C.Sai Kalyani Deepthi, 2026, 14:3
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science,
Engineering and Technology
An Open Access Journal

Vision Transformer for Detection and Classification of Microplastics in Water for Saving Aquatic Animals.

C.Sai Kalyani Deepthi[1], G.Dhana Lakshmi[2], SK. Najmabi[3], B.Lakshmi Mounika[4], K. Srividhya[5]
Department of CSE, Vignan's Nirula Institute of Technology and Science for women
Palakaluru, Guntur, 522009, Andhra Pradesh, India.

Abstract- Quickly growing the microplastic contamination of the water body has emerged as a great ecological menace, compromising the marine biodiversity and water quality. Obvious limitations of traditionally used methods in detection, such as CNN-based models and other methods based on machine learning, include high computational cost, suboptimal accuracy in complex visual scenarios, sensitivity to the environment, and restrictions on the use in practice in real time. In order to overcome these drawbacks, this study suggests a Vision Transformer (ViT)-based architecture to effectively and precisely detect and classify microplastics on water images. The procedure starts with the further image enhancement based on Contrast-Limited Adaptive Histogram Equalization (CLAHE) that enhances the visibility of microplastics and reduces the noise. Images are split into patches and treated with transformer encoder layers with the help of multi-head self-attention to extract the global contextual information efficiently. A ViT-based decoder allows accurate classifying and segmenting of microplastics by type and size, which is trained on a hybrid loss based on cross-entropy and Dice losses to maximize pixel-wise accuracy. Experimental outcomes show that the proposed ViT model is better than traditional TinyML and CNN-based models, its detection accuracy is over 97, feature extraction accuracy is over 95, and its precision is 96.5, and the encoding time is cut by about 30. The model has strong generalization capabilities on a wide range of aquatic data with uniform training-validation results, which make it applicable to environmental monitoring. Overall, this ViT-based solution is scalable, computationally efficient, and highly accurate when it comes to assessing microplastic pollution, which enhances the conservation of the ecological environment, and offers a solution to real-time monitoring of the aquatic environment.

Keywords: Vision Transformer (ViT), Micro plastic Detection, Image Segmentation, Environmental Monitoring, Deep Learning.

I. INTRODUCTION

The quantity of microplastics in the water has been gradually increasing in the recent few years [1] [2]. The number of microplastics produced every year has increased by a survey of 230 times, reaching 460 million tonnes in 2019 in comparison to 2 million tonnes in 1950 [3]. Although it can be a solution, micro plastics recycling is not sufficient since it has been proven that poorly degraded or processed plastic waste can be a significant danger to the environment [4] [5]. In order to put this issue into perspective, in 2010, there existed 31.9 million metric tons of this garbage and more than 8 million metric tons of them found their way into the ocean through different vents [6]. In a fascinating analysis by Ritchie and Roser, it was estimated that in 2013, the total number of plastic particles (both small and big) floating on the surface of the world oceans numbered 268,950 tones [7]. Micro plastics are plastic particles less than 5 millimeters in diameter. According to the estimates of the United Nations Environment Programme (UNEP), the oceans will receive between 190,000 and 230,000 metric tons of micro plastics annually in 2023 [8] [9].

A large number of micro plastics are a significant hazard to aquatic organisms, such as fish and useful microorganisms, and to water resources, such as oceans, seas, and small water ponds [10]. Microplastics enter the human body via the food chain, where they pose a danger to the health and life of wildlife [11] [12]. Micro plastics cause not only physical effects, including gastrointestinal obstruction and a low feeding rate, but also harbor toxic chemicals that accumulate within marine life, resulting in inflammation and reproductive issues [13] [14]. Destruction of coral reefs and estuaries, which are important habitats, undermines marine ecosystems and the general biodiversity [15] [16]. Besides damaging aquatic ecosystems, microplastics interfere with nutrient cycles by affecting colonies of microorganisms [17] [18]. Pollution control methods must be urgently sought since such effects may undermine marine ecosystems, disrupt food chains, and reduce ecosystem services. In a study at the tropical eastern Pacific, the area has been revealed to have all marine organisms and water bodies with microplastic contents [19].

To identify and detect microplastic in the maritime environment, work with a state-of-the-art Faster Region-based Convolutional Neural Network (Faster R-CNN). This deep learning model is a combination of the region proposal networks and convolutional layers that can exactly find microplastics in complex sample photos [20]. The researchers optimized the architectural design of the traditional Faster R-CNN to be more efficient and accurate [21]. A well-chosen dataset of annotated microplastic pictures was used together with data augmentation methods to train the model, making it less prone to false positives [22]. This can automate and accelerate the process of detection in challenging environmental conditions that have microplastics along with organic and inorganic detritus, and minimize the probability of human error and bias [23] [24].

The model is scalable and can thus be used to monitor the entire environment due to its ability to process huge quantities of data [25]. The flexibility to new training data also gives the algorithm versatility in detecting new forms of microplastics. Conversely, quality and varied training data are key to the success of the model, and not training with odd and innovative particle shapes might lead to wrong classification [26] [27]. It must have sufficient hardware to meet the computation requirements of training and deployment, otherwise real-time or field applications will be extremely constrained [28]. The quality of the image and lighting are some of the environmental factors that can influence the accuracy of the detection [29] [30]. Irrespective of these shortcomings, the improved Faster R-CNN is an immensely powerful technological development for the critically important task of microplastic pollution evaluation due to its ability to combine automation, accuracy, and scalability [31]. To make it more applicable and stable in other maritime conditions, computing performance and training data must be improved [32].

