SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V12I9P109

Volume 12 Issue 9, 108-117, September 2025

Original Article

Path Aggregation Network based WATT-EffNet for Unmanned Aerial Image Classification

Nakkala Geetha[1], Gurram Sunitha[2]

[1]*School of Computing, Mohan Babu University, Tirupati, Andhra Pradesh, India.*
[2]*Department of CSE, School of Computing, Mohan Babu University, Tirupati, Andhra Pradesh, India.*

[1]*Corresponding Author : geetha8923@gmail.com*

Received: 09 July 2025 Revised: 11 August 2025 Accepted: 10 September 2025 Published: 29 September 2025

Abstract - *High-resolution aerial images captured by UAVs are critical in various real-world applications. However, effective classification of these images is challenging due to scale variation, occlusions, and complex scene structures. Existing deep learning models often face a trade-off between computational efficiency and classification accuracy. To address this issue, a novel Path Aggregation Network-based Wider Attention EfficientNet (PANet-WATT-EffNet) is proposed. PANet-WATT-EffNet employs EfficientNet as a lightweight backbone, combined with wider attention layers to capture salient regions. PANet is used for multi-scale feature fusion. PANet-WATT-EffNet's design improves the extraction of fine-grained and global features, enabling accurate recognition of small and complex objects in aerial imagery. Experimental evaluation on UAV benchmark datasets shows that the model achieves 97.71% accuracy. A significant gain is observed in F-measure, MCC, and reduced RMSE values of PANet-WATT-EffNet, while also lowering computational time compared to existing methods. The results confirm the robustness of the approach in handling diverse aerial imagery. The lightweight architecture further supports deployment on resource-constrained edge devices, making it suitable for applications in precision agriculture, urban infrastructure monitoring, disaster management, and defence surveillance.*

Keywords *- Remote sensing, Path aggregation network, EfficientNet, Multi-scale feature fusion, Remote sensing, Lightweight deep learning.*

1. Introduction

Remote sensing has become an important method in agriculture, range management, and environmental monitoring. It supports applications such as precision farming, biodiversity detection, and vegetation change tracking [1]. Remote sensing helps in better and timely decision-making for crop management, land use planning, ecological conservation, etc. Historically, satellites and manned aircraft collected most remote sensing imagery. However, Unmanned Aerial Vehicles (UAVs) have emerged as an effective alternative for capturing aerial data at higher spatial resolutions. UAVs are more cost-effective than hiring manned aircraft. They also offer greater schedule flexibility. They can capture centimetre-level resolution imagery, which is essential for detecting small-scale changes in landscapes.

UAVs have been increasingly adopted for agricultural monitoring. They are used for assessing crop health, detecting plant stress, estimating yield, etc [2]. UAV imagery has also been integrated with satellite platforms to upscale local measurements for regional-scale analysis [3]. UAV-mounted LiDAR and multispectral sensors enable accurate vegetation structure and species composition mapping. Studies have shown their effectiveness in monitoring savanna vegetation, crop phenology, plant canopy structures, etc [4]. This multi-scale integration is important for long-term environmental monitoring and resource management.

In precision agriculture, UAV-based imaging supports targeted resource allocation, irrigation planning, early detection of pests and diseases, etc [5]. In urban planning, UAV imagery assists in infrastructure inspection, traffic monitoring, land use classification, etc. In environmental surveillance, UAVs help track deforestation, monitor wetland changes, and assess coastal erosion [6]. These applications require high classification accuracy to ensure reliability in operational decision-making.

Despite advances in UAV sensing technologies, processing high-resolution aerial imagery remains challenging. Large image sizes demand high computational resources. Conventional neural networks often struggle to balance computational efficiency with classification accuracy [7]. Few models fail to capture complex spatial relationships, while deeper models can become computationally expensive. This problem becomes more

severe in time-sensitive applications, such as disaster response, security monitoring, etc, where faster classification is necessary [6].

Recent research has explored various deep learning architectures for UAV image classification. UAV images contain objects of varying sizes and shapes. Multi-scale feature fusion is an essential factor in such datasets. EfficientNet has gained attention due to its scalable architecture and balanced depth, width, and resolution parameters [8]. However, EfficientNet alone may not fully exploit fine-grained spatial features in the UAV imagery. Attention mechanisms have been introduced to focus on important image regions. However, integrating them with scalable architectures for aerial image classification requires careful design.

This research proposes Path Aggregation Network-based WATT-EffNet (PANet-WATT-EffNet), a novel architecture for UAV imagery classification. PANet-WATT-EffNet combines EfficientNet's scalability and efficiency with attention mechanisms and path aggregation [9, 10]. Wider Attention (WATT) modules capture fine-grained spatial features [9]. WATT-EffNet acts as a backbone, integrating a lightweight and fast architecture with an attention mechanism. The idea is to capture fine-grained global information from aerial imagery. PAN enables multi-scale feature fusion [10]. The design of PANet-WATT-EffNet is targeted to achieve high accuracy in complex aerial imagery classification.

Motivation for this research comes from the growing need for accurate and efficient UAV image classification. The proposed PANet-WATT-EffNet uses advanced attention modules and hierarchical feature aggregation. This improves feature extraction and classification accuracy while keeping computational cost low. The approach aims to enable reliable performance in operational applications.

2. Literature Survey

This survey explores various innovative methods for unmanned aerial image classification. UAV is designed to autonomously identify and accept two distinct loads of varying weights in the cargo reception area, utilizing image processing approaches. Moreover, based on their weight, UAV uses a weight sensor to transport loads to a designated drop location at a predefined GPS position. Trial outcomes indicate that the UAV successfully identified the loads, lifted them off the ground, and transported them to the designated load-shedding region. In terms of load recognition and classification, the prototype achieved an average success rate of 90% and 82.5%, respectively.

