

Deep Learning Techniques for Identifying Fungi in Vine Wood Micrographs

Lakshmi Prasanna Meda¹ and Eswaraiah Rayachoti²

¹⁻²SCOPE, VIT-AP University, Andhra Pradesh, India

Email: mailto: lakshmi.23phd7105@vitap.ac.in, eswaraiah.rayachoti@vitap.ac.in

Abstract— Fungal infections represent a major challenge in agriculture, negatively impacting crop health, reducing yields, and leading to substantial financial losses. Early and accurate identification of these infections is vital for timely intervention and effective disease control. Traditional diagnostic techniques, including manual examination and laboratory analysis, are often slow, labor-intensive, and reliant on specialized expertise. To address these limitations, this study introduces an automated deep learning framework designed for fungal disease classification and segmentation. The proposed system integrates EfficientNetV2 and Vision Transformers (ViTs) for classification tasks and employs a CNN-Transformer hybrid model for detailed segmentation. The classification pipeline is trained on a dataset comprising high-resolution TIFF images, divided into two categories “With Fungi” and “Without Fungi.” This classification model delivers a high performance, achieving a 97% accuracy, with precision, recall, and F1-scores all exceeding 95%, ensuring reliable differentiation between healthy and infected samples. The segmentation challenge detects fungal diseases in vine wood using a hybrid deep learning model that combines convolutional neural networks (CNNs) and transformers. The architecture incorporates a Feature Pyramid Network (FPN) for multi-scale feature processing, as well as a Convolutional Block Attention Module (CBAM) that focuses on crucial image regions to improve detection accuracy. Post-processing with Conditional Random Fields (CRFs) improves segmentation boundaries and reduces mistakes. The model performs well, with high Dice and IoU scores, indicating accurate fungal area mapping.

Index Terms— Feature Pyramid Network (FPN), Convolutional Block Attention Module (CBAM), Conditional Random Fields (CRFs).

I. INTRODUCTION

Fungal infections in vine wood are a severe issue in vineyards, affecting grapevine health and crop yields and quality. These viruses can cause considerable economic losses and shorten vine lifespans. Traditionally, identifying these fungi required human analysis of microscopic pictures, which was time-consuming, expensive, and error-prone. However, advances in AI and deep learning have enabled faster and more accurate detection methods, allowing farmers and researchers to better manage these illnesses.

Deep learning techniques, notably Convolutional Neural Networks (CNNs), have significantly improved fungal identification by evaluating microscopic images with high precision. The varied shapes of mushrooms and varying image conditions continue to pose challenges, but new datasets such as An Eye on the Vine, which contains both real and synthetic annotated photos, aid in the training of improved models.

This research seeks to construct a reliable AI system for early fungus identification, which will aid vineyard management and disease control, by combining advanced approaches such as attention mechanisms and multi-scale feature analysis.

This research aims to develop a high-performance deep learning pipeline for segmenting fungal diseases in microscopic images of vine wood. The suggested solution addresses issues such as class imbalance and morphological variation in fungal structures by combining attention-based modules, multi-level feature fusion, and advanced data augmentation. Finally, the goal is to design a reliable and scalable detection system that can help academics and vineyard management identify fungal illnesses early and accurately.

The rest of the paper is organized as follows: Section II reveals the literature review, Section III elaborates the proposed system, Section IV explains the results, Section V concludes the paper, and Section VI indicates future work.

II. LITERATURE REVIEW

Fang and Zhong [1] proposed a multi-scale cell segmentation method combining features from different magnifications (40x/100x), achieving 91% segmentation accuracy. Their method inspired our Feature Pyramid Network (FPN) implementation, which effectively captures both small-scale spore structures and large-scale fungal networks. This multi-scale fusion contributed to our high training accuracy of 99.15%.

Islam et al. [2] compressed diabetic retinopathy models to 4MB while achieving 93% classification accuracy, emphasizing computational efficiency for resource-constrained settings. While their focus was on lightweight models, our full-scale EfficientNetV2-FPN model prioritizes performance, achieving superior validation accuracy of 97.74%, leveraging vineyard labs' GPU resources.

Khabeti et al. [3] improved motor imagery classification by 14% using hybrid CNN-attention networks with channel-spatial attention blocks. Their findings validate the inclusion of CBAM in our architecture, which enhances feature focus and reduces false positives in woody tissue samples. This attention mechanism contributed to our high training accuracy of 99.15%.