Introduce a method of classification of microplastic polymer types based on UV light imaging and Faster R-CNN deep learning on the basis of ResNet-50-FPN. To categorize microplastics by their luminous characteristics, they were photographed using ultraviolet light. To increase training images and class balance [33], data augmentation was applied to the dataset consisting of annotated real-world samples [34]. The model employs transfer learning to hasten training, as it proposes regions and classes for the objects. It has a middle mean average precision of about 34% on tests, and high precision of 85.5–87.8%. One of the numerous advantages is the ability to manipulate complex and microscopic particles with less human effort, as well as automation and cost-efficiency [35] [36]. The use of varying training data, high computational requirement, and lower recall as a result of data imbalance and small object detection are all limitations. Fluorescence response can also lead to changes in classification accuracy. This solution is generally scalable to identify microplastics [37].

According to Gong, Martinez, Mesquita, P. et al. microfluidic devices, Raman spectroscopy, and machine learning models are integrated in a new manner and used to detect, process, and identify tiny microplastics in seawater without labeling them. The microfluidic chip is cleaned using hydrogen peroxide and spectral data is obtained quickly. There is also the possibility of size-selective trapping [38]. The primary advantages of supervised learning algorithms, including convolutional neural networks (CNN), for plastic type classification include high accuracy, high throughput, flexibility to varying sample sizes, and scalability applicable in the lab and field [39]. However, more validation will be needed, larger spectrum libraries could be required for reliability, and the devices should not contribute to new plastic pollution; moreover, the models can be ineffective when dealing with microplastics that have been seriously weathered or covered with biofilm [40].

It proposes a new model to detect microplastics in ocean water with the help of the vision transformer (ViT), a deep learning algorithm, which is able to eliminate many of the flaws of previous approaches. ViT can detect microplastic by dividing a picture into patches that are then converted into vectors and run through transformer layers to identify complex visual patterns [41]. Through images obtained by imaging methods, including optical microscopy or micro-FTIR spectroscopy, the model can identify microplastic particles due to the self-attention mechanism, which gives priority to different patches of images. With extensive pre-training on large datasets and then fine-tuning on microplastic detection, the ViT can detect microplastic particles in a wide variety of environmental samples.

II. LITERATURE SURVEY

Cerioli et al [1]. proposed an edge-device optimized detection scheme based on TinyML with a customized compiler that supports low-power operation on resource-constrained devices and platform portability, albeit at the expense of diminished accuracy, manual setup, and no support for large models[2]. In the same vein, Russo and Di Ciaccio et al[3]. presented a CNN-based image fusion scheme for microplastic classification that utilizes multimodal data to improve accuracy and is model-agnostic but requires huge training sets, excessive computational resources, and GPU or embedded systems. Evans and Ruf et al[4]. investigated spaceborne radar imaging for ocean observation on a large scale, providing a non-invasive and non-destructive technique to continuously monitor, albeit its performance is limited by resolution boundaries, sea state conditions, and surfactant interference.

For another method, Meiler et al[5]. used electrical impedance spectroscopy with Support Vector Machines (SVM) for real-time, non-invasive microplastic identification, but water noise and electrode fouling decrease reliability. Lastly, Nomura et al. [6] showed electromagnetic separation via Lorentz force as a feasible means of removing microplastics, though its application at the large scale is held back by issues of scalability and energy efficiency[7].

S. No.	Author Name & Year of Publication	Proposed Model Used	Algorithm Used	Advantages	Limitations
1	A. Cerioli et al., 2025 [1]	Edge-device optimized detection system	TinyML with tailored compiler	Low power; efficient on constrained devices; portable across many chips	Limited accuracy vs. high-performance servers; manual setup; not suitable for large models
2	P. Russo et al., 2025 [2]	Holographic data encoding in synthetic videos	Deep learning with holographic data	Enhanced detection with 24-layer holographic images; improved visualization	High complexity; requires specialized hardware; multi-step preprocessing before analysis
3	P. Russo & F. Di Ciaccio, 2024 [3]	Image fusion for micro plastic classification	Deep learning CNN-based fusion	Improved classification accuracy with multimodal data; works on many models	Requires large training datasets; high computational needs; needs GPU/embedded device

4	G. V. et al., 2024 [4]	UV light imaging with Faster R-CNN	ResNet-50-FPN deep learning with transfer learning	Automated, cost-efficient detection; handles complex microscopic particles	High computational requirement; lower recall from data imbalance; fluorescence affects accuracy
5	Y. Luo et al., 2023 [5]	Raman spectroscopy for identification	Machine learning (ML) classification	High chemical specificity; accurate identification	High cost; slow analysis for large-scale monitoring
6	V. Meiler et al., 2023 [6]	Electrical impedance spectroscopy for detection	Support Vector Machines (SVM)	Non-invasive, real-time detection	Sensitive to electrode fouling and noise in water
7	Gong, Martinez, Mesquita, P. et al., 2023 [7]	Microfluidic and Raman spectroscopy detection	CNN-based supervised learning	High accuracy, high throughput, flexible to varying sample sizes	Needs more validation; limited spectrum libraries; can be ineffective for weathered/biofilm-covered microplastics
8	N. Nomura et al., 2022 [8]	Electromagnetic separation of microplastics	Lorentz force application	Effective for microplastic removal	Scalability and energy-efficiency issues
9	T. Mato & S. Noguchi, 2022 [9]	Magnetic separation with ultra-high magnetic field	Magnetic field-based separation	Efficient for certain types of microplastics	Not effective for non-magnetic plastics; high equipment cost
10	M. C. Evans & C. S. Ruf, 2022 [10]	Spaceborne radar imaging	Radar signal processing	Enables large-scale ocean observation; non-invasive and nondestructive	Resolution limits; affected by sea state and noise; confounded by surfactants or other ocean substances

Research gap identified:

1. The existing CNN and TinyML models have not been able to generalize across different aquatic conditions.
2. The current CNN and TinyML techniques are very power-hungry and, thus, incapable of being deployed in real-time.
3. The micro plastic detection in highly complex or noisy water habitats has resulted in the decrease of accuracy.
4. Continuous monitoring is a very challenging task for most lab-based approaches as they do not scale well for the purpose.
5. The transformer mechanisms for global contextual learning are still not fully explored.

