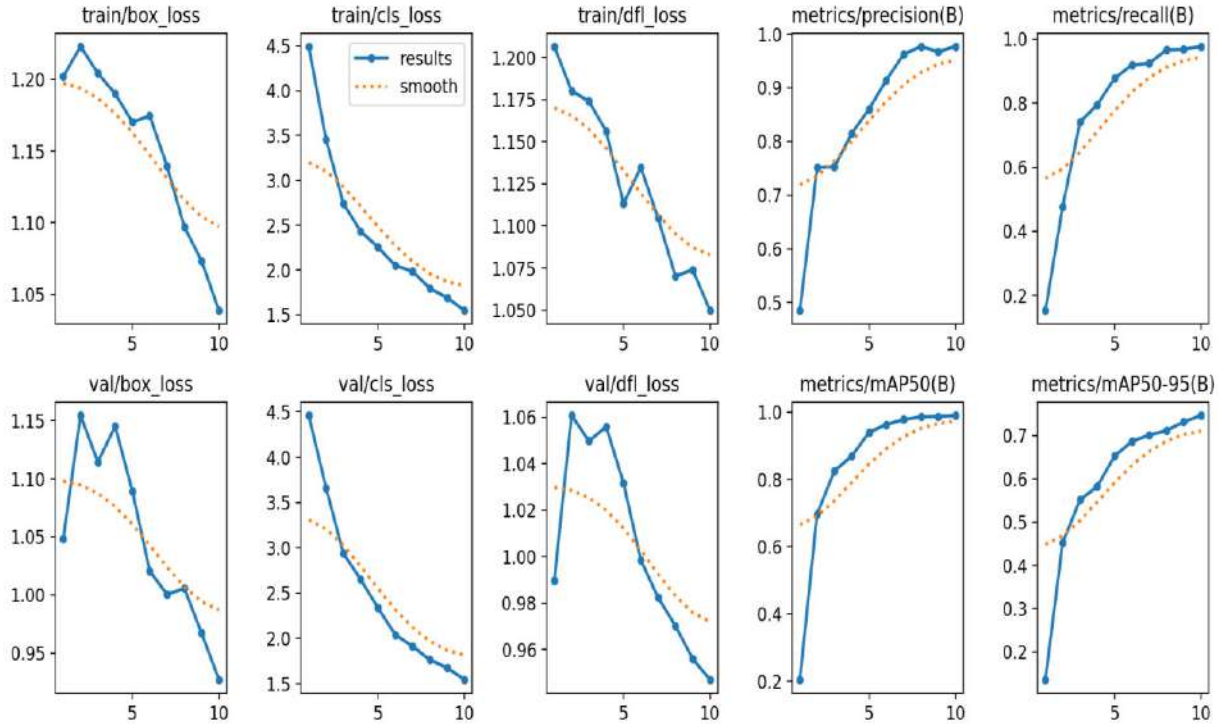


**Fig. 10** Line graph representation of Precision, Loss. Recall and mAP results of RVUMR-14 original Naval Mine Dataset with background

