Microplastic Classification in Holographic Images Using Swin Transformer V2

Sireesha Moturi[1], Sampath Kumar Mabanu[2], Polisetty Rishitha[3], Mohana Sri Krupa Nagasarapu[4], Lakshmi Rayala[5]
[1]sireeshamoturi@gmail.com, [2]sampath.kumar8106@gmail.com, [3]rishithapolisetty7@gmail.com,
[4]mohanasrikrupa@gmail.com, [5]lakshmirayala62@gmail.com

***Abstract*—Classifying microplastics with deep learning methods on integrated holographic images is more successful.By representing the microcosm along with its hologram phase and amplitude as an HSL color space, microplastic samples become rich in both spatial and color information. The Swin Transformer V2 which is famously known for its hierarchical utilization of self-attention and its success on minute and complex images, is chosen to perform exact classification. Results show the accuracy and F1-Score of 91.65% and 91.79%, respectively, thus reflecting a high level of true positive identification and good false positive avoidance. This study highlights the potential of vision transformers for environmental monitoring tasks and provides a framework for the automated microplastics detection via digital holography.**

I. Introduction

The major challenge of illegally dumping plastic waste has now emerged as one of the most significant environmental concerns this day and age. The convenience and overuse of plastic products have resulted in tremendous amounts of waste. Gradually, this waste will undergo a decomposition process and be turned to microplastics. These microplastics have spread to bodies of water and even infiltrated living organisms.

Although these particles are too small to see, they can have devastating impacts on the health of the ocean, and human health. Manual approaches to identifying these particles are inefficient and too slow, too costly, and too reliant on human effort to be practical for large datasets. With microplastic contamination becoming a more prominent concern, there is a greater urgency to find reliable and robust methodologies to deal with the problem.

The advancement in imaging and the analysis of data has prompted researchers to find better methods to keep track of environmental pollutants. The aforementioned developments can potentially lead to enhanced and more reliable methods of microplastic detection.

A specific category of Deep Learning models, Convolutional Neural Networks (CNNs) [1], are adapted to interpret visual data because of the similarity between their design and the processes of the human brain. Unlike traditional neural networks, CNNs [2] have been specifically structured to efficiently and self-sufficiently learn the spatial hierarchies of features through successive layers of convolutions, activations, and pooling. CNNs [3] are particularly well suited to image-based tasks, including image and object classification, detection and analysis, as well as interpretation of medical images. The fundamental principle CNNs [4] , [5]are constructed upon is their capacity to capture local patterns with the use of small filters, or kernels, that slide through the input image. These filters are capable of recognizing the simpler features, for example, edges and textures in the foremost layers while more advanced patterns like shapes and objects in deeper layers. These features are combined as data progresses through the network, resulting in a more detailed comprehension of the image understanding.

979-8-3315-9524-1/25/$31.00 ©2025 IEEE

This research paper is primary model for classifying the holographic images of microplastics was the Swin Transformer V2. The phase and amplitude imaging methods synergize to improve the holography of microplastics' images, which enhances model training. With the adjusted model configuration and training, all its performance metrics, including accuracy, precision, and recall, were significantly improved. Furthermore, the model's architecture is favorable for parallel processing which is important for large datasets, making it suitable for rapid and robust detection of microplastics in environmental monitoring.

II. Related Work

Russo et al. [6] developed a technique applying deep learning for microplastic classification. They synthesized holographic images of amplitude and phase, representing them as HSL. Their method performed well, exhibiting high recall of 96. 4% and precision of 93. 2% with models such as ViT, ResNet18, and DenseNet121. Their approach did not incorporate segmentation.

Lee et al. [7] developed a microplastic imagery dataset meant for detection and segmentation tasks.With segmentation and detection, the team achieved 89.2% F1 score and 63.14% mIoU using EfficientNetV2, U-Net and MRFM, though struggled with intricate small structures and background clutter.

Dils et al. [8] address the issue of insufficient microplastic image datasets by applying GANs to create additional images. This approach improved the F1 rate from 0.82 to 0.91.