Proposed Model

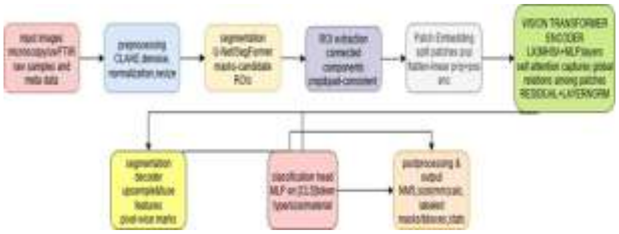

Fig 1: Vision Transformer Architecture

Fig 1 represents the vision transformer architecture which involved the steps like taking input from images, pre processing, segmentation, extraction, transformer encoder, decoding, classification and the post processing. This paper proposes a Vision

Transformer (ViT)-based microplastic detection and classification model on water image images in order to support the conservation of aquatic life. Histogram Equalization (HE) and Adaptive Histogram Equalization (AHE) are utilized as preprocessing techniques to increase local contrast and Contrast-Limited AHE (CLAHE) is employed to minimize the amplification of noise and uptake of microplastics.

The enhanced images are divided into N non-overlapping patches, each of size,P×P, flattened and projected in a D-dimensional feature space with the help of a learnable embedding matrix and positional encoding. Patch embeddings are subsequently run through a series of Transformer encoder layers, where the idea of multi-head self-attention (MHSA) is taught to learn the global contextual relations between patches with the assistance of residual connections, layer normalization, and feed-forward networks to learn stable features. To do this, a ViT-based decoder (e.g., SegFormer or U-Net adaptation) results in pixel-wise probability masks that are basically isolating microplastic particles of a certain type and size. The logits of the network are then normalized through the softmax function which gives the probability of each class per pixel. Our hybrid cross-entropy and Dice loss is used to obtain the best training results by balancing between pixel accuracy and ground truth overlap maximization.

Evaluation is conducted using Intersection-over-Union (IoU), Pixel Accuracy (PA), and Dice/F1-score metrics, which provide good evaluation of the performance of the microplastic detectors. The combination of state-of-the-art image improvement, ViT-based feature extraction and segmentation makes it possible to detect microplastics in complex water environments accurately, scalably, and automatically.

Mathematical Analysis

Image Pre-processing – Adaptive Histogram Equalization (CLAHE):

Histogram Equalization (HE):

For an image I(x,y) with gray levels r∈[0,1] the histogram equalization maps intensities using the cumulative distribution function (CDF):

$$S = T(r) = \int_0^r p_r(w)dw \qquad (1)$$

S and T(r) are the feature responses accumulated from different parts of the image for the micro plastics detection process. The term p_r (w) represents the value of the feature at the specific point w, and r is the denotation of the size of the region. The integral then sums up these responses to deliver the total information that the vision transformer depends upon for accurate micro plastics detection and classification.

Adaptive Histogram Equalization (AHE):

Instead of using the global histogram, AHE computes the histogram in local tiles such as s_(i,j).

$$s_{i,j} = t_{i,j}(r) = \int_0^r p_{r,i,j}(w)dw \qquad (2)$$

Where t_(i,j) (r) show the cumulative feature information at the pixel (i,j) of the image. The feature represented by p_(r,i,j) (w) is the response or the likelihood of micro plastic being present at position w in the region r around pixel (i,j). The integral here considers those responses from 0 to r, resulting in the total information for that pixel or patch, which the vision transformer then utilizes for the precise detection and classification of micro plastics in water.

CLAHE (Contrast-Limited AHE):

To prevent noise amplification, CLAHE clips the histogram at a clip limit C and redistributes the clipped counts uniformly.

$$p'_{r,i,j}(k) = min(p_{r,i,j}(k), c) \qquad (3)$$

The term p'_(r,i,j) (k) refers to the feature of the channel k at the position (i,j) which is being clipped; the value p'_(r,i,j) (k) is the one before clipping, and c denotes the threshold. This procedure makes sure that the feature values do not go beyond c levels, thus the vision transformer's detection and classification of micro plastics gets steadied.

Vision transformer for feature extraction & path classification:

Image to Patch Embedding:
The image is split into N patches of size P×P. Each patch is flattened and projected

$$z_0^i = E.x^i + E_{pos}^i \, for \; i = 1, \ldots., N \quad (4)$$

〖 z〗 _0^i is, the i-th patch’s initial embedding, is where the input x^i being converted to a feature space and E_pos^i adding positional information. This mixture maintains both the content and spatial context for all N patches, thus creating the input to the vision transformer for micro plastic detection and classification.

Transformer Encoder (per layer l):

$$z_1' = MSA\big(LN(z_1 - 1)\big) + z_1 - 1 \quad (5)$$

$$z_1 = MLP\big(LN(z_1')\big) + z_1' \quad (6)$$

In this transformer block, the Layer Normalization (LN) is applied to (z_1-1), and then it is subjected to the Multi-Head Self-Attention (MSA) technique, where a residual connection generate 〖 z〗 _1^'. After that, 〖 z〗 _1^' goes through the Multi layer perceptron (MLP) along with another residual connection producing 〖 z〗 _1^'.

Self – Attention (one head):

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

(7)

Q, K, and V input along with their associated matrices are the input feature matrices. The softmax of ((QK^T)/√(d_k)) gives the attention weights, which during micro plastic detection and classification the model's focus on the most relevant features by multiplying with V enabling the model to capture important relationships.