Two fully convolutional network frameworks are investigated [11]. One framework employed a dilated convolution layer without down-sampling. The other utilized a learned up-sampling convolution layer combined with down-sampling. Three basic classes were identified: structures, a mixed class comprising barren land and low vegetation, and a higher vegetation class.

A neural network was designed with higher-resolution KOMPSAT-3 satellite images to classify land cover with greater accuracy [12]. Training data was generated after obtaining satellite imagery of the coastal areas near Gyeongju City. Land cover was categorized into three classes: flora, land, and water, using neural network algorithms. The accuracy of classification results using deep learner models was 92.0%.

To improve the model's diagnosis and classification capabilities, as well as its generalization potential, a UAV dataset was created based on various UAV structural forms [13]. Subsequently, six classic deep convolutional neural network detection models are empirically evaluated using the transfer learning technique. The collected UAV test dataset is then subjected to experimental analysis. Compared to traditional recognition models, the transfer learning-based image classification method utilized in this study demonstrated significantly improved accuracy, recall, and precision.

Several image-based machine learning processes are investigated for UAV image diagnosis and classification [14]. The topic has garnered significant interest recently due to the exponential growth in the availability of UAVs, which are used for applications ranging from entertainment to defense missions, alongside the associated risks. Presently, the most commonly used technologies for UAV diagnosis and classification include acoustic, optical, radar, and radio frequency sensing systems. The study's findings highlight the efficiency of machine learning-based UAV classifications and the important separate contributions in this field.

A UAV is prototyped for autonomous load transportation [15]. It is a low-cost, high-mobility, autonomously flying rotary-wing UAV. An image processing module has been onboarded on the prototype. This UAV prototype achieved a success rate of 90% in load recognition and 82.5% in classification and transport tasks.

A parallel neural network model was designed for UAV Hyperspectral Image (HIS) classification [16]. An ICA-2D-CNN branch with a 3D-CNN branch was combined. ICA-2D-CNN branch extracts spatial features. 3D-CNN branch processes both spectral and spatial features simultaneously. Using UAV HSI data with a spatial resolution of 5 cm as an instance, the model achieved a total accuracy of 71.87%, an average accuracy of 72.9%, and a Kappa coefficient of 0.639. Performance of 1D-CNN, 2D-CNN, and 3D-CNN improved by 3.6%, 9.3%, and 6.5%, respectively.

An efficient approach was introduced to assess the readiness of rose crops for cultivation [17]. A deep learning model analyzed UAV footage to automatically distinguish between open and closed rosebuds across large cultivation areas. This technique integrated various algorithms to enhance detection and counting precision. This approach improved tracking stability and reduced duplicate counts. This model achieved a mAP of 94.1%, which highlights its robustness and practical efficiency. This method significantly decreases the time and effort required for rose crop monitoring, yield estimation, and cultivation management.

Despite these advancements, several research challenges remain in unmanned aerial image classification. Current deep learning models often face real-time processing and scalability limitations, especially when applied to large and high-resolution UAV datasets. In addition, many algorithms show poor adaptability to environmental variability, such as fluctuations in lighting, weather conditions, or object occlusions. Another important gap is the limited exploration of multimodal integration, where visual imagery could be combined with thermal, multispectral, or hyperspectral data for enhanced feature representation. Further, deep learning methods require large labeled datasets for effective training, but the creation of such datasets is costly and time-intensive. These factors emphasize the need for unsupervised and semi-supervised learning strategies that can reduce dependency on labeled data while improving classification robustness. Addressing these gaps is critical to increasing the dependability, efficiency, and practical applicability of aerial image classification systems in agriculture, environmental monitoring, and urban planning.

3. Proposed Methodology

In this research, PANet-WATT-EffNet, a novel Path Aggregation Network-based Wider Attention EfficientNet framework is designed (Figure 1). The target is to optimize feature extraction and classification tasks of aerial images. WATT-EffNet serves as the backbone, integrating a lightweight and fast architecture with attention mechanisms, to extract fine-grained and global information since aerial data [8, 9].

PANet incorporates multi-scale feature fusion [10]. This enables aggregation of multi-scale features across various levels of the backbone network. This integration enhances PANet-WATT-EffNet's capability to recognize complex patterns at various scales typically found in aerial images.