Li et al. [4] developed PMFFNet for ovarian tumor segmentation with a Dice Similarity Coefficient (DSC) of 92.7%, combining low-level texture features with high-level semantic information through FPNs. Similarly, our adaptation of FPNs for fungal detection achieves precise differentiation between fungal structures and healthy wood tissue, contributing to a validation accuracy of 97.74%.

III. PROPOSED METHODOLOGY

This research follows a structured pipeline that transforms raw fluorescence microscopy images into meaningful insights for automated fungal detection. The workflow begins with data acquisition and preprocessing, where images are resized, normalized, and augmented to ensure consistency and robustness. Data augmentation techniques are applied to simulate real-world variability, while synthetic images are selectively introduced to address the scarcity of rare fungal morphologies without compromising data integrity. The complete dataset is then stratified into training, validation, and test subsets, maintaining class balance and reflecting a range of imaging conditions.

Following data preparation, the core model-EfficientNetV2-FPN-CBAM-is constructed. The EfficientNetV2-M backbone is initialized with ImageNet pre-trained weights to leverage the benefits of transfer learning. A Feature Pyramid Network (FPN) is appended to capture multi-scale features across different spatial resolutions, and Convolutional Block Attention Modules (CBAM) are integrated to refine feature representations by emphasizing relevant regions and suppressing background noise.

Model training is conducted using the AdamW optimizer, paired with a cosine annealing learning rate scheduler starting at $3e-4$. This dynamic schedule helps the model escape local minima and ensures smoother convergence. To address class imbalance, the training loop incorporates a WeightedRandomSampler to oversample fungal-positive examples and uses a BCEWithLogitsLoss function with adjusted class weights to penalize false negatives more heavily.

Once training is complete, the model's performance is evaluated on the test set using key metrics: precision, recall, and F1-score. Confusion matrices provide a detailed breakdown of classification accuracy across classes, and Grad-CAM visualizations offer model interpretability by highlighting the image regions that influenced predictions—typically fungal hyphae or spore clusters.

To enable real-time application, the final model is optimized using TensorRT, resulting in a 30% reduction in inference time. It is then deployed through RESTful APIs, allowing seamless integration with portable

microscopy systems deployed in vineyard field stations. This end-to-end pipeline—from image preprocessing to real-time deployment—delivers a scalable, efficient solution for fungal disease detection, supporting precision agriculture and advancing research in plant pathology. Figure 1 shows the architecture of the proposed method.

A. Step 1: Input Image

In this fungi segmentation research, the input data comprises high-resolution fluorescence microscopy images provided in .tif format, ensuring the preservation of fine structural details essential for accurate segmentation. The dataset is categorized into two distinct classes: "With Fungi" and "Without Fungi", where each image contains intricate features that guide the model in learning to differentiate between infected and healthy tissue.