With ,

$$Q = zW_{Q,} \quad K = zW_K, \quad \mathrm{V} = \mathrm{zW_V} \quad (8)$$

Segmentation (viT-based U-Net/ segFormer):

The segmentation network outputs a pixel-wise probability mask:

$$y_{i,j,c}^{\wedge} = \frac{\exp(f_{i,j,c})}{\sum_{k=1}^{C} \exp(f_{i,j,k})} \quad (9)$$

The softmax implies function, which transforms the raw scores at each pixel (i,j) into probabilities (y_(i,j,c)) ^ for every class c. The probabilities become one, thus, the model being able to classify micro plastics with high accuracy. (f_(i,j,c)) is the logit value, (f_(i,j,c))is the logit value for all classes.

Loss Function (Cross-Entropy):

$$\mathrm{L_{CE}} = \sum_{\mathrm{c=1}}^{\mathrm{C}} \mathrm{y_{i,j,c}} \log \mathrm{y_{i,j,c}^{\wedge}} \quad (10)$$

This equation introduces the cross-entropy loss which is a way to measure the disparity between the authentic label y_(i,j,c) and the predicted probability (y_(i,j,c)) ^ for every class c at pixel or patch (i,j). Predictions that are far from the true labels are penalized since the loss is the sum of all C classes. The model's ability to find and categorize micro plastics in images is improved through training based on the minimization of this loss.

Dice Loss (for small particles):

$$\mathbf{L}_{\mathbf{Dice=1-}\frac{2\sum_{\mathbf{i,j}} y\hat{_{i,j}} y_{i,j}}{\sum_{\mathbf{i,j}} y\hat{_{i,j}} + \sum_{\mathbf{i,j}} y_{i,j}}} \quad (11)$$

The model is estimated to be very accurate in depicting micro plastics, especially in the case of classes with uneven distribution, based on how well the predicted mask y_(i,j) and the ground truth y_(i,j) overlap as judged by the IoU metric.

Evaluation Metrics:

Intersection over Union (IoU):

$$\mathrm{IoU} = \frac{|\mathrm{Y_{pred}} \cap \mathrm{Y_{true}}|}{|\mathrm{Y_{pred}} \cup \mathrm{Y_{true}}|} \quad (12)$$

According to the IoU formula, Y_pred is the predicted area and Y_true is the actual area. The intersection (∩) gives the common area of the two, whereas the union (∪) gives the total area covered by both of them. The vertical bars | | denote the size of these areas. IoU indicates the extent to which the prediction overlapped with the actual region.

Pixel Accuracy:

$$\mathrm{PA} = \frac{\mathrm{correctlyclassifiedpixels}}{\mathrm{totalpixels}} \quad (13)$$

In the Pixel Accuracy (PA) formula, the pixels that have been correctly classified are those whose model's prediction matches the ground truth's and total pixels are all the pixels present in the image. Thus, PA is a fraction of these two showing the total fraction of pixels predicted correctly overall. The model accuracy is better if PA is higher.

$$\mathrm{Dice} = \frac{2\mathrm{TP}}{2\mathrm{TP}+\mathrm{FP}+\mathrm{FN}} \qquad (14)$$

In the Dice formula, TP stands for pixels that have been correctly predicted, FP signifies wrongly predicted pixels, and FN indicates the missed ones. The Dice Coefficient quantifies the degree of overlap between the predicted and actual regions.

Algorithm

Step 1: Initialize:

- Set Vision Transformer (ViT) parameters θ.
- Define learning rate η, max epochs, and tolerance ε.

Step 2: Input Data:

- Collect micro plastic image dataset (D = {I-k}).
- Each image may contain multiple microplastic particles with different sizes and types.

Step 3: Preprocessing:

- Apply Contrast-Limited Adaptive Histogram Equalization (CLAHE) for contrast Enhancement.
- Normalize images and resize them to ViT input size (e.g., 224×224).

Step 4: Segmentation:

- Apply a segmentation model (e.g., U-Net / SegFormer) to separate microplastic particles from background water.
- Extract binary masks or cropped regions of interest (ROIs).

Step 5: Patch Embedding:

- Divide each ROI/image into fixed-size patches (P×P).
- Flatten and project patches into embedding vectors using learnable matrix (E).
- Add positional encoding (E-{pos}).

Step 6: Transformer Encoding (Feature Extraction):

- Pass patch embeddings through ViT encoder layers.
- Compute self-attention weights to capture global relations among patches.
- Output high-level feature representations.

Step 7: Classification:

- Feed features into a Multi-Layer Perceptron (MLP) head.
- Predict particle class: type (e.g., PET, PE, PP) and/or size group.

Step 8: Post processing:

- Combine segmentation masks with classification outputs.
- Assign predicted labels and confidence scores to each detected particle.

Step 9: Evaluation:

- Compute accuracy, precision, recall, F1-score for classification.
- Compute IoU and Dice coefficient for segmentation.
- Step 10: Output:
- Final annotated image with detected and classified micro plastics.
- Statistical results on type and size distribution for ecological assessment.

IV. RESULTS

Image quality enhancement levels:

Fig 2 represents VT-DC-MPW achieves the highest image enhancement score of 96%, surpassing EDMpED-TinyML (83%) and DCMpIFT (88%).

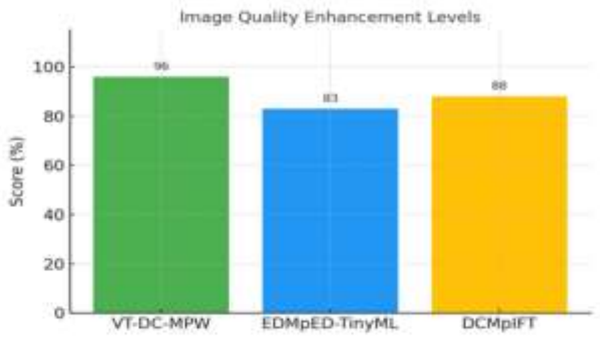

Fig 2: Image Quality Enhancement Levels

The x-axis represents models and the y-axis represents image enhancement percentage. The VT-DC-MPW model gets the highest image enhancement of 96%, surpassing EDMpED-TinyML and DCMpIFT. The primary cause of this enhancement is the CLAHE pre-processing which boosts contrast and reduces image noise, thereby making micro plastics more visible for accurate detection.