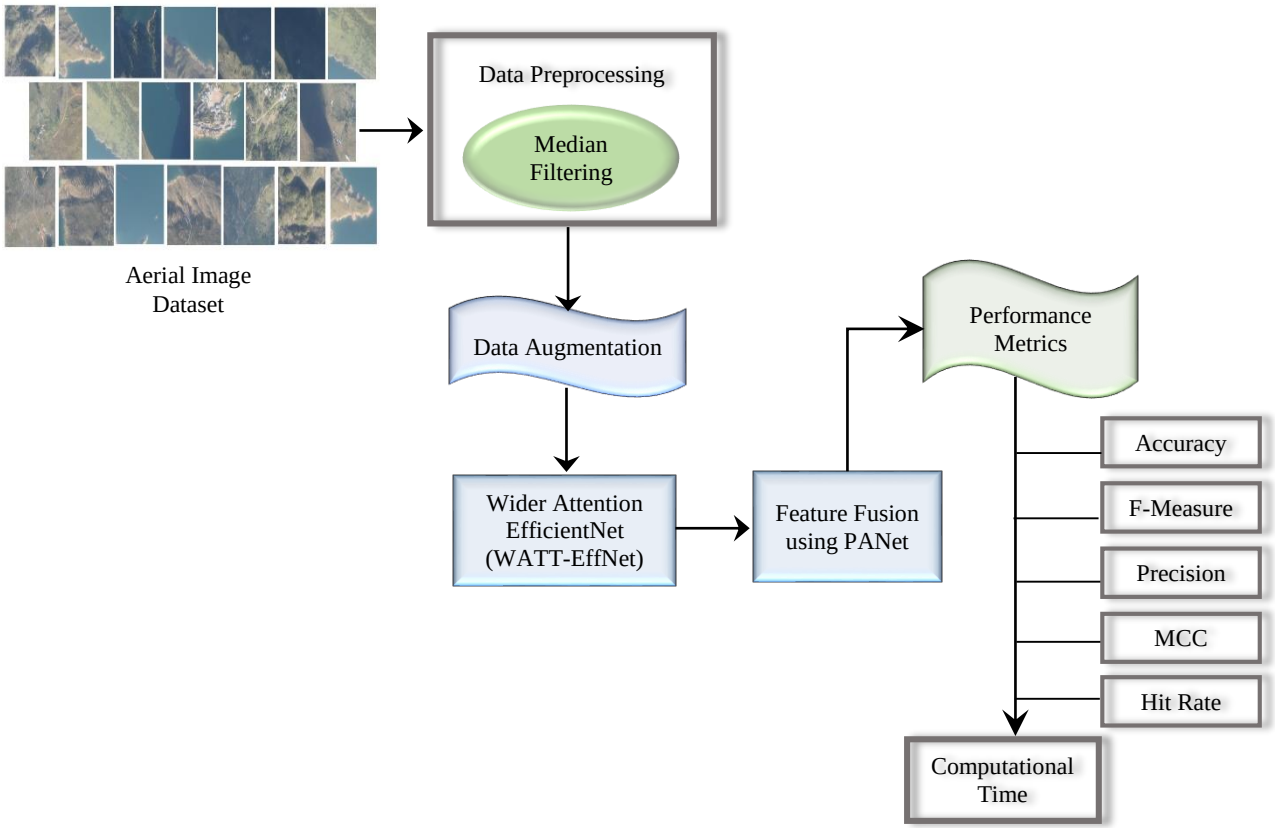

Fig. 1 Proposed PANet-WATT-EffNet learning model for unmanned aerial image classification

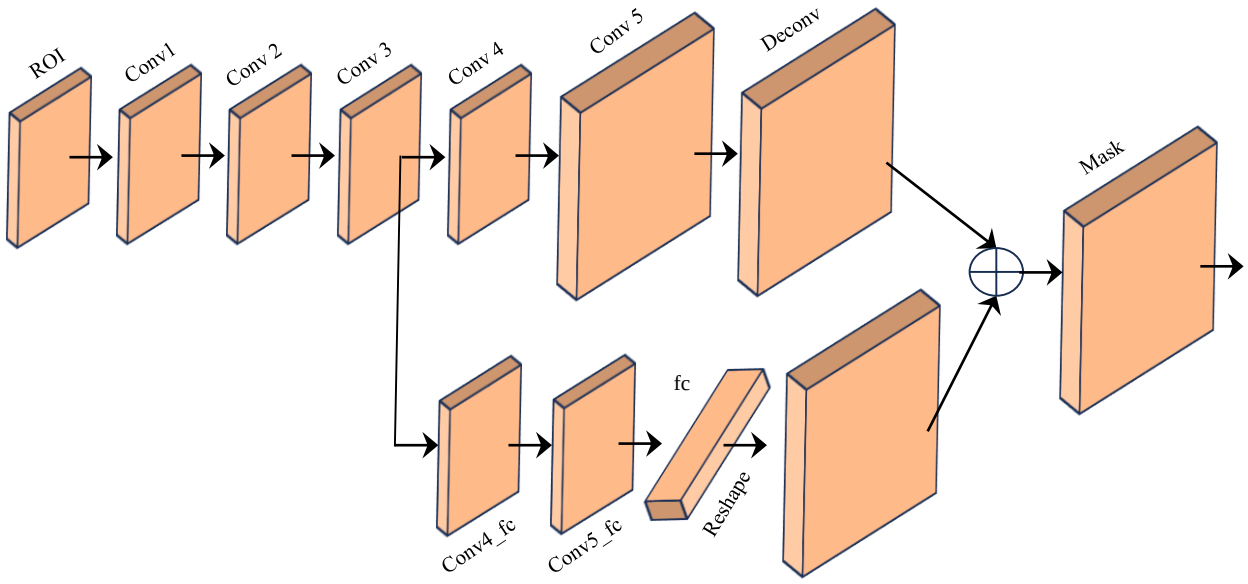

Fig. 2 Architecture of PANet learning model

3.1. Data Preprocessing and Augmentation

The Median Filtering (MF) approach is used to enhance image quality. The most recent indicators classify this method as a nonlinear signal processing paradigm. The median values of neighborhoods, known as masks, are used to correct erroneous digital representations. To replace inaccurate results, the median scores of a group are stored after the pixel is graded for its gray level. Any shape, such as a cross, square, round, or linear, can be used for the mask. Since MF is a nonlinear filter, performing numerical analysis on images containing random noise is quite challenging. The noise variance of MF would have a mean equal to zero if the image is categorized as having noise below the standard distribution.

$$\sigma_{med}^2 = \frac{1}{4nf^2(\overline{n})} \approx \frac{\sigma_i^2}{n+\frac{\pi}{2}-1} \cdot \frac{\pi}{2} \quad (1)$$

Where σ_i^2 describes input noise power, n signifies the size of MF, and $f(n)$ is noise intensity.