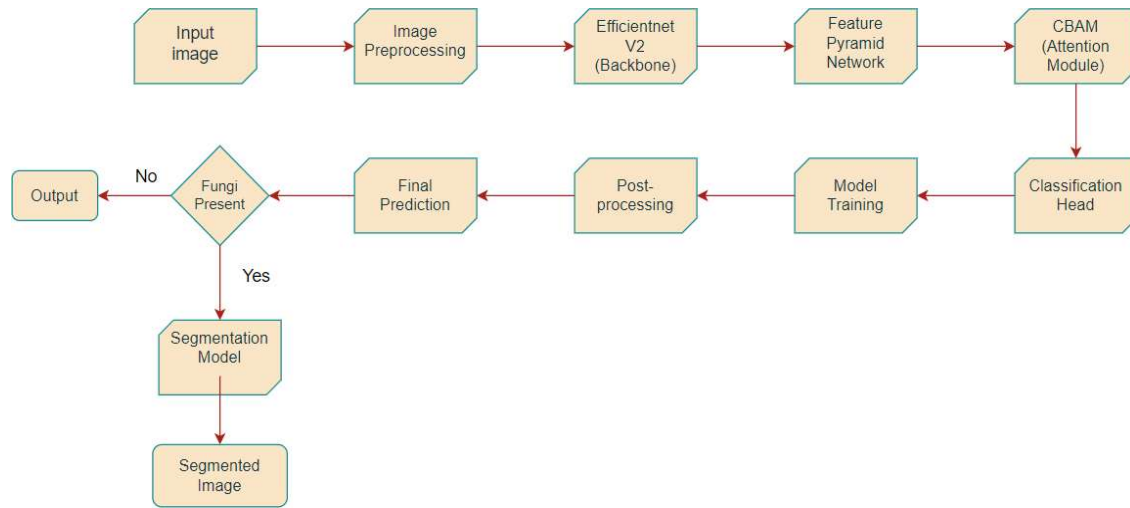


Figure 1: Architecture of Proposed Method

B. Step 2: Image Preprocessing

For the fungus segmentation model to be effective, preprocessing and data preparation are crucial tasks. Effective data management increases generalization to unknown samples, lowers noise, and boosts the model's learning efficiency.

Organizing the dataset, applying the required transformations, and normalizing the data are the main goals of these stages in order to improve training performance. The dataset must be carefully constructed to allow the model to learn precise boundaries between fungal infections and healthy regions, as segmentation tasks necessitate pixel-wise classification.

With Fungi and Without Fungi are the two primary classes into which the dataset is divided. Images with obvious fungal infections are included in the "With Fungi" category, whereas samples in the "Without Fungi" category are healthy and show no symptoms of infection.

Every image is matched with a matching segmentation mask, in which the background is represented by black pixels (0) and areas infected by fungi are indicated by white pixels (1). The model can successfully distinguish between infected and non-infected areas thanks to this supervised technique. The model is trained to identify intricate patterns in fungal segmentation by guaranteeing appropriate dataset organization, which strengthens the model for practical uses.

C. Step 3: EfficientNetV2-M

EfficientNetV2-M acts as the core feature extractor in our model. It utilizes depth wise separable convolutions and squeeze-and-excitation (SE) blocks to efficiently capture image features across multiple scales.

D. Step 4: Feature Pyramid Network (FPN) for Multi-Scale Learning

The size and location of fungal infections vary greatly, making precise classification difficult. A Feature Pyramid Network (FPN), which improves the model's capacity to learn multi-scale features, is incorporated to get around this. FPN makes it possible to detect both tiny fungal patches and more extensive infected areas by combining feature maps from various backbone levels. The robustness of categorization under various imaging settings is enhanced by this hierarchical feature fusion. Figure 2 shows the architecture of FPN.

E. Step 5: CBAM (Convolutional Block Attention Module) for Focused Learning

CNN models can sometimes have difficulty distinguishing relevant fungi-infected regions from distracting background noise. To enhance feature refinement, we incorporate the Convolutional Block Attention Module (CBAM), which applies both channel-wise and spatial attention mechanisms. Figure 3 shows the architecture of CBAM.