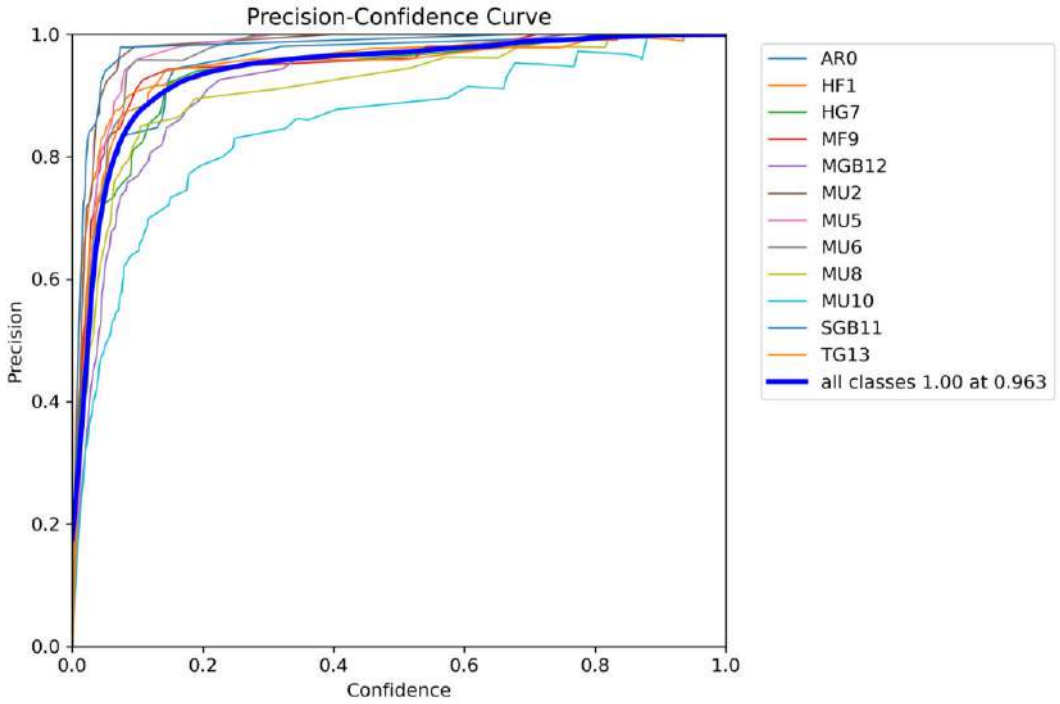


**Fig.11** F1-Confidence curve of RVUMR-14 original Naval Mine Dataset with background

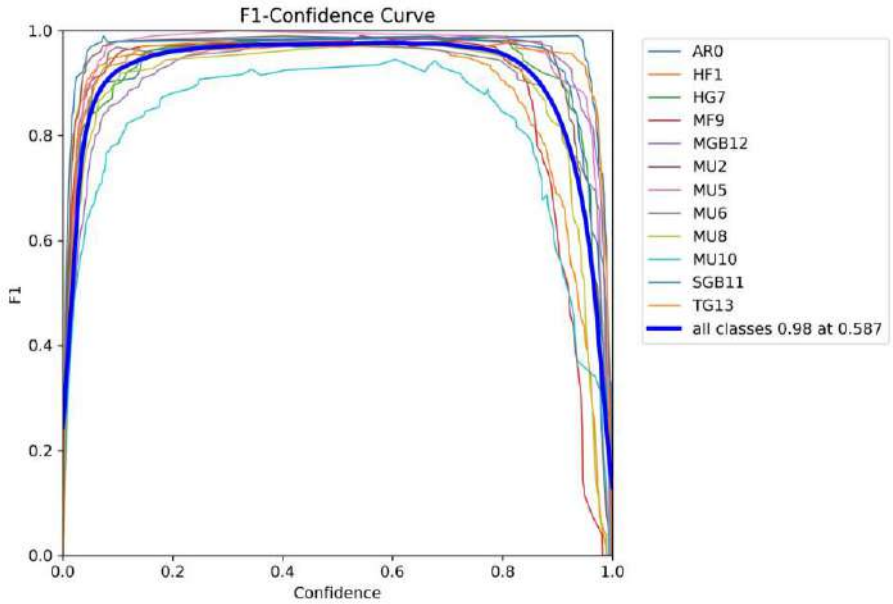


**Fig. 12** Precision-Confidence curve of Naval mine dataset RVUMR-14 original Naval Mine Dataset with background

Figures 11 and 12 provide a visual depiction of the F1-Confidence curve and Precision-Confidence curve, respectively, for the dataset incorporating noise. Noteworthy insights emerge as the F1-Confidence curve attains an impressive 98%, underscoring the model's adeptness in achieving a robust balance between precision and recall even in the presence of noise. Similarly, Figure 12 reveals a Precision-Confidence curve standing firm at 96.3%, affirming the model's capacity to maintain a high level of precision across varied confidence levels.