The average filtering noise variance is represented in Equation (2).

$$\sigma_0^2 = \frac{1}{n}\sigma_i^2 \quad (2)$$

Equations (1) and (2) are compared to determine the MF functions. By eliminating noise, the MF method significantly outperforms average filtering. Therefore, the MF method is more effective when the pulse width is smaller and the narrow pulse is farther away from impulse noise. When used with an average filtering model, the purpose of the MF technique is to maximize performance. Subsequently, a data augmentation procedure is performed with two operations: rotation and flipping.

3.2. PANet

PANet is a feature pyramid network architecture designed to enhance instance segmentation by improving feature representation through bottom-up path aggregation and an adaptive pooling mechanism (Figure 2).

PANet efficiently propagates information across all levels of the feature pyramid, which is particularly beneficial for detecting objects at different scales and handling complex instance segmentation tasks [18, 19].

3.2.1. Bottom-Up Path Augmentation

PANet enhances the traditional top-down Feature Pyramid Network (FPN) by introducing a bottom-up pathway. This addition strengthens low-level features with high-level semantic information and vice versa, thereby improving feature alignment and robustness.

Given a feature map F_l at level l of the pyramid, the original FPN computes top-down features F_l^{td} using Equation (3).

$$F_l^{td} = Upsample(F_{l+1}^{td}) + F_l \quad (3)$$

PANet introduces the bottom-up path F_l^{bu} by connecting features from low to high levels. Equation (4) shows the mathematical modeling of the bottom-up pathway.

This bottom-up augmentation integrates detailed spatial information from lower levels with abstract semantics from higher levels.

$$F_l^{bu} = Downsample(F_{l-1}^{bu}) + F_l \quad (4)$$

3.2.2. Adaptive Feature Pooling

To ensure that the features are well-aligned for Region of Interest(RoI) operations, PANet employs adaptive pooling across multiple feature pyramid levels. This process generates a unified feature representation for each RoI by pooling features from different levels.

Let *R* represent a region of interest. The adaptive pooling operation *AdaptivePool (R, F_l)* combines information across all levels.

$$F_R = \sum_l \omega_l \,.\, Pool(R, F_l) \quad (5)$$

Here, F_l is a pooled feature for region *R*, ω_l is the learned weight for each level *l*, and *Pool(R, F_l)* extracts features of *R* from level *l*.

3.2.3. Fully Connected Fusion

To further enhance instance segmentation, PANet employs a fully connected layer for feature fusion, ensuring improved information propagation. Final output is computed as in Equation (6).

$$F_{final} = FC(Concat(F_{l_1}, F_{l_2}, \ldots, F_{l_k})) \quad (6)$$

Here, *Concat()* aggregates features from all levels $l_1,\ldots,l_k$; *FC* is a fully connected layer that learns the optimal combination of these features.

3.2.4 Loss Functions

PANet uses standard loss functions, such as segmentation, combining classification loss Lcls, box regression loss Lbox, and mask prediction loss Lmask, as shown in Equation (7).

$$\boldsymbol{L = L_{cls} + L_{box} + L_{mask}} \quad (7)$$

These components are optimized together to achieve better segmentation and detection performance. By integrating these mechanisms, PANet outperforms traditional FPN-based architectures in tasks such as object diagnosis and instance segmentation

3.3. Wider Attention EfficientNet (WATT-EffNet)

Wider Attention EfficientNet (WATT-EffNet) is a modification of the EfficientNet architecture that incorporates wider layers and attention mechanisms to enhance representational efficiency and accuracy.

3.3.1. Baseline: EfficientNet Overview

EfficientNet scales a baseline model by optimizing width ω, depth *d*, and resolution *r* using compound scaling as shown in Equation (8).

$$EfficientNet(\phi) = \{\omega = \alpha^{\phi}, d = \beta^{\phi}, r = \gamma^{\phi}\} \quad (8)$$

Here, φ is the compound scaling factor; α, β, γ are constants satisfying $\alpha, \beta^2, \gamma^2 \approx 1$.

3.3.2. Wider Layers in WATT-EffNet

Wider Attention EfficientNet increases the width of layers while adhering to the scaling principles of EfficientNet.

$$\omega_{new} = k.\omega_{base} \quad (9)$$

Here, K>1 ensures the model incorporates more neurons per layer, leading to better feature extraction capacity. Width scaling preserves computational efficiency by maintaining a balance with depth and resolution.

3.3.3. Incorporation of Attention Mechanisms

Squeeze-and-Excitation (SE) Block, CBAM attention mechanisms are incorporated into PANet-WATT-EffNet in order to enhance focus on salient features. A typical attention module operates by computing attention weights.

SE Block

a) Squeeze: Global average pooling aggregates spatial features into channel descriptors.

$$z_c = \frac{1}{H.W}\sum_{i=1}^{H}\sum_{j=1}^{W} X_{i,j,c} \quad (10)$$

Here, $X_{i,j,c}$ is the feature map at spatial location *(i, j)* and channel *c*.

b) Excitation: A gating mechanism adjusts the scaling of channel activations. Where W_1, W_2 are learnable weights and σ is the sigmoid function.

$$s_c = \sigma(W_2.ReLU(W_1.z_c)) \quad (11)$$

c) Reweighting: Allows the network to focus on the most informative channels while ignoring irrelevant ones.

$$\hat{X}_{i,j,c} = s_c.X_{i,j,c} \quad (12)$$

Convolutional Block Attention Module (CBAM).