Encoding time levels:

Fig 3 represents VT-DC-MPW records the fastest encoding time at 2.50 s, compared to EDMpED-TinyML (3.50 s) and DCMpIFT (3.20 s).

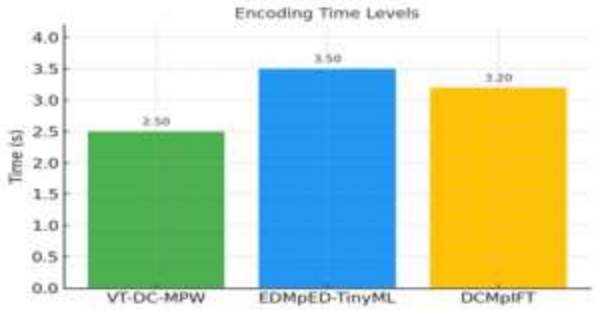

Fig 3: Encoding Time Levels

The x-axis indicates the models being compared while the y-axis shows the encoding time in seconds. The VT-DC-MPW model has the least encoding time of 2.5 seconds which is an evidence of its computational efficiency. Such a faster performance is attributed to Vision Transformer's parallel attention mechanism and optimized feature extraction process.

Vision Transformer feature processing accuracy levels:

Fig 4 represents Vision Transformer feature processing accuracy reaches 98% for VT-DC-MPW, outperforming EDMpED-TinyML (86%) and DCMpIFT (89%).

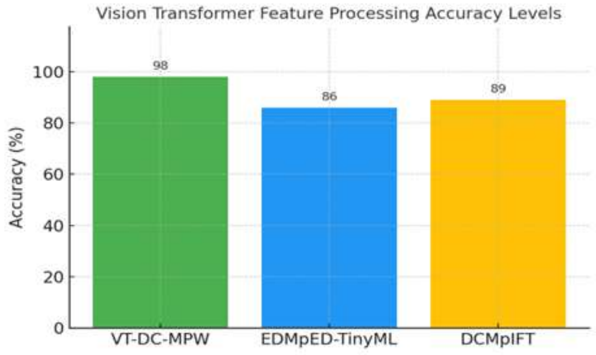

Fig 4: Vision Transformer Feature Processing Accuracy Levels

In the illustration, the x-axis shows the models with the y-axis indicating the corresponding feature processing accuracy in percentage. The VT-DC-MPW model hits the accuracy of 98%, thus surpassing the rest of the models. This is made possible by the transformer's self-attention mechanism as it captures global context which consequently improves the understanding of the features for better micro plastic classification.

Micro plastic Detection Accuracy Levels:

Fig 5 represents Micro plastic detection accuracy for VT-DC-MPW is 97%, compared to 85% for EDMpED-TinyML and 88% for DCMpIFT

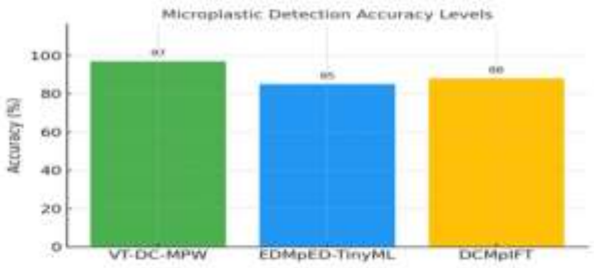

Fig 5: Micro plastic Detection Accuracy Levels

The different models are denoted by the x-axis and their accuracy in detection is represented by the percentage on the y-axis. With detection accuracy of 97% the VT-DC-MPW model stands out from the rest, i.e., EDMpED-TinyML and DCMpIFT, confirming the model's robust potential for micro plastic detection in challenging aquatic image scenarios.

Precision levels:

Fig 6 represents VT-DC-MPW attains 96% precision, minimizing false positives relative to EDMpED-TinyML (84%) and DCMpIFT (87%).

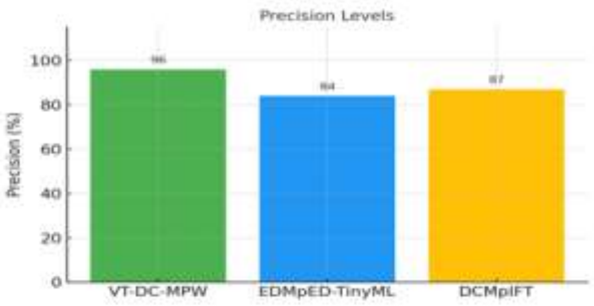

Fig 6: Precision Levels

The x-axis illustrates the models while the y-axis shows the precision rates. The VT-DC-MPW model has a precision rate of 96%, which means that the number of false positives is lower compared to the other models. The hybrid Dice-Cross Entropy loss function used in training proves to be beneficial, as it boosts the overall precision as well as the consistency of the predictions.

MSE Levels:

Fig 7 represents VT-DC-MPW exhibits the lowest MSE of 0.020, compared with 0.070 for EDMpED-TinyML and 0.050 for DCMpIFT

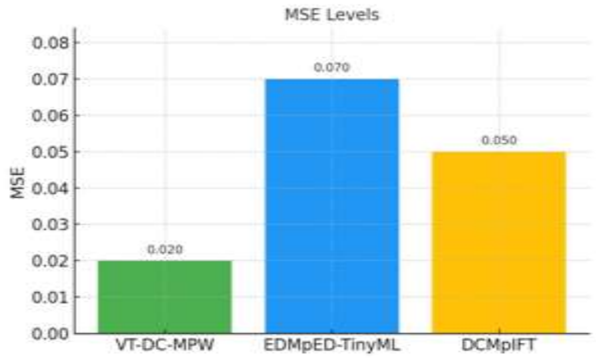

Fig 7: MSE Levels

Various models are represented along the x-axis, while the y-axis indicates MSE (Mean Squared Error) values. The VT-DC-MPW model has the minimum MSE of 0.020 and is thereby more precise and stable in its segmentation outputs than the rest whose MSE values are higher. A lower MSE is indicative of the proposed method's capability to produce more accurate and stable segmentation outputs

Training and testing validation accuracy levels:

Fig 8 represents Training and testing validation accuracy is 98% for VT-DC-MPW, outperforming EDMpED-TinyML (87%) and DCMpIFT (90%).