### 7.3 Results of RVUMR-14 with CLAHE Underwater Background

**Table-3** Precision, Loss. Recall and mAP of RVUMR-14 Pre-processed Naval Mine Dataset with an underwater background

| epoch | train/box_loss | train/cls_loss | train/dfl_loss | metrics/precision | metrics/recall | metrics/mAP50 |
|---|---|---|---|---|---|---|
| **1** | 1.1992 | 4.4258 | 1.2067 | 1 | 0.03582 | 0.169 |
| **2** | 1.2342 | 3.4768 | 1.1998 | 0.67414 | 0.46451 | 0.59526 |
| **3** | 1.2157 | 2.815 | 1.1907 | 0.7554 | 0.71762 | 0.82842 |
| **4** | 1.2027 | 2.4129 | 1.1867 | 0.81082 | 0.82476 | 0.87515 |
| **5** | 1.2005 | 2.2973 | 1.154 | 0.80756 | 0.8586 | 0.90811 |
| **6** | 1.1676 | 2.1068 | 1.1517 | 0.90781 | 0.91903 | 0.94734 |
| **7** | 1.1523 | 2.0252 | 1.126 | 0.89452 | 0.86356 | 0.95388 |
| **8** | 1.1614 | 1.8097 | 1.1129 | 0.96174 | 0.95818 | 0.98724 |
| **9** | 1.1025 | 1.6908 | 1.106 | 0.96753 | 0.97158 | 0.99195 |
| **10** | 1.0695 | 1.5662 | 1.0707 | 0.98558 | 0.97734 | 0.9916 |

Table 3 showcases the outcomes derived from the model trained on a dataset devoid of noise, following a rigorous 10-epoch training regimen. Notably, the model demonstrates exceptional performance with an accuracy rate of 98.5% and a recall rate of 97.73%. These results surpass the corresponding metrics observed in the dataset with noise, highlighting the model's heightened accuracy and proficiency when operating in a noise-free environment. The superior precision and recall metrics underscore the model's precision in distinguishing between classes and

its ability to effectively retrieve relevant instances from the dataset. This robust performance on the clean dataset further accentuates the model's capability to excel under ideal conditions. Fig. 9 shows the results in line graphs

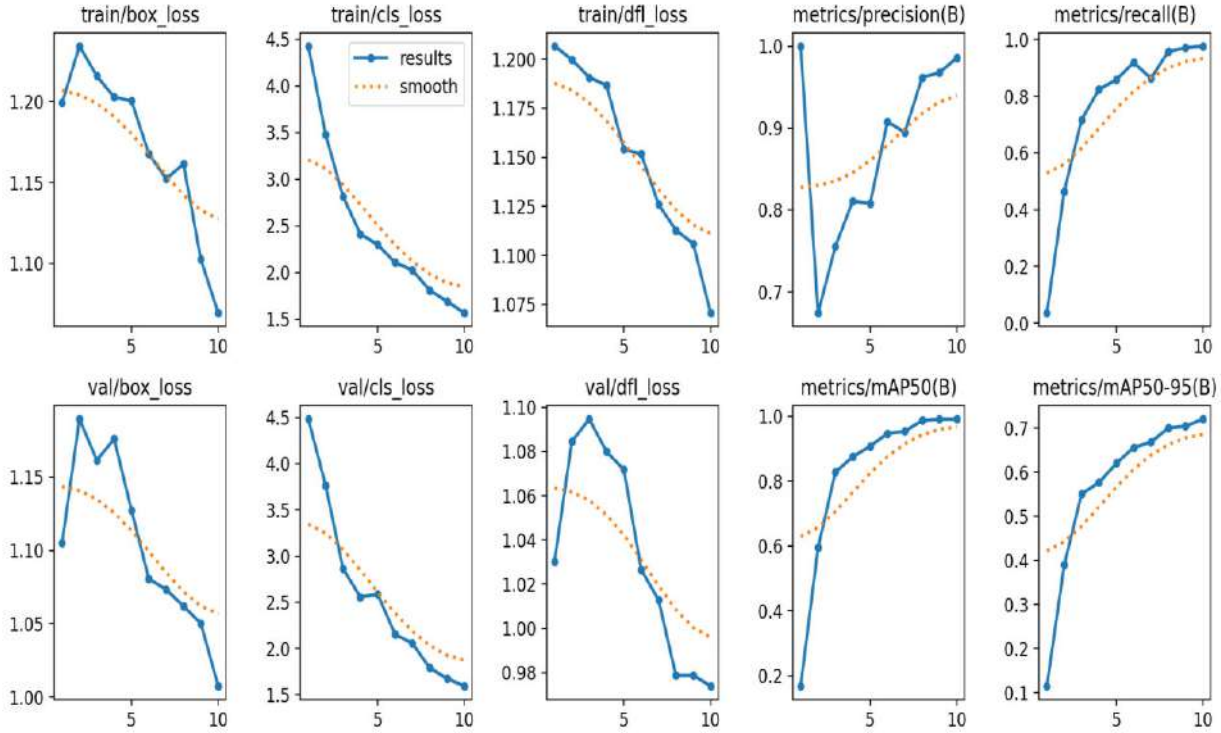


**Fig. 13** Line graph representation of Precision, Loss. Recall and mAP results of RVUMR-14 Pre-processed Naval Mine Dataset with an underwater background