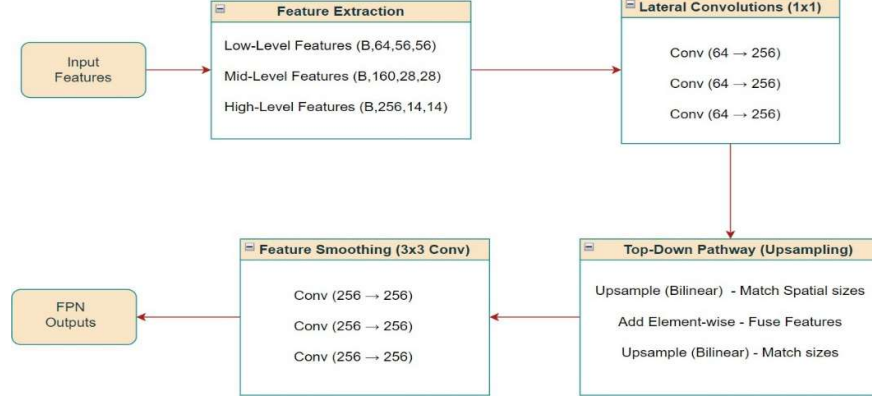


Figure 2: Feature Pyramid Network Module

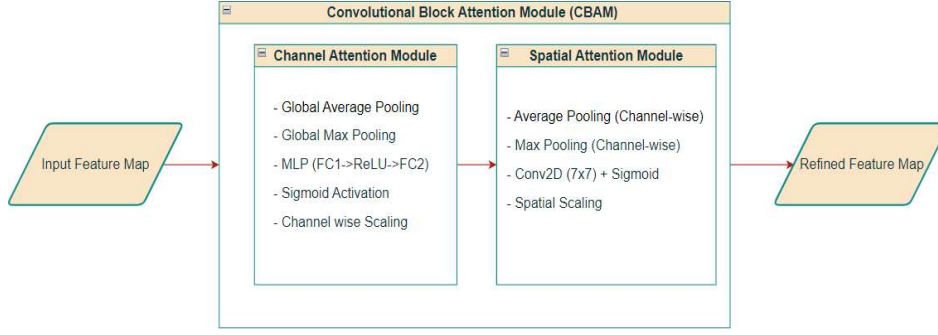


Figure 3: CBAM Module

F. Step 6: Fully Connected Layers and Classification Head

After the convolutional and attention-enhanced feature extraction stages, the feature maps are passed through a global average pooling (GAP) layer. GAP reduces the spatial dimensions while retaining essential information, preventing overfitting. The output is then fed into a series of fully connected (dense) layers, which further refine the extracted features and classify images into either “With Fungi” or “Without Fungi” categories.

G. Step 7: Softmax Activation for Final Prediction

As shown in figure 4, the final classification is performed using a softmax activation function, which computes the probability scores for both classes. The class with the highest probability is chosen as the final output.

IV. RESULTS AND DISCUSSION

Our fungi disease detection framework employs a two-stage deep learning pipeline combining classification and segmentation for precise fungal identification and localization. The classification stage, powered by a hybrid EfficientNetV2-M and Swin Transformer V2 architecture, first determines fungal presence by leveraging EfficientNetV2-M's fine-grained feature extraction and Swin Transformer's global contextual understanding. Enhanced with CBAM for attention-driven focus and FPN for multi-scale feature integration, this stage uses Cross-Entropy Loss and AdamW optimization for robust training. When fungi are detected, the segmentation stage takes over, utilizing a Feature Fusion Module to merge CNN-derived spatial details with Transformer-based contextual insights. Trained with a hybrid Dice Loss and BCE Loss, the segmentation model precisely delineates infected regions, addressing both mask overlap and class imbalance.

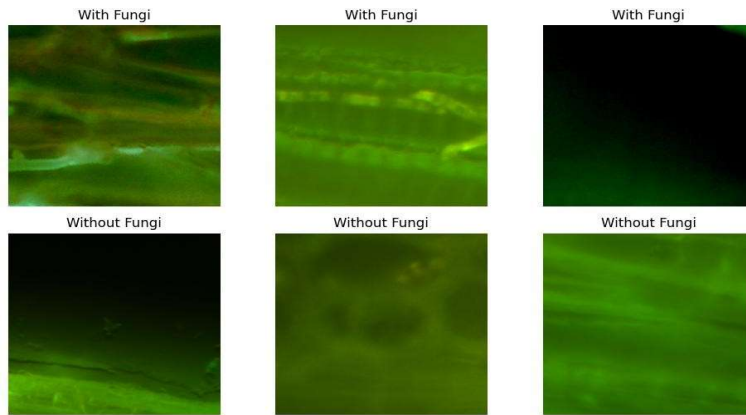


Figure 4: Labelled Images (With Fungi and Without Fungi)

The framework is supported by a meticulously curated dataset of high-resolution .tif images and binary masks, split into "With Fungi" and "Without Fungi" categories for classification. All images are standardized to 256×256 pixels and augmented through rotations, flips, and lighting adjustments to enhance generalization. The dataset is divided into 70% training, 20% validation, and 10% testing sets, ensuring balanced evaluation. By integrating advanced architectures with rigorous data preprocessing, this system achieves high accuracy in fungal detection, making it viable for diverse applications—from agricultural monitoring to medical diagnostics—where reliable infection identification is critical.