Attention is applied to both the spatial and channel dimensions. Channel attention is similar to SE Block. Spatial attention is defined in Equation (13).

$$M_s(X) = \sigma\left(f^{7\times7}([AvgPool(X); MaxPool(X)])\right) \quad (13)$$

Here, f^{7x7} is a convolution filter.

3.4. Final WATT-EffNet Architecture

The output of each block integrates attention-enhanced features.

$$Y_{block} = f(X;\Theta).M_a(X) + B \quad (14)$$

Here, f(X;Θ)represents the widened convolutional layers, M_a (X)is the attention map, and B is the bias.

The network structure balances the improved representational capacity of wider layers and attention, while scaling efficiently as defined in Equation (15).

$$WATT - EffNet(\phi) = \{k\omega, \beta^{\phi} d, \gamma^{\phi} r\} \quad (15)$$

In PANet-WATT-EffNet, PANet improves hierarchical feature fusion for better multi-scale feature handling in unmanned aerial image data. The WATT-EffNet model uses EfficientNet's lightweight design to achieve computational efficiency with high classification accuracy. Wider attention layers in WATT-EffNet highlight salient image regions, improving the capture of critical features for aerial image classification. The combined use of PANet and attention mechanisms increases robustness in processing composite scenes with varied backgrounds and occlusions.

4. Experimental Results and Discussion

4.1. Dataset Description

The dataset considered for experimentation comprises high-resolution aerial imagery captured by UAVs [20]. The dataset is designed to advance small object detection tasks using DL techniques. This version is based on the original dataset [21]. This dataset offers an organized split for effective model building and assessment. The dataset includes 717 training examples, 84 validation samples, and 43 testing samples, enabling researchers to systematically develop and compare algorithms for identifying and categorizing tiny, poorly represented objects in aerial photography. Sample aerial images are presented in Figure 3.

4.2. Environmental Setup

Windows 10 was used for the experiments, and the system specifications included an Intel(R) Xeon(R) W-2223 CPU, an Nvidia GeForce RTX 3060 GPU, and 80 GB of RAM. Python 3.8 was utilized as the programming language. CUDA 11.7 served as the acceleration environment, PyCharm was the development platform, and PyTorch 2.0.1 was the framework. This research employs existing methods, including Mask-RCNN [22], EDL-MMLCC [23], MLCG-OCNN [24], and DRL-GAN [25].

4.3. Results and Discussion

The proposed PANet-WATT-EffNet is evaluated against the following learning models.

Mask-RCNN [22]: To segment apple blossoms by instance, the Mask-RCNN algorithm was used. Various image enhancement methods were applied, and their effects on flower detection were evaluated.

Fig. 3 Sample images from aerial dataset

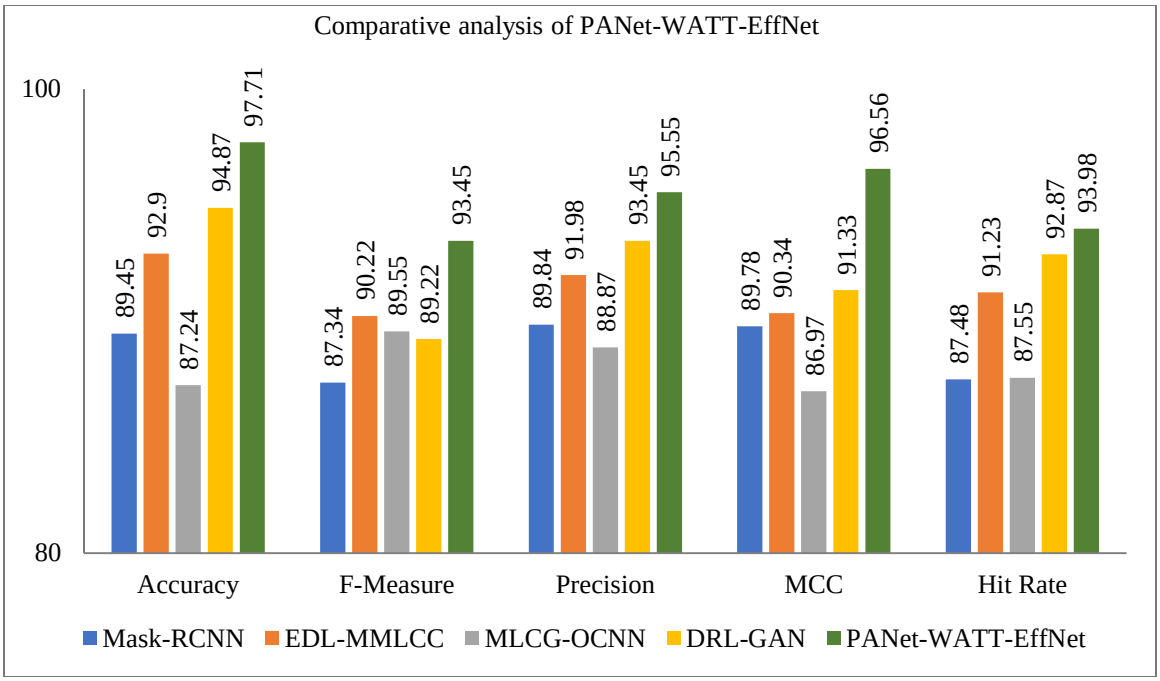

Fig. 4 Comparative analysis of PANet-WATT-EffNet

Table 1. Comparative analysis of PANet-WATT-EffNet

Learning Model	Accuracy	F-Measure	Precision	MCC	Hit Rate
Mask-RCNN	89.45	87.34	89.84	89.78	87.48
EDL-MMLCC	92.90	90.22	91.98	90.34	91.23
MLCG-OCNN	87.24	89.55	88.87	86.97	87.55
DRL-GAN	94.87	89.22	93.45	91.33	92.87
PANet-WATT-EffNet	97.71	93.45	95.55	96.56	93.98

EDL-MMLCC [23]: The goal of the EDL-MMLCC method is to categorize remote sensing images based on land cover, clouds, and shadows. Data augmentation techniques and preprocessing methods, such as median filtering, are commonly used.