Simonyan et al. [9] created the VGG network, which is characterized by a deep architecture that employs small convolutional filters. Though it struggled with holistic image comprehension, the VGG network became an important model for microplastic detection.

Dosovitskiy et al. [10]implemented self-attention to process images by dividing them into smaller components called patches (tokens). ViTs understand spatial relationships across different components of an image and outperform traditional CNNs.

Liu et al. [11] developed the Swin Transformer.Its attention mechanism is unique in that it is hierarchical in nature; it dynamically shifts focus from one region of an image to another. It is particularly useful in analyzing images with high resolution and is very helpful in identifying microplastics.

Zhang et al. [12] applied a one-dimensional CNN to classify microplastics using Raman spectra, outperforming traditional machine learning techniques.

Zhu et al. [13] applied a simple form of transfer learning with an input compression block, which reduced the channel dimensions and assisted significantly in the classification of microplastics with deep neural networks.

Valentino et al. [14] enhanced the classification accuracy of microplastics over microscopic plankton by using features of fractal geometry to improve digital holographic images.

Han et al. [15] demonstrated a capability to detect, classify, and segment large marine microplastics and assigned categories such as fiber, fragment, pellet, or rod to the classified samples.

III. Proposed Methodology

A. Data Preprocessing

The approach for identifying microplastics employing Swin Transformer V2 begins with acquiring amplitude and phase images from the HMPD dataset.

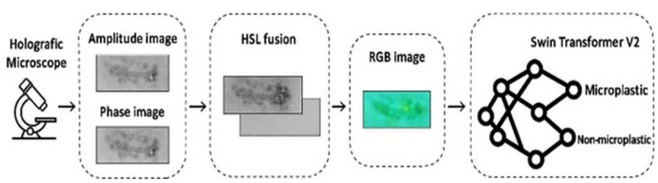

Fig. 1: Block Diagram of the Proposed Swin Transfer V2 Architecture.

As shown in Figure 1, these images are processed into grayscale first. During HSL image fusion, amplitude becomes the hue while phase becomes the saturation and lightness is constant.

$$H = A, \quad S = P, \quad L = 0.5$$

where A is the amplitude image, P is the phase image, and L is set to a fixed value of 0.5 for uniform brightness.

The training goal is defined using cross-entropy loss:

$$L = -\frac{1}{N}\sum_{i=1}^{N} y_i \log(\hat{y}_i)$$

where L is the cross-entropy loss, y_i is the true label, and $\hat{y}_i$ is the predicted probability.

The evaluation metrics used are defined as:

$$\text{Precision} = \frac{TP}{TP + FP}$$
$$\text{Recall} = \frac{TP}{TP + FN}$$
$$F_1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

This enhances texture and color, making the images more akin to RGB images which are more accessible for deep learning models. After fusion, the images are stored and then split into an 80-20 training and validation set. The first step in the process is Capturing with a Holographic Microscope which Figure 1 accompanies with image and color detection.

The images undergo HSL fusion to create a single color image, and then processed in the Swin Transformer V2 for microplastic detection. The figure illustrates all image processing steps along with data arrangement. Images are additionally enhanced by rotating, adjusting, and flipping them to color balance. Swin Transform V2, a vision transformer that uses shifted windows, processes the images to learn small and larger details. During and after multiple rounds of training, the model is evaluated using multiple metrics such as accuracy, precision, recall, F1-score, confusion matrix, and ROC curve. This approach helps the model to be reliable with different types of data instead of overfitting to particular training examples.

B. Dataset Descripion

This study utilizes the Holographic Microplastic Dataset (HMPD), which contains paired amplitude and phase images captured via digital holography, each labeled as microplastic or non-microplastic.

To enhance feature representation, amplitude and phase images were fused using the HSL color space—mapping amplitude to Hue, phase to Saturation, and fixing Lightness at 0.5—producing a single composite image that highlights microplastic texture and structure. These HSL images were then converted to RGB format using OpenCV, resized to 256×256 pixels, and split into 80% training and 20% validation sets for use with Swin Transformer V2.