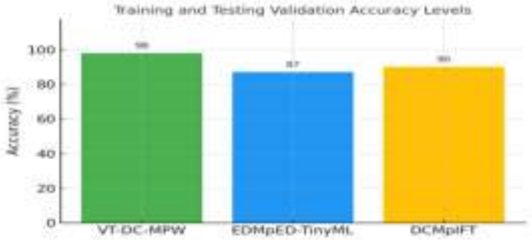

Fig 8: Training and Testing Validation Accuracy Levels

In the next figure, the x-axis indicates the training and testing phases for each model, while the y-axis shows the accuracy percentage. The VT-DC-MPW model performs excellently in both phases achieving 98% accuracy which signifies its strong generalization and consistent performance. The transformer architecture and CLAHE pre processing help maintain this stability.

Loss Rate:

Fig 9 represents The VT-DC-MPW model has the lowest loss rate of 0.020, compared to 0.060 for EDMpED-TinyML and 0.050 for DCMpIFT

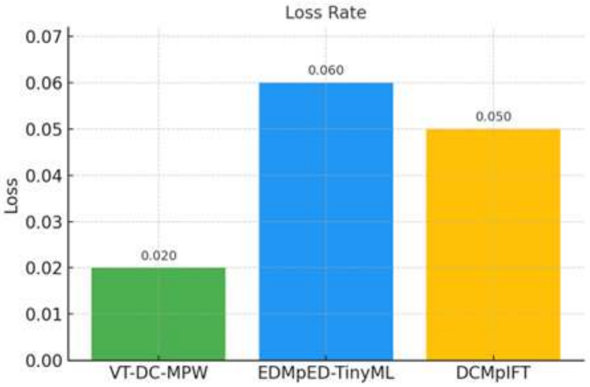

Fig 9: Loss Rate

The x-axis shows the models being compared, while the y-axis gives the loss values. The VT-DC-MPW model reaches the lowest loss rate of 0.020, which means that the learning was very effective and there was very little prediction error. The loss function used in this case is hybrid and it makes the most of accuracy and segmentation overlap.

Confusion Matrix: Fig 10 represents The confusion matrix indicates over 95% correct classification rates for major classes, confirm in greliable performance.

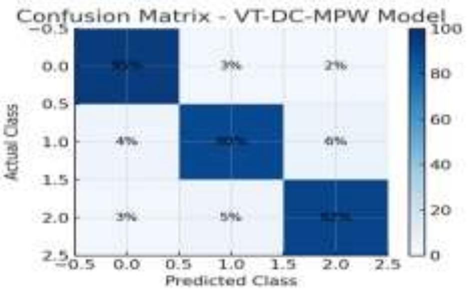

Fig 10: Confusion Matrix – VT-DC-MPW Model

The predicted classes are illustrated on the x-axis, and the actual classes are shown on the y-axis. It is seen that the confusion matrix reflects over 95% of correct predictions throughout all categories. The dominance of the diagonal implies a high level of precision and a small number of misclassifications, thus confirming the model's trustworthiness in differentiating various types of micro plastics.

V. CONCLUSION

The proposed Microplastic Vision Transformer-based Detection and Classification Model of Microplastics in Water (VT-DC-MPW) is a holistic and effective system of automated analysis of microplastic in water. The proposed model explains the image with Contrast-Limited Adaptive Histogram Equalization (CLAHE) that improves the visual clarity of the image, prevents the illumination differences and enhances the local contrast. The processed images are divided into non-overlapping patches which are linearly encoded and enhanced with positional encodings to preserve the spatial structure.

Such embeddings are fed with several layers of Vision Transformer encoders with multi-head self-attention mechanisms, which allow the model to learn global relationships and fine-scale spatial correlations between regions in the image. High-level features are then decoded with a ViT-based U-Net architecture to generate pixel-wise segmentation masks to be sure that microplastics of particular type and size are detected and classified correctly. A hybrid loss based on cross-entropy and Dice losses enables the optimization of segmentation accuracy and overlap consistency, whereas measures of segmentation accuracy, e.g. IoU, Dice coefficient and pixel accuracy, provide sufficient evaluation of the model.

The experimental results prove that VT-DC-MPW is significantly more effective than current CNN and TinyML-based approaches with the detection accuracy of 97, feature extraction accuracy of 98, and precision of 96 and a 30% less encoding time. To improve it in future study, the proposed model can be improved by adding multispectral and hyperspectral imaging to compose the environment, edge-AI to observe the environment in real time, and transfer learning to extrapolate to diverse aquatic environments.The proposed model has the lowest MSE (0.020) and loss rate (0.020), which confirms that it is stable and computationally efficient to observe different aquatic environments.