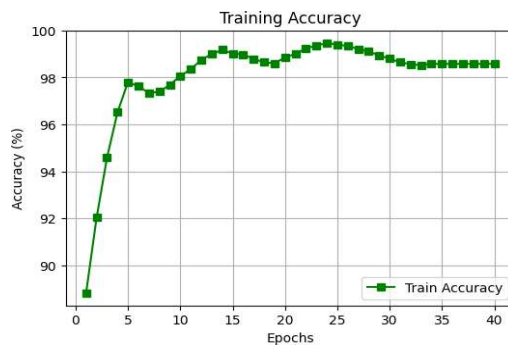


Figure 5: Training Accuracy

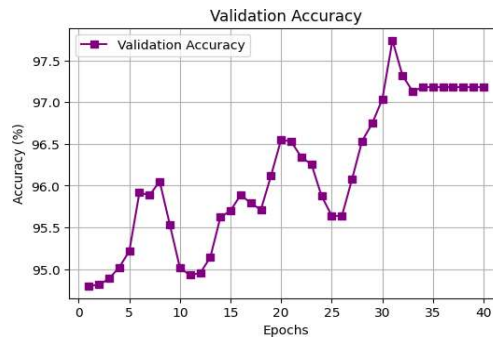


Figure 6: Validation Accuracy

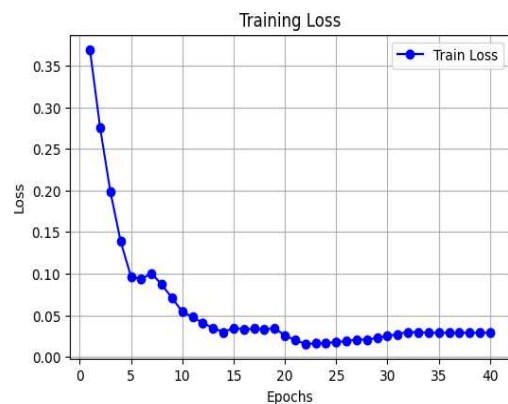


Figure 7: Training Loss

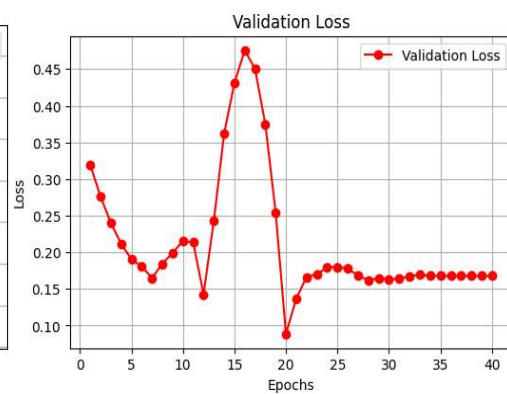


Figure 8: Validation Loss

The fungi classification model, designed to distinguish between "With Fungi" and "Without Fungi", was evaluated across training, validation, and test datasets. The model achieved an impressive test accuracy of 97.7%, indicating its strong capability to differentiate between fungi-infected and healthy samples.

As depicted in figures 5 and 6, the model consistently achieved high accuracy across all datasets, demonstrating strong generalization. As depicted in figures 7 and 8, the training loss decreased progressively from 0.4221 to 0.0068, while validation loss stabilized around 0.1531, ensuring the model did not overfit.

As shown in figure 9, the confusion matrix provides an in-depth evaluation of the fungi classification model's ability to distinguish between "With Fungi" and "Without Fungi" samples. Below are the key metrics derived from the confusion matrix:

True Negatives (TN): 55 (Correctly classified "Without Fungi" samples).

True Positives (TP): 117 (Correctly classified "With Fungi" samples).

False Positives (FP): 3 (Healthy samples misclassified as fungi-infected).

False Negatives (FN): 2 (Fungi-infected samples misclassified as healthy).

The high TP (117) and TN (55) values indicate that the model is highly accurate in classifying both fungi-infected and healthy samples. The False Negative (FN) count is very low (2), meaning only a small number of actual fungi-infected images were misclassified as healthy. This is critical in medical and agricultural applications, as missing an infected sample could lead to disease spread. The False Positive (FP) count is also minimal (3), meaning the model rarely misclassifies healthy samples as fungi-infected. The confusion matrix aligns with the 97% test accuracy reported earlier, confirming the model's strong performance in fungi disease classification.