Figure 2 shows dataset used in this study contains a total of 6,586 images of microplastics and nonmicroplastics, with each class having 3,293 images. As seen in the presented bar graph, both classes are represented equally, which ensures a balanced dataset. Having an unbiased dataset is important when training machine learning models, because it helps avoid any class being favored over others alongside improving the model's generalization when used in various and different tasks.

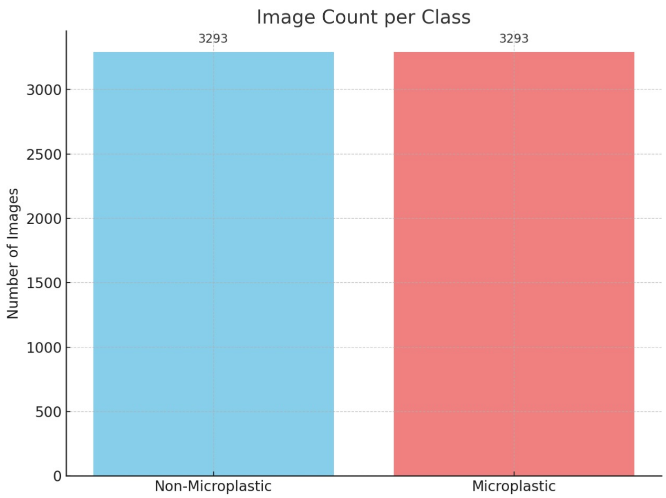

Fig. 2: Block Diagram of the Proposed System

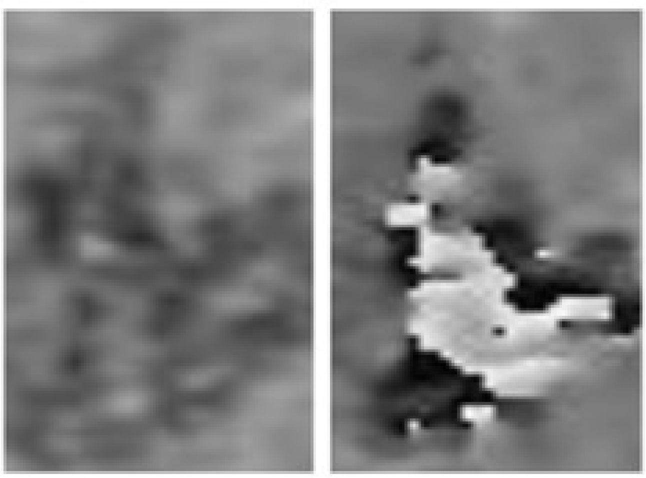

Fig. 3: Before preprocessing

Figure 3, presents the original amplitude and phase data of the microplastic dataset. It captures the image untitled multiclass datset v2. At this stage, the images look visually unappealing, untidy, devoid of clarity, and lack organization, very structured, and devoid of order which obscures any meaningful insights. The images are in the need of, HSL fusion techniques and data augmentation, which are crucial for classification.

Figure 4, illustrates an example of a microplastic sample that underwent HSL fusion and enhancement techniques. HSL fusion improves the images' contrast, making distinguishing the hues easier, and rendering the borders of features more defined, which enhances the visibility of the microplastic's details. Such adjustments optimize learning and enhance classification accuracy during model training, especially for the Swin Transformer V2 model.