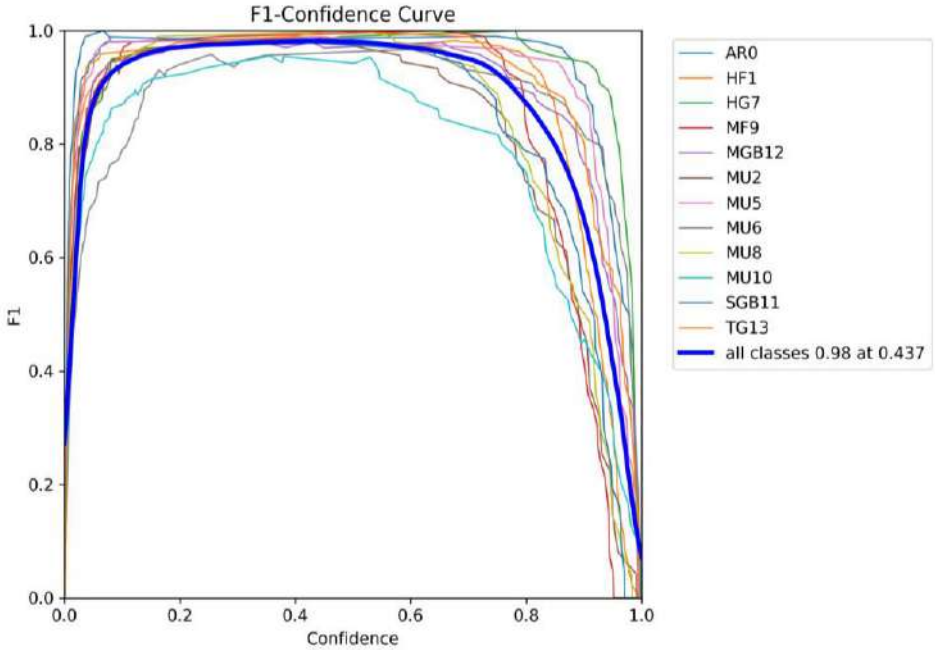


**Fig. 14** F1-Confidence curve of RVUMR-14 Pre-processed Naval Mine Dataset with an underwater background

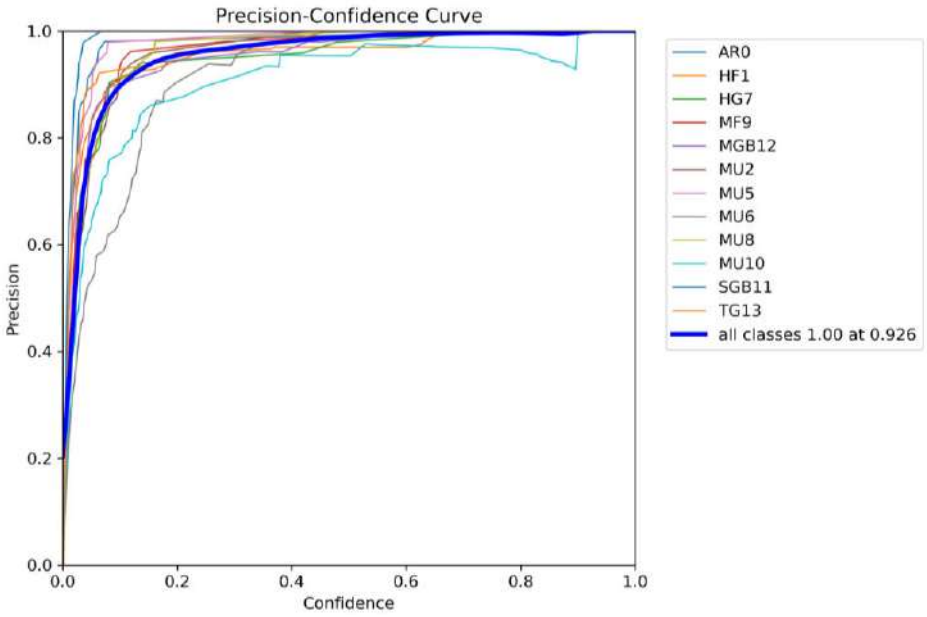


**Fig. 15** Precision - Confidence curve of RVUMR-14 Pre-processed Naval Mine Dataset with an underwater background

The F1-Confidence curve and Precision-Confidence curve illustrated in Fig. 14 and 15, respectively, provide an all-encompassing perspective of the dataset, showcasing its performance without noise. Notably, the F1 score, a key metric for balancing precision and recall, achieves an impressive 98%. This indicates a high level of accuracy and effectiveness in the model's ability to correctly recognize genuine instances while minimizing incorrect identifications.

Furthermore, the Precision-Confidence curve highlights a precision score of 92.6%, underscoring the model's proficiency in accurately classifying positive predictions. This precision metric is particularly valuable in scenarios where minimizing false positives is crucial, emphasizing the reliability of the model in making positive predictions with a high level of confidence

### 7.4 Comparison of the Results with:

**Table - 4** Comparison of results with three different combinations of underwater Naval Mine dataset

| Dataset | Data Preprocessing | Classification Model | Metrics | |
|---|---|---|---|---|
| RVUMR-14 Naval Mine images without an underwater background | Without CLAHE | YOLO v8 | Precision | 97.10% |
| | | | Accuracy | 95.70% |
| | | | Recall | 90.60% |
| | | | F1 | 93% |
| RVUMR-14 original Nava Mine images with an underwater background | Without CLAHE | YOLO v8 | Precision | 96.30% |
| | | | Accuracy | 97.70% |
| | | | Recall | 97.70% |
| | | | F1 | 98% |
| RVUMR-14 pre-processed Naval Mine images with an underwater background | With CLAHE | YOLO v8 | Precision | 92.60% |
| | | | Accuracy | 98.50% |
| | | | Recall | 97.73 |
| | | | F1 | 98% |

Analyzing Table 4 reveals compelling insights into the performance metrics of the YOLO model across different datasets. The accuracy metric, a pivotal indicator of overall model effectiveness, attains its zenith in the dataset devoid of noise. This observation suggests that the YOLO model performs optimally when confronted with a clean and noise-free dataset. The significance of this discovery is paramount for applications where high accuracy is paramount.