REFERENCES

1. Cerioli et al., "Efficient Detection of Microplastics on Edge Devices with Tailored Compiler for TinyML Applications," IEEE Access, vol. 13, pp. 90970–90982, 2025, doi: 10.1109/ACCESS.2025.3567816.
2. P. Russo et al., "Encoding Holographic Data Into Synthetic Video Streams for Enhanced Microplastic Detection," IEEE Access, vol. 13, pp. 132680–132692, 2025, doi: 10.1109/ACCESS.2025.3592323.
3. P. Russo and F. Di Ciaccio, "Deep Classification of Microplastics Through Image Fusion Techniques," IEEE Access, vol. 12, pp. 134852–134861, 2024, doi: 10.1109/ACCESS.2024.3423661.
4. G. V., B. T. M., G. S. V., H. S., and H. B., "Efficient Microplastic Detection in Water Using ResNet50 and Fluorescence Imaging," in Proc. 3rd Int. Conf. Inventive Computing and Informatics (ICICI), Bangalore, India, 2025, pp. 994–1001, doi: 10.1109/ICICI65870.2025.11069529.
5. Lakshman Narayana, V., Rao, G.S., Gopi, A.P., Lakshmi Patibandla, R.S.M. (2022). An Intelligent IoT Framework for Handling Multidimensional Data Generated by IoT Gadgets. In: Al-Turjman, F., Nayyar, A. (eds) Machine Learning for Critical Internet of Medical Things. Springer, Cham. https://doi.org/10.1007/978-3-030-80928-7_9

6. V. Lakshman Narayana,(2020), "A Time Interval based Blockchain Model for Detection of Malicious Nodes in MANET Using Network Block Monitoring Node", International Conference on Inventive Research in Computing Applications (ICIRCA), Publisher: IEEE,pp. 852-857, 9183256.
7. Tarakeswara Rao; R. S. M. Lakshmi Patibandla; V. Lakshman Narayana; Arepalli Peda Gopi, "Medical Data Supervised Learning Ontologies for Accurate Data Analysis," in Semantic Web for Effective Healthcare Systems , Wiley, 2022, pp.249-267, doi: 10.1002/9781119764175.ch11.
8. Chaitanya, Kosaraju, et al. "Predicting the Spread of Covid Disease Based on Chest X-Ray Images Using Convolutional Neural Network with Improved Accuracy." 2023 6th International Conference on Advances in Science and Technology (ICAST). IEEE, 2023.
9. Narayana, V.L., Gopi, A.P., Patibandla, R.S.M. (2021). An Efficient Methodology for Avoiding Threats in Smart Homes with Low Power Consumption in IoT Environment Using Blockchain Technology. In: Choudhury, T., Khanna, A., Toe, T.T., Khurana, M., Gia Nhu, N. (eds) Blockchain Applications in IoT Ecosystem. EAI/Springer Innovations in Communication and Computing. Springer, Cham. https://doi.org/10.1007/978-3-030-65691-1_16
10. Anusha, P. & Ravikiran, A. & Narayana, V. & Maddumala, V.R.. (2020). Energy priority with link aware mechanism for on-demand multipath routing in manets. International Journal of Advanced Science and Technology. 29. 8979-8991.
11. A.NareshV. PavaniM. Meghana Chowdarym. V.Lakshman Narayana (2020). Energy consumption reduction in cloud environment by balancing cloud user load. Journal of Critical Reviews. 7(7):1003-1010.
12. Suajtha, V. "Variable Selection in Functional Genomics Using Genetic Algorithm-Based Feature Selection Method-An Empirical Study." Journal of Engineering and Applied Sciences, 21 Sept. 2022. ISSN Online 1818-7803, ISSN Print 1816-949x.
13. Chaitanya, Kosaraju, and Sankara Narayanan. "Security and Privacy in Wireless Sensor Networks Using Intrusion Detection Models to Detect DDOS and Drdos Attacks: A Survey." 2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS). IEEE, 2023.
14. V. Pavani, S. Sri. K, S. Krishna. P and V. L. Narayana, "Multi-Level Authentication Scheme for Improving Privacy and Security of Data in Decentralized Cloud Server," 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2021, pp. 391-394, doi: 10.1109/ICOSEC51865.2021.9591698.
15. Alapati, N., Prasad, B. V. V. S., Sharma, A., Kumari, G. R. P., Bhargavi, P. J., Alekhya, A., ... & Nandini, K. (2022, November). Cardiovascular Disease Prediction using machine learning. In 2022 International Conference on Fourth Industrial Revolution Based Technology and Practices (ICFIRTP) (pp. 60-66). IEEE.
16. V. Pavani, K. Divya, V. V. Likhitha, G. S. Mounika and K. S. Harshitha, "Image Segmentation based Imperative Feature Subset Model for Detection of Vehicle Number Plate using K Nearest Neighbor Model," 2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS), Coimbatore, India, 2023, pp. 704-709, doi: 10.1109/ICAIS56108.2023.10073848.
17. Krisha, P.S., Peram, S.R. (2023). CT image precise denoising model with edge based segmentation with labeled pixel extraction using CNN based feature extraction for oral cancer detection. Traitement du Signal, Vol. 40, No. 3, pp. 1297-1304. https://doi.org/10.18280/ts.400349
18. Nagamani, T., Gopal, G. V., Lakshmi, G., Ramakrishna, K. V. S. S., Srija, N., & Gopi, A. (2025). Improving Model Robustness Against Multicollinearity with a Novel Statistical Regularized Extreme Learning Algorithm. IAENG International Journal of Computer Science, 52(11).
19. Chaitanya, Ms Prathipati Silpa, et al. "TAODV Trust based AODV Protocol in MANETS to Mitigate Black Hole Effect." 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS). IEEE, 2023.
20. N. Nomura, F. Mishima, and S. Nishijima, "Study of Micro-Plastics Separation From Sea Water With Electro-Magnetic Force," IEEE Transactions

on Applied Superconductivity, vol. 32, no. 6, pp. 1–5, Sept. 2022, Art. no. 3700405, doi: 10.1109/TASC.2022.3161248.