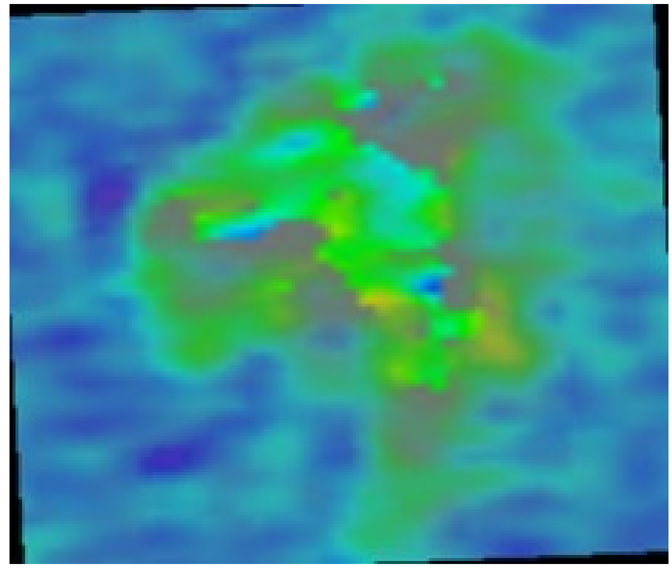

Fig. 4: After preprocessing

TABLE I: Methods of Preprocessing

Step	Method	Purpose
1	Resize (256×256)	Standardizes input dimensions for Swin V2.
2	Horizontal Flip (50%)	Introduces mirrored views to reduce overfitting.
3	Vertical Flip (30%)	Adds diversity via top-down perspective changes.
4	Color Jittering	Simulates lighting variation with contrast, brightness, and saturation.
5	Rotation (±15°)	Improves robustness to image tilts.
6	Tensor Conversion	Prepares images for model input in tensor format.

In the Table 1, illustrate the processes conducted to prepare the data for the model. These processes augment the variability of the images and enable the model to function optimally in different situations. These steps standardize the images and ensure that the modifications included are the ones that aid the model in learning optimally without overfitting to particular instances.

IV. Model Architecture

Swin Transformer V2 utilizes shifted windows to compute local self-attention while preserving hierarchical features. The image is divided into patches, processed through transformer blocks, and finally passed through a linear classification layer. This architecture allows the model to efficiently learn spatial patterns and classify complex microplastic structures.

Figure 5 ,illustrates the entire system used for the classification microplastics work that Image HSL (Hue, Saturation, Lightness) data is used in). Image HSL data begins with the formation of an HSL merged holoAmplitude and holoPhase images captured by a holographic microscope which is known as an image of HSL fusion.

This combination of images helps to raise the contrast and detail-thus the resolving power. It is very important for the verification of the organization and light properties of the particles that are extremely small in size.

The reason for converting pictures into HSL and mixing them with the model is to provide the Swin Transformer V2 model a more vivid and detailed view of the sample images, which allows a proper detection of the features that are very small and not easily discernable.More complex and detailed designs can be recognized as additional normalization layers and feed-forward networks are incorporated. Blocks use a shifted windowing technique wherein the attention windows move between blocks.

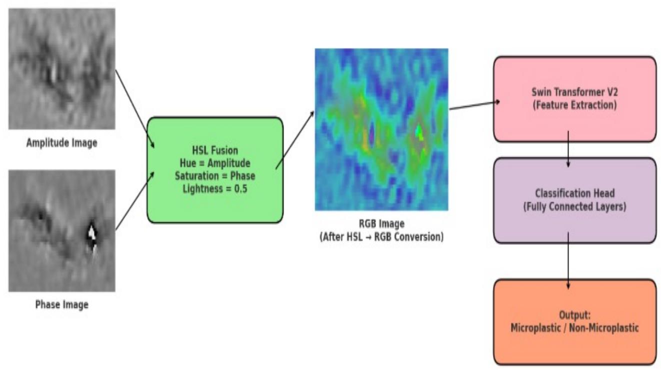

Fig. 5: Architecture of Swin Transformer V2 for Microplastic Classification

The images are fed into the Swin Transformer V2 for feature extraction following RGB conversion and HSL fusion. The first step in the model is patch partitioning, which divides the input image into smaller, more manageable chunks known as patches (tokens).

Every patch is processed separately within the transformer blocks and given metadata pertaining to its location. The Swin Transformer learns the spatial relationships between individual patches while processing them independently. This is crucial for capturing the fine-grained holographic textures of microplastic particles.