Delving deeper into precision, the dataset devoid of background emerges as the frontrunner, showcasing the highest precision scores. Precision, which gauges the model's ability to correctly identify positive instances among its predictions, reaches its pinnacle when the dataset is stripped of background elements. This emphasizes the significance of dataset composition in achieving optimal precision, especially in scenarios where false positives are to be minimized.

On the other hand, when considering recall—a metric that assesses the model's capacity to capture all positive instances within the dataset—the dataset without noise stands out as the leader. The highest recall values in this particular dataset imply that the YOLO model excels at comprehensively identifying relevant objects when noise is absent. This is particularly crucial in applications where the exhaustive detection of positive instances is of paramount importance.

Combining these findings, it becomes clear that the dataset is devoid of noise and emerges as the most preferable option for achieving the highest accuracy and recall, whereas the dataset with no background proves optimal for precision-centric tasks. These nuanced distinctions in performance across different datasets emphasize the significance of tailoring dataset characteristics to the specific goals and priorities of the YOLO model application. As such, meticulous consideration of dataset composition is essential for harnessing the full potential of the YOLO model in diverse real-world scenarios.

## CONCLUSION

In conclusion, the comprehensive evaluation of the YOLO model across various datasets has provided valuable insights into its performance characteristics. The meticulous analysis of Table IV demonstrates that the choice of dataset significantly influences the model's accuracy, precision, and recall metrics.

The dataset without noise emerges as the optimal choice for maximizing overall accuracy and recall. This implies that in scenarios where a clean and noise-free environment is prioritized, the YOLO model excels in accurately detecting objects and minimizing false negatives. This finding is crucial for applications where completeness in identification is paramount.

Conversely, for precision-centric tasks, particularly those where false positives must be minimized, the dataset with no background proves to be the most favorable. The YOLO model achieves its highest precision when background elements are absent, emphasizing the importance of dataset composition in mitigating false positive predictions.