MLCG-OCNN [24]: To achieve precise object classification, a feature-fusing OCNN is proposed. It learns higher-level features by integrating spectral patterns, object-level contextual information, and geometric attributes. Additionally, it incorporates object deformation coefficients as a supplement to the object contour-preserving mask technique. DRL-GAN [25]: A DRL-GAN utilizes high-resolution images as a guide to generate low-resolution data. It generates clearer and more detailed images that are better suited for recognition tasks. It also recovers missing information in both low-frequency and high-frequency components. This capability ensures that reconstructed images maintain structural consistency as well as fine details.

Table 1 and Figure 4 present a comparative analysis of various performance metrics across different models, showcasing their respective effectiveness in terms of Accuracy, F-Measure, precision, Matthews Correlation Coefficient (MCC), and Hit Rate.

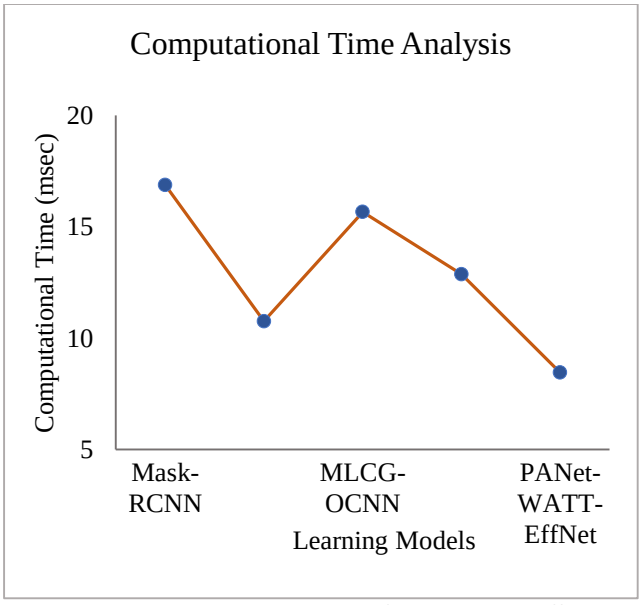

Fig. 5 Computational time analysis of PANet-WATT-EffNet

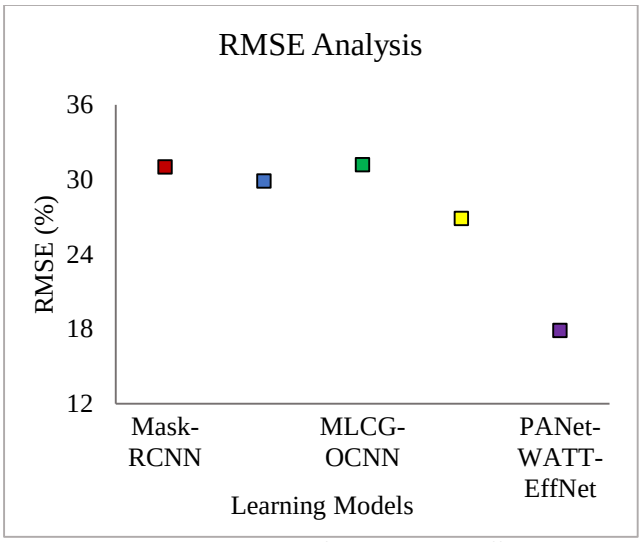

Fig. 6 RMSE analysis of PANet-WATT-EffNet

The proposed model outperforms all others, achieving the highest Accuracy (97.71%), F-Measure (93.45%), precision (95.55%), MCC (96.56%), and Hit Rate (93.98%). Among the remaining models, DRL-GAN follows closely, demonstrating strong performance, particularly in precision (93.45%) and Hit Rate (92.87%).

EDL-MMLCC also show solid results, excelling in Accuracy (92.9%), Hit Rate (91.23%), and F-Measure (91.98%). MLCG-OCNN, though performing reasonably well, exhibits the lowest values across all metrics, particularly in accuracy (87.24%). Overall, the proposed model demonstrates a significant improvement over the existing methods across all evaluated metrics.

Figure 5 presents computational time analysis of PANet-WATT-EffNet in comparison with other models. The PANet-WATT-EffNet model demonstrates the fastest computational time at 8.456 ms, outperforming all other models. Mask-RCNN takes the longest time to process, with a computational time of 16.876 ms. EDL-MMLCC and DRL-GAN show relatively efficient performance, taking 10.761 ms and 12.876 ms, respectively. MLCG-OCNN requires 15.665 ms to complete its computations. Overall, the PANet-WATT-EffNet demonstrated efficiency in computational time. Thus, making it a promising choice for real-time applications.

Figure 6 presents RMSE analysis of PANet-WATT-EffNet in comparison to other models. PANet-WATT-EffNet has the lowest RMSE of 17.87%, indicating superior accuracy and smaller prediction errors. In contrast, DRL-GAN follows with an RMSE of 26.87%, a relatively good presentation. EDL-MMLCC and Mask-RCNN report RMSE values of 29.87% and 30.99%, respectively, reflecting slightly higher prediction errors. Meanwhile, MLCG-OCNN records the highest RMSE at 31.18%, indicating a relatively larger deviation between forecast and actual values. Overall, the significantly lower RMSE of PANet-WATT-EffNet highlights its efficacy in reducing forecast errors.