Because of the model's hierarchical attention mechanism that enables it to garner both local and global features, it is very successful in analysing complex and extensive holographic images. To sum up, such a design makes it possible for the automated, precise, and efficient identification of the smallest particles, thus easing pollution management and allowing the continuous monitoring of the environment.

The attention mechanism in Swin Transformer V2 helps the model to focus on the most valuable areas of the image, increasing robustness and lowering false positives because holographic images may contain background noise and illumination variations.The complexity of traditional vision transformers is quadratic with the size of the image. Swin Transformer V2 brings this down to a linear one by performing self-attention only on smaller windows.

V. Results and Discussion

The model achieved F1-score and precision of 91.7% and 94.6% respectively is evidence of this. This indicates that vision transformers, especially while employing data fusion techniques, are much better suited for capturing the intricate details of microplastics.The model was trained on a varied dataset of microplastic images from different environmental settings. This approach ensures it remains strong under various conditions

Furthermore, the proposed method is efficient and easy to scale to different environments, which is advantageous for long-term and real-time ecological assessments where reliability and precision are crucial.Training And Accuracy.

TABLE II: Performance Comparison of Microplastic Classification

Model	Ch.	Acc.	Rec.	Prec.	F1
AlexNet	A	82.4	87.5	79.3	83.1
AlexNet	P	83.7	84.2	83.4	83.8
ResNet18	A	80.9	85.9	78.2	81.7
ResNet18	P	83.5	85.0	82.6	83.8
VGG11	A	84.0	86.8	82.1	84.4
VGG11	P	85.5	89.3	82.9	86.7
Swin Transformer V2	HSL (A+P)	**91.65**	**89.3**	**94.6**	**91.7**

The Table 2, indicates the effectiveness of various deep learning models in classifying microplastics is shown in a comparison table that incorporates three channels: Amplitude (A), Phase (P), and Combined (HSL). Standard CNN models such as AlexNet, ResNet18, and VGG11 achieve reasonable results. Notably, VGG11 with the phase channelobic was the highest single channel model with an F1-score of 86. 7%. But, the Swin Transformer V2 model, using fused HSL data from both amplitude and phase images, outperformed every other model by a significant margin.

To further illustrate the performance gap, Table 2 compares the F1-scores of all evaluated models. CNN-based models like AlexNet and ResNet18 offer reasonable accuracy (approximately 82–84%) but lack reliability in recognizing holographic textures. VGG11 manages to get a better recall of 89.3% but limit the precision.

In comparison with previous studies who got a higher recall (96.4%) but a lower precision (93.2%), our method is more balanced. This not only allows a reduction in false alarms but also keeps high detection accuracy. This equilibrium is crucial for the condition of water bodies in real-time, where false positives can provoke a situation of a forced next step, which would be intervention.

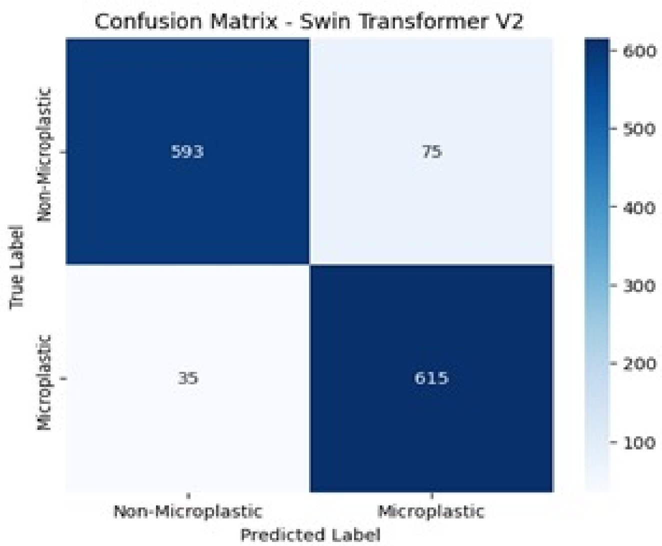

Fig. 6: Confusion Matrix of Swin Transformer V2

The confusion matrix in Figure 6 provides a clear indication of the effectiveness of the Swin Transformer V2 model in distinguishing microplastic particles from non-microplastic particles. It displays four key numbers: true positives (615), true negatives (593), false positives (75), and false negatives (35). These figures suggest the model does very well in confirming 615 microplastics and 593 non-microplastics, demonstrating the model's strong reliability for both microplastics and non-microplastics.