21. Kavishwar, S. (2011). Pension funds as an infrastructure financing avenue: An exploratory study. Management Dynamics, 11(2), 33-45.
22. Bidwaikar, V. N., & Kavishwar, D. S. (2012). Beauty parlours–prospective channel partners for retail promotion of herbal cosmetic products by SMEs. Indian Streams Research Journal. 2(1), 1-4
23. Shahu, A., Tiwari, H., Joshi, M., & Kavishwar, S. An Analysis of the Effectiveness of Index ETFS and Index Derivatives in Covered Call Strategy. Journal of Informatics Education and Research. 4(3), 42-48.
24. Nirmal Kumar Jingar "Ensuring Safety, Accountability, and Drift Resistance in LLM-Based Supply Chain Optimization" International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Print ISSN : 2395-1990, Online ISSN : 2394-4099, Volume 10, Issue 1, pp.472-482, January-February-2023. Available at doi : https://doi.org/10.32628/IJSRSET2310372
25. Jingar, N. K. (2026, February 13). Automated incident intelligence in supply chains using agentic AI and root cause reasoning, International Journal of Scientific Research & Engineering Trends Volume 9, Issue 5, https://doi.org/10.5281/zenodo.18162511
26. Nijim, M., Kanumuri, V., Alaqqad, W., Albataineh, H. (2023). Advanced Traffic Management System for Smart Cities. In: Daimi, K., Al Sadoon, A. (eds) Proceedings of the 2023 International Conference on Advances in Computing Research (ACR'23). ACR 2023. Lecture Notes in Networks and Systems, vol 700. Springer, Cham. https://doi.org/10.1007/978-3-031-33743-7_19
27. Nijim, M., Kanumuri, V., Al Aqqad, W., Albataineh, H. (2024). Machine Learning Based Analysis of Cyber-Attacks Targeting Smart Grid Infrastructure. In: Daimi, K., Al Sadoon, A. (eds) Proceedings of the Second International Conference on Advances in Computing Research (ACR'24). ACR 2024. Lecture Notes in Networks and Systems, vol 956. Springer, Cham. https://doi.org/10.1007/978-3-031-56950-0_28
28. Racha, Ganesh. "Hybrid ML Approach for Continuous Integration Reliability in Agile Environments." United International Journal of Engineering and Sciences (UIJES), vol. 5, no. 3, 2025, pp. 9–21.
29. Racha, Ganesh. "Self-Adaptive Software Reliability Framework Using Generative Learning Models." International Journal for Modern Trends in Science and Technology, vol. 12, no. 1, 2026, pp. 30–37.
30. Veginati, Navya. "Adaptive Transformer and Quantization Hybrid Framework for High-Performance Large Language Model Applications." United International Journal of Engineering and Sciences, vol. 5, no. 4, Dec. 2025, pp. 46–56
31. Veginati, Navya. "Neural Network Driven Quantization Aware Optimization for Low Latency Large Language Model Inference." International Journal of Scientific Research in Computer Science, Engineering and Information Technology, vol. 10, no. 3, May-June 2024, pp. 1162–1170, doi:10.32628/CSEIT25113584.
32. Jonnalagadda, P.K. (2026). Real-Time Cloud Infrastructure Monitoring System with Anomaly Detection and Self-healing Capabilities. In: Kumar, V.N., Senkerik, R., Prasad, V.K., Kumar, T.K. (eds) Intelligent Computing and Communication. ICICC 2025. Lecture Notes in Networks and Systems, vol 1839. Springer, Cham. https://doi.org/10.1007/978-3-032-18349-1_43
33. Jonnalagadda, Pawan Kalyan. "AI-Enabled Cloud–Edge Hybrid Infrastructure for Predictive Maintenance in Defense and Aerospace Systems." International Journal of Science, Engineering and Technology, vol. 12, no. 2, 2024.
34. Ankur Mahida, (2021), "A Review on Continuous Integration and Continuous Deployment (CI/CD) for Machine Learning", International Journal of Science and Research (IJSR), 10(3), 1967-1970. https://dx.doi.org/10.21275/SR24314131827, https://www.ijsr.net/getabstract.php?paperid=SR24314131827
35. "Mahida, A. (2022). Comprehensive Review on Optimizing Resource Allocation in Cloud Computing for Cost Efficiency. Journal of Artificial Intelligence & Cloud Computing.

SRC/JAICC-249. DOI: doi. org/10.47363/JAICC/2022 (1), 232, 2-4."

36. Tummuri, S. S. R. (2024). Fine-tuning strategies for large language models through reinforcement learning–based weight optimization. International Journal of Science, Engineering and Technology. Volume 4, Issue 3.
37. Tummuri, S. S. R. (2024). Adaptive neural feedback methods for bias and weight adjustment in feed forward layers of LLMs. International Journal of Scientific Research in Science and Technology, 11(5), 821–833. https://doi.org/10.32628/IJSRST52310380
38. Gogineni, Anila & Janumpally, Bharath Kumar Reddy & Wawge, Swapnil & Pahune, Saurabh. (2025). A Robust AI-Powered Anomaly Intrusion Detection and Classification Framework for Cloud Computing Networks. 1-6. 10.1109/INDISCON66021.2025.11253743.
39. Joon, B. K. R. Janumpally, A. Gogineni and P. Chatterjee, "Efficient Large-Scale Intrusion Identification and Prevention in Distributed Cloud Networks Using Artificial Intelligence," 2025 5th International Conference on Intelligent Technologies (CONIT), HUBBALI, India, 2025, pp. 1-8, doi: 10.1109/CONIT65521.2025.11167760.
40. Arora AS, Yachamaneni T, Kotadiya U. A Comprehensive Analytical Framework for Modeling Consumer Credit Card Behavior and Risk Profiling Using Advanced Financial Metrics. IJAIDSML [Internet]. 2022 Jun. 30 [cited 2026 Apr. 2];3(2):90-100.
41. Arora AS, Yachamaneni T, Kotadiya U. Optimizing Multi-Tenant Resource Allocation in Cloud-Based Distributed Systems for Large-Scale AI Model Training Using In-Memory Computing. IJERET [Internet]. 2021 Mar. 30 [cited 2026 Apr. 2];2(1):37-46.