PANet-WATT-EffNet's performance is evaluated against existing research using several key metrics. PANet-WATT-EffNet model achieved an accuracy of 97.71%, when compared to other models, Mask-RCNN (89.45%), EDL-MMLCC (92.9%), MLCG-OCNN (87.24%), and DRL-GAN (94.87%). PANet-WATT-EffNet model acquired F-Measure (93.45%), Precision (95.55%), and MCC (96.56%). These results showcase PANet-WATT-EffNet's capability to make balanced and reliable forecasts compared to other models. PANet-WATT-EffNet's Hit Rate (93.98%) emphasized its robustness. PANet-WATT-EffNet achieved computational effectiveness, with the lowest computational time (8.456 ms). RMSE (17.87) highlights its suitability for real-time requirements. Experimental results show that PANet-WATT-EffNet delivers higher classification accuracy.

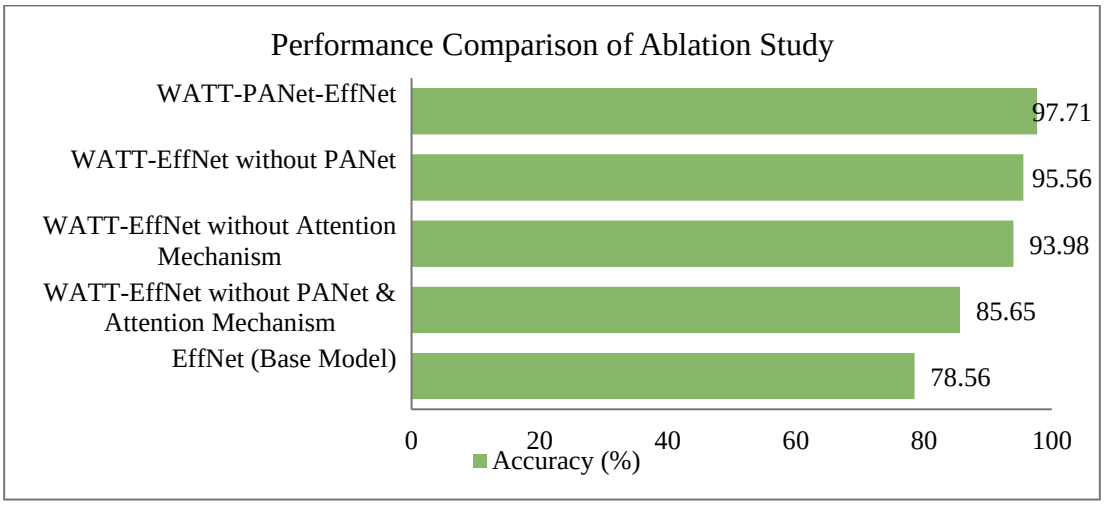

Fig. 7 Experimental results of the ablation study on PANet-WATT-EffNet

4.4. Ablation Study

This section presents the results of ablation tests conducted on PANet-WATT-EffNet. Figure 7 illustrates the results of the ablation study, highlighting the incremental contributions of various components to PANet-WATT-EffNet. PANet-WATT-EffNet achieved an accuracy of 97.71%, demonstrating the efficacy of its holistic architecture. Removing the PANet module reduced accuracy to 95.56%, while excluding the attention module further lowered it to 93.98%. When PANet and attention modules are removed together, accuracy dropped significantly to 85.65%. The base model, EffNet, achieved the lowest accuracy of 78.56%.

This result shows that the plain baseline struggled to capture complex patterns present in aerial imagery. When compared, it is clear that additional components, Wider Attention and PANet, played a vital role. These modules strengthened the feature extraction process and pushed the accuracy higher.

4.5. Influence of WATT-EffNet

The influence of WATT-EffNet lies in its capability, particularly in complex image classification tasks. By incorporating a wider attention mechanism alongside the efficient architecture of EfficientNet, WATT-EffNet increases PANet-WATT-EffNet's capacity to focus on critical features while reducing computational overhead.

A broader attention mechanism enabled WATT-EffNet to efficiently capture spatial hierarchies and context. This makes it highly suitable for applications such as UAV image classification.

5. Conclusion

This research addressed the challenge of classifying high-resolution UAV images. Conventional DL models often struggle to balance computational efficiency with accuracy. To overcome these limitations, a novel Path Aggregation Network-based Wider Attention EfficientNet (PANet-WATT-EffNet) was proposed. PANet-WATT-EffNet integrates the scalable design of EfficientNet with wider attention layers and a path aggregation mechanism. Thus, enabling multi-scale feature fusion and enhancing focus on salient image regions. A combination of these components strengthens PANet-WATT-EffNet's capability to capture both local and global information. This feature is crucial for handling complex aerial objects at various scales. Experimental results demonstrated that PANet-WATT-EffNet outperformed existing methods, achieving the highest accuracy of 97.71%, superior F-measure, MCC, and significantly lower RMSE. Reduced computational time further highlights the efficiency of PANet-WATT-EffNet, making it suitable for real-time aerial image classification tasks. Ablation study confirmed individual and collective contributions of wider attention and path aggregation modules. PANet-WATT-EffNet framework provides a robust and efficient solution for aerial image classification. Its capability to achieve high accuracy with reduced computational time makes it highly relevant for real-time applications, including precision agriculture, urban infrastructure monitoring, disaster response, etc. This research contributes to advancing UAV-based image analysis by combining scalable architectures with attention-driven feature refinement and effective multi-scale fusion, ensuring both reliability and efficiency in practical deployments.