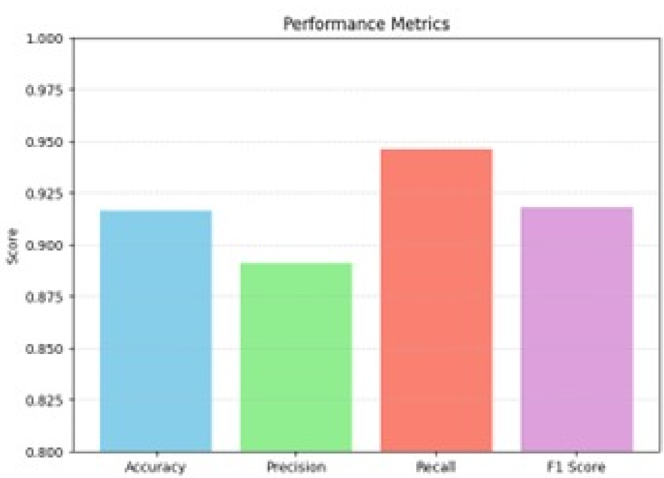

Fig. 7: Performance Metrics of Swin Transformer V2 Model for Microplastic Classification.

Figure 7 illustrates the classification performance of Swin Transformer V2 in distinguishing microplastics from other small particles. The model's performance metrics include Accuracy, Precision, Recall, and the F1 score. Accuracy is just above 91%, suggesting that the model largely captures the correct predictions. Precision, sitting just below 90%, indicates that the model's microplastics identifications, although containing errors, are mostly accurate. Recall is the highest at close to 95%, showing that the model performs well in the identification of microplastics.

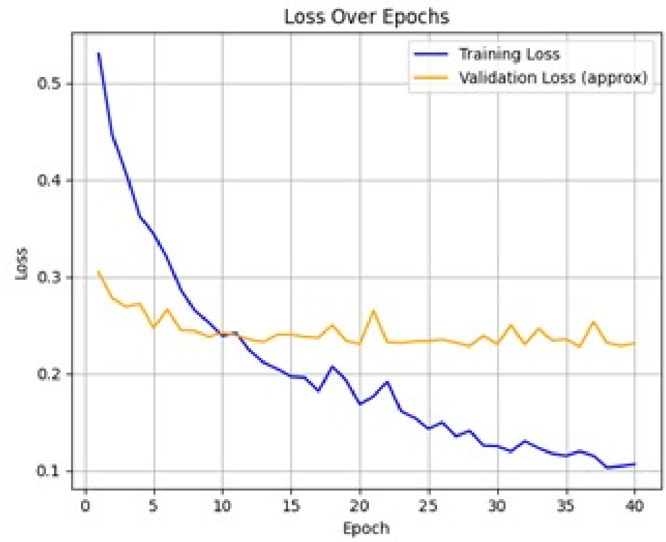

Fig. 8: Training and Validation loss Trend Across 40 Epochs Using Swin Transformer V2