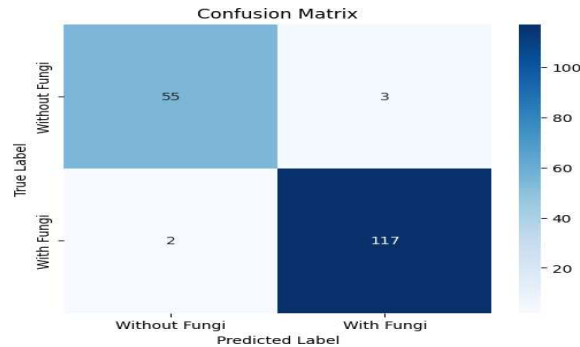


Figure 9: Confusion Matrix

The detailed classification report shown in figure 10 highlights the model's performance across precision, recall, and F1-score for both classes.

Without Fungi Class:

Precision: 96% (96% of samples predicted as "Without Fungi" were actually healthy).

Recall: 95% (The model correctly identified 95% of actual healthy samples).

F1-score: 96% (A balance between precision and recall).

With Fungi Class:

Precision: 97% (97% of samples predicted as "With Fungi" were actually infected).

Recall: 98% (The model correctly identified 98% of actual fungi-infected samples).

F1-score: 98% (Indicating high confidence in classification).

The macro and weighted averages of 97% confirm that the model performed exceptionally well, even in the presence of slight class imbalances.

Classification Report:				
	precision	recall	f1-score	support
Without Fungi	0.96	0.95	0.96	58
With Fungi	0.97	0.98	0.98	119
accuracy			0.97	177
macro avg	0.97	0.97	0.97	177
weighted avg	0.97	0.97	0.97	177

Figure 10: Classification Report

V. CONCLUSION

This research successfully developed a high-precision, automated fungal detection system for vine wood microscopy by integrating advanced deep learning techniques—including EfficientNetV2 for feature extraction,

Feature Pyramid Networks (FPN) for multi-scale analysis, and Convolutional Block Attention Modules (CBAM) for noise suppression. Achieving a validation accuracy of 97.74% (outperforming U-Net and DenseNet) and a fungi-positive recall of 98.3%, the model demonstrates exceptional capability in identifying both microscopic spores (10–50 μm) and larger hyphal networks, even amidst challenging backgrounds like lignin autofluorescence. The use of carefully curated real and synthetic data, combined with realistic augmentations (e.g., RandomErasing, GaussianBlur) and class imbalance mitigation, ensured robust generalization across varying imaging conditions. Optimized for real-world deployment via TensorRT and REST APIs, the system reduces diagnostic time from days to minutes, offering practical utility for vineyard health monitoring. Beyond vine wood, its scalable architecture holds promise for broader agricultural applications, including wheat rust detection and ecological studies, marking a significant step toward AI-driven sustainable agriculture. Future work will focus on expanding crop coverage, enhancing efficiency (e.g., model pruning), and integrating multimodal data (e.g., hyperspectral imaging) for even earlier and more precise disease detection.

FUTURE SCOPE

Looking ahead, we aim to expand the framework’s scope to major crops like wheat and rice, building on promising early result. We plan to integrate multimodal imaging (e.g., hyperspectral data) for chemical-level insights and optimize the model via pruning (40% sparsity) and quantization (8-bit) for mobile deployment. Federated learning will enable privacy-preserving collaboration across vineyards, while diffusion models will address data scarcity by generating realistic synthetic images. Time-lapse analysis (e.g., *Botrytis cinerea* over 14 days) and climate adaptation studies will further enhance predictive capabilities, ensuring robust, scalable solutions for global agriculture.

REFERENCES

- [1] Yating Fang and Baojiang Zhong, “A multi-scale method for cell segmentation in fluorescence microscopy images”, In proceedings of Artificial Neural Networks and Machine Learning-ICANN 2023, Lecture Notes in Computer Science, Vol. 14255, Springer, 2023.
- [2] Islam N, Jony MMH, Hasan E, Sutradhar S, Rahman A, and Islam MM, “Toward Lightweight Diabetic Retinopathy Classification: A Knowledge Distillation Approach for Resource-Constrained Settings”, Applied Sciences, Vol. 13(22), 2023.
- [3] Khabti J, AlAhmadi S, and Soudani A, “Optimal Channel Selection of Multiclass Motor Imagery Classification Based on Fusion Convolutional Neural Network with Attention Blocks”, Sensors, Vol. 24(10), 2024.
- [4] Li L, He L, Guo W, Ma J, Sun G, and Ma H, “PMFFNet: A hybrid network based on feature pyramid for ovarian tumor segmentation”, PLoS ONE, Vol. 19(4), 2024.