References

[1] Emmanuel K. Raptis et al., "End-to-End Precision Agriculture UAV-based Functionalities Tailored to Field Characteristics," *Journal of Intelligent & Robotic Systems*, vol. 107, pp. 1-26, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[2] Peter B. Boucher et al., "Flying High: Sampling Savanna Vegetation with UAV-Lidar," *Methods in Ecology and Evolution*, vol. 14, no. 7, pp. 1668-1686, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[3] Thaísa F. Bergamo et al., "From UAV to PlanetScope: Upscaling Fractional Cover of an Invasive Species Rosa Rugosa," *Journal of Environmental Management*, vol. 336, pp. 1-12, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[4] Chenglong Zhang et al., "Feasibility Assessment of Tree-Level Flower Intensity Quantification from UAV RGB Imagery: A Triennial Study in an Apple Orchard," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 197, pp. 256-273, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[5] Parthasarathy Velusamy et al., "Unmanned Aerial Vehicles (UAV) in Precision Agriculture: Applications and Challenges," *Energies*, vol. 15, no. 1, pp. 1-19, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[6] Jean-Paul Yaacoub et al., "Security Analysis of Drones Systems: Attacks, Limitations, and Recommendations," *Internet of Things*, vol. 11, pp. 1-39, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[7] Chrysanthos Maraveas, "Incorporating Artificial Intelligence Technology in Smart Greenhouses: Current State of the Art," *Applied Sciences*, vol. 13, no. 1, pp. 1-35, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[8] Mingxing Tan, and Quoc Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," *Proceedings of 36th International Conference on Machine Learning* (ICML), vol. 97, pp. 6105-6114, 2019. [Google Scholar] [Publisher Link]

[9] Gao Yu Lee et al., "WATT-EffNet: A Lightweight and Accurate Model for Classifying Aerial Disaster Images," *IEEE Geoscience and Remote Sensing Letters*, vol. 20, pp. 1-5, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[10] Shu Liu et al., "Path Aggregation Network for Instance Segmentation," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 8759-8768, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[11] Nicholus Mboga et al., "Fully Convolutional Networks for Land Cover Classification from Historical Panchromatic Aerial Photographs," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 167, pp. 385-395, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[12] Gab Su Moon, Kyoung Seop Kim, and Yun Jea Choung, "Land Cover Classification based on High Resolution KOMPSAT-3 Satellite Imagery using Deep Neural Network Model," *Journal of the Korean Association of Geographic Information Studies*, vol. 23, no. 3, pp. 252-262, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[13] Wei Meng, and Meng Tia, "Unmanned Aerial Vehicle Classification and Detection Based on Deep Transfer Learning," *2020 International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI)*, Sanya, China, pp. 280-285, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[14] Nidhish Dubey, Nanduri Mahathi Sai Nithin, and Shrivishal Tripathi, "Analysis and Comparison of Image-Based UAV Detection and Identification," *2022 IEEE 9th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*, Prayagraj, India, pp. 1-6, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[15] Ahmet Furkan Büyükkelek et al., "Image Processing Based Transportation in Unmanned Aerial Vehicles," *2020 28th Signal Processing and Communications Applications Conference (SIU)*, Gaziantep, Turkey, pp. 1-4, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[16] Linhuan Jiang, Zhen Zhang, and Yabin Hu, "Unmanned Aerial Vehicle Hyperspectral Remote Sensing Image Classification Based on Parallel Convolutional Neural Network," *2023 11th International Conference on Information Systems and Computing Technology (ISCTech)*, Qingdao, China, pp. 526-530, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[17] David Herrera et al., "Combining Image Classification and Unmanned Aerial Vehicles to Estimate the State of Explorer Roses," *AgriEngineering*, vol. 6, no. 2, pp. 1008-1021, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[18] Shuangshuang Li, and Wenming Cao, "SEMPANet: A Modified Path Aggregation Network with Squeeze-Excitation for Scene Text Detection," *Sensors*, vol. 21, no. 8, pp. 1-16, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[19] Yunzuo Zhang et al., "Full-Scale Feature Aggregation and Grouping Feature Reconstruction-based UAV Image Target Detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1-11, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[20] Sovit Ranjan Rath, UAV - Small Object Detection Dataset, Kaggle, 2023. [Online]. Available: https://www.kaggle.com/datasets/sovitrath/uav-small-object-detection-dataset

[21] Weihancug, 10-Category UAV Small-Weak Object Detection Dataset (UAVOD10), GitHub Repository, 2022. [Online]. Available: https://github.com/weihancug/10-category-UAV-small-weak-object-detection-dataset-UAVOD10

[22] Uddhav Bhattarai et al., "Automatic Blossom Detection in Apple Trees using Deep Learning," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 15810-15815, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[23] Gyanendra Prasad Joshi et al., "Ensemble of Deep Learning-Based Multimodal Remote Sensing Image Classification Model on Unmanned Aerial Vehicle Networks," *Mathematics*, vol. 9, no. 22, pp. 1-17, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[24] Chenxiao Zhang et al., "A Multi-Level Context-Guided Classification Method with Object-Based Convolutional Neural Network for Land Cover Classification using Very High Resolution Remote Sensing Images," *International Journal of Applied Earth Observation and Geoinformation*, vol. 88, pp. 1-13, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[25] Yue Xi et al., "DRL-GAN: Dual-Stream Representation Learning GAN for Low-Resolution Image Classification in UAV Applications," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 1705-1716, 2021. [CrossRef] [Google Scholar] [Publisher Link]