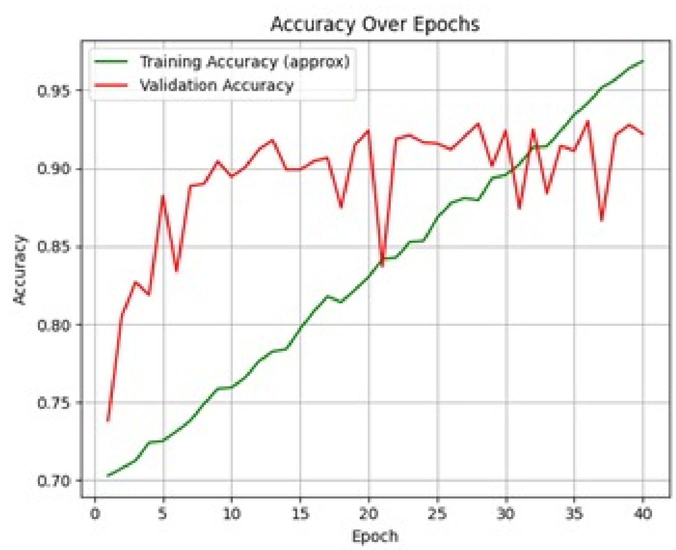

Fig. 9: Training and Validation Accuracy Progression over 40 Epochs Using Swin Trandformer

Figure 8 shows the changes in training and validation loss over 40 epochs for the Swin Transformer V2 model. The training loss, represented with a blue line, continues its downtrend signaling the model is improving with additional epochs. The validation loss, represented with an orange line, remains constant with minute fluctuations around 0.22.

The Figure 9, depicts the changes in training and validation accuracy for the Swin Transformer V2 model over the 40 epochs. The training accuracy, depicted with a green line, increases steadily from a little over 70% to nearly 98%, indicating the model is performing well. The validation accuracy, shown as a red line, remains high for the majority of the time, oscillating between 85% and 93%.

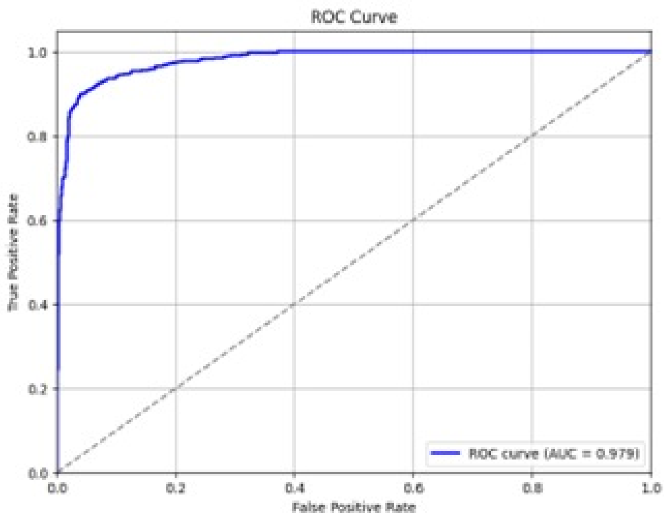

Fig. 10: ROC Curve of Swin Transformer V2 Model with indicating High Classification Performance

The ROC curve in the Figure 10, shows good classification ability with an AUC of 0.979. The model attains a high true positive rate while maintaining a low false positive rate. This implies that there is a strong capacity to differentiate between various types of microplastics.

While the HMPD dataset is well-balanced and curated, it is inherently limited. The dataset consists mainly of laboratory-captured holographic images, which do not necessarily capture the diversity of microplastics in natural water systems. The dataset also only recognizes two classes, microplastic and non-microplastic, whereas in real-world applications, there can be several types, shapes, and stages of degradation of microplastics.

VI. Conclusion

Real-time microplastic classification using deep learning is made easier in this framework with the introduction of Swin Transformer V2 as its backbone. Its strong performance on high-resolution scientific imaging is due to the enhanced model stability, efficiency, spatial reasoning, generalization, and high-level feature integration. The system captures fine and broader details with fusion of HSL and multi-level attention mechanisms.

Using these techniques, the system achieved an accuracy of 91.65% and an F1 score of 91.79%, outperforming VGG11 and CNN-based models that had a maximum accuracy of 85.5%. This is a 6.15% improvement and demonstrates the rigor of transformer models in microplastic detection. This project seeks to further develop a real-time capability for the microplastic classification system based on Swin Transformer V2 for environmental monitoring on multi water bodies.