These nuanced observations underscore the need for a thoughtful and tailored approach to dataset selection when deploying the YOLO model in real-world applications. Varied objectives and priorities of a particular task will dictate, that practitioners should carefully consider whether accuracy, precision, or recall holds greater significance, and choose their dataset accordingly.

The results of this research add to not only to a deeper understanding of the YOLO model's behavior but also offer practical guidance for optimizing its performance in diverse scenarios. As the field of object detection continues to evolve, these insights will be instrumental in refining and adapting YOLO-based applications to meet the unique demands of various domains and industries.

## REFERENCES

[1] Diana, Moină, et al. "Marine Mine Detection Using Deep Learning." *2022 26th International Conference on System Theory, Control and Computing (ICSTCC)*. IEEE, 2022.

[2] Munteanu, Dan, et al. "Sea mine detection framework using YOLO, SSD and EfficientDet deep learning models." *Sensors* 22.23 (2022): 9536.

[3] Lalitha, V. P., and Shanta Rangaswamy. "Automatic object detection in aerial image using bent identity-convolutional neural network and fine tuning algorithm." *Multimedia Tools and Applications* 81.7 (2022): 9713-9740.

[4] Jiao, Licheng, et al. "A survey of deep learning-based object detection." *IEEE access* 7 (2019): 128837-128868.

[5] Wu, Wentong, et al. "Application of local fully Convolutional Neural Network combined with YOLO v5 algorithm in small target detection of remote sensing image." *PloS one* 16.10 (2021): e0259283.

[6] Zhao, Liquan, and Shuaiyang Li. "Object detection algorithm based on improved YOLOv3." *Electronics* 9.3 (2020): 537.

[7] Musa, Purnawarman, Farid Al Rafi, and Missa Lamsani. "A Review: Contrast-Limited Adaptive Histogram Equalization (CLAHE) methods to help the application of face recognition." *2018 third international conference on informatics and computing (ICIC)*. IEEE, 2018.

[8] Hożyń, Stanisław. "A review of underwater mine detection and classification in sonar imagery." *Electronics* 10.23 (2021): 2943.

[9] Liu, Chengji, et al. "Object detection based on YOLO network." *2018 IEEE 4th information technology and mechatronics engineering conference (ITOEC)*. IEEE, 2018.

[10] Zhao, Zhong-Qiu, et al. "Object detection with deep learning: A review." *IEEE transactions on neural networks and learning systems* 30.11 (2019): 3212-3232.

[11] Basha, SH Shabbeer, et al. "Impact of fully connected layers on performance of convolutional neural networks for image classification." *Neurocomputing* 378 (2020): 112-119.

[12] Kolisnik, Brendan, Isaac Hogan, and Farhana Zulkernine. "Condition-CNN: A hierarchical multi-label fashion image classification model." Expert Systems with Applications 182 (2021): 115195.

[13] Tripathi, Milan. "Analysis of convolutional neural network based image classification techniques." Journal of Innovative Image Processing (JIIP) 3.02 (2021): 100-117.

[14] Domingos, Lucas CF, et al. "An investigation of preprocessing filters and deep learning methods for vessel type classification with underwater acoustic data." IEEE Access 10 (2022): 117582-117596.

[15] Chen, Dong, Ben Huang, and Fei Kang. "A Review of Detection Technologies for Underwater Cracks on Concrete Dam Surfaces." Applied Sciences 13.6 (2023): 3564.

[16] Choi, Jongkwon, Youngmin Choo, and Keunhwa Lee. "Acoustic classification of surface and underwater vessels in the ocean using supervised machine learning." Sensors 19.16 (2019): 3492.

[17] Wang, Pin, En Fan, and Peng Wang. "Comparative analysis of image classification algorithms based on traditional machine learning and deep learning." Pattern Recognition Letters 141 (2021): 61-67.

[18] Terven, Juan, Diana-Margarita Córdova-Esparza, and Julio-Alejandro Romero-González. "A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas." Machine Learning and Knowledge Extraction 5.4 (2023): 1680-1716.

[19] Santos, Carlos, et al. "A New Approach for Detecting Fundus Lesions Using Image Processing and Deep Neural Network Architecture Based on YOLO Model." Sensors 22.17 (2022): 6441.