This model could be integrated into compact devices or specialized systems designed for on-site microplastic detection and continuous monitoring of water quality. In addition, the model could be further improved for real world application by incorporating a greater variety of microplastic types and real-world samples into the dataset.

While this study does not advance algorithms development, Swin Transformer V2 is successfully applied to the classification of holographic microplastics. Amongst earlier models like AlexNet, ResNet18, and VGG11, the Swin Transformer V2 delivers improved accuracy and F1-scores.

This not only reaffirms the superiority of transformer models on complex holographic image data, but also highlights the uniqueness of this study with respect to the performance gains attained in this area.

References

[1] S. L. Jagannadham, K. L. Nadh, and M. Sireesha, "Brain tumour detection using cnn," in *5th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, Palladam, India, 2021, pp. 734–739.

[2] B. Greeshma, M. Sireesha, and S. N. T. Rao, "Detection of arrhythmia using convolutional neural networks," in *Proceedings of Second International Conference on Sustainable Expert Systems*, ser. Lecture Notes in Networks and Systems, S. Shakya, K. L. Du, and H. Wang, Eds. Springer, 2022, vol. 351.

[3] K. V. N. Reddy, Y. Narendra, M. A. N. Reddy, A. Ramu, D. V. Reddy, and S. Moturi, "Automated traffic sign recognition via cnn deep learning," in *IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*, Gwalior, India, 2025, pp. 1–6.

[4] S. Moturi, S. Tata, S. Katragadda, V. P. K. Laghumavarapu, B. Lingala, and D. V. Reddy, "Cnn-driven detection of abnormalities in pcg signals using gammatonegram analysis," in *First International Conference for Women in Computing (InCoWoCo)*, Pune, India, 2024, pp. 1–7.

[5] D. Venkatareddy, K. V. N. Reddy, Y. Sowmya, Y. Madhavi, S. C. Asmi, and S. Moturi, "Explainable fetal ultrasound classification with cnn and mlp models," in *First International Conference on Innovations in Communications, Electrical and Computer Engineering (ICICEC)*, Davangere, India, 2024, pp. 1–7.

[6] P. Russo and C. D. Ciaccio, "Deep learning classification of microplastics in holographic images represented in hsl color space," *Marine Pollution Bulletin*, vol. 179, p. 113684, 2022.

[7] Z. Lee, S. Lee, and H. Kim, "Microplastic segmentation and detection using deep learning with an annotated image dataset," *IEEE Access*, vol. 10, pp. 100 732–100 744, 2022.

[8] C. Dils, M. Metz, and G. Montavon, "Gan-based synthetic data generation to enhance microplastic image classification," *Sensors*, vol. 21, no. 14, pp. 4764–4775, 2021.

[9] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *International Conference on Learning Representations (ICLR)*, 2015.

[10] A. D. et al., "An image is worth 16x16 words: Transformers for image recognition at scale," in *International Conference on Learning Representations (ICLR)*, 2021.

[11] Z. L. et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *IEEE International Conference on Computer Vision (ICCV)*, 2021, pp. 10 012–10 022.

[12] Y. Zhang, B. Zhao, and T. Li, "Microplastic classification using 1d cnn on raman spectroscopy data," *Chemosphere*, vol. 298, p. 134274, 2022.

[13] H. Zhu, X. Wang, and L. Feng, "Microplastic detection via transfer learning with input compression block," *IEEE Access*, vol. 10, pp. 122 341–122 351, 2022.

[14] M. Valentino, G. Amato, and F. Carrara, "Improved microplastic classification in digital holography using fractal features," *Applied Sciences*, vol. 11, no. 23, pp. 11 184–11 198, 2021.

[15] Y. Han, J. Zhang, and S. Liu, "Automated detection and categorization of large marine microplastics using deep neural networks," *Remote Sensing*, vol. 14, no. 7, pp. 1543–1559, 2022.