# A Lightweight Convolutional Neural Network with Hierarchical Multi-scale Features Fusion for Gastrointestinal Disease Classification in Endoscopic Images

Adama DEMBELE [1] Waweru Mwangi [2] Ananda Omutokoh Kube3 [3]

[1]PAUSTI, Kenya [2,3]School of Computing and Information Technology,Kenya

## Abstract

Endoscopic images of gastrointestinal diseases are difficult to analyze due to low contrast, noise, illumination variations, and variability of disease manifestations. This work proposes a lightweight convolutional neural network with hierarchical multi-scale features Fusion for Gastrointestinal Disease Classification in Endoscopic Images. The proposed method uses a lightweight convolutional neural network architecture optimized for resource-constrained environments. Hierarchical Multi-scale Features Fusion improves disease classification by capturing fine-grained details and contextual information. The proposed neural network promises accurate and efficient endoscopic diagnosis of gastrointestinal diseases, improving patient care and medical decision-making.

## Introduction

Gastrointestinal diseases are a major public health concern worldwide, necessitating timely and accurate diagnosis in order to provide effective treatment and improve patient outcomes[2]. However, due to the complexity and variability of disease manifestations, manual interpretation of endoscopic images is time-consuming and prone to error.

In addition to these challenges, automated approaches using deep learning commonly face resource constraints, such as:

- Computational demands,
- Memory requirements,
- Inference Speed.

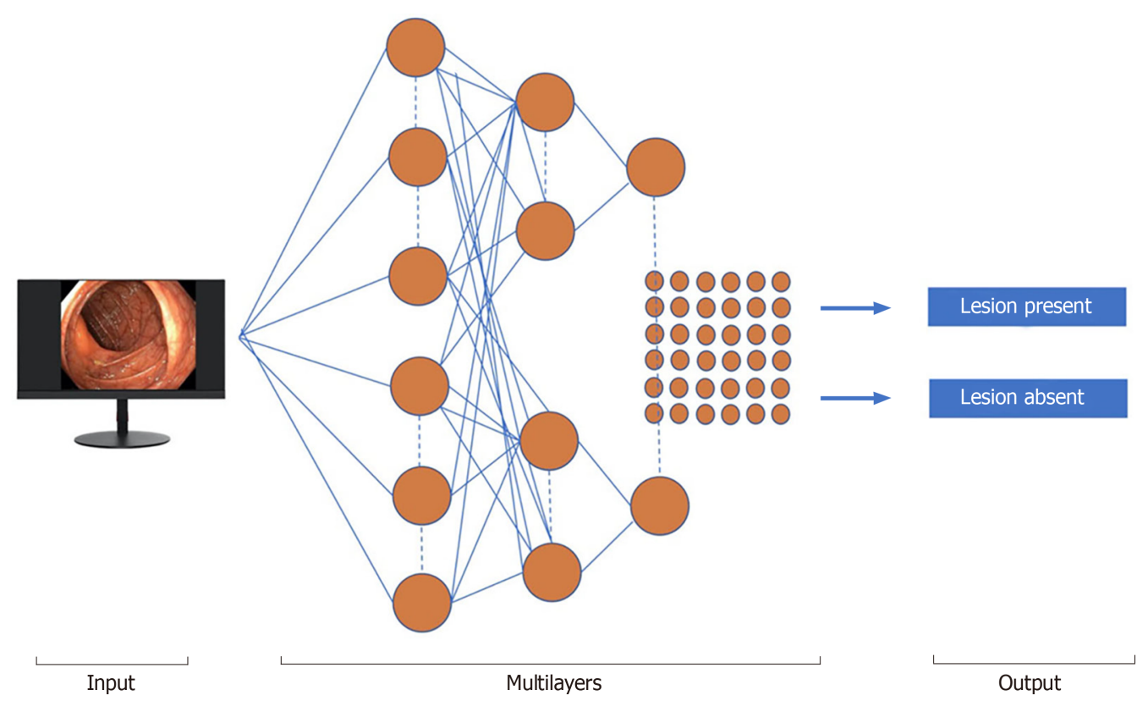

Figure 1. Illustration of classifying Endoscopic Image of Gastrointestinal Disease using CNN model

To address these challenges effectively, We proposed approach leverages a lightweight CNN architecture and hierarchical multi-scale features fusion to achieve efficient and accurate disease classification.

The research aims to provide healthcare professionals with a cost-effective and automated solution for improved gastrointestinal disease diagnosis.

## Methodology

The proposed method combines MobileNetV1 with the HMFF module, enhancing feature representation through multi-scale features fusion. MobileNetV1[1] is a lightweight CNN designed for mobile and embedded vision tasks, utilizing depth-wise separable convolutions to reduce complexity. The HMFF module extracts multi-scale information from MobileNetV1's layers and enables dense connectivity for improved information flow, capturing local and global patterns. This integration results in a powerful and efficient architecture, ideal for accurate image analysis in resource-constrained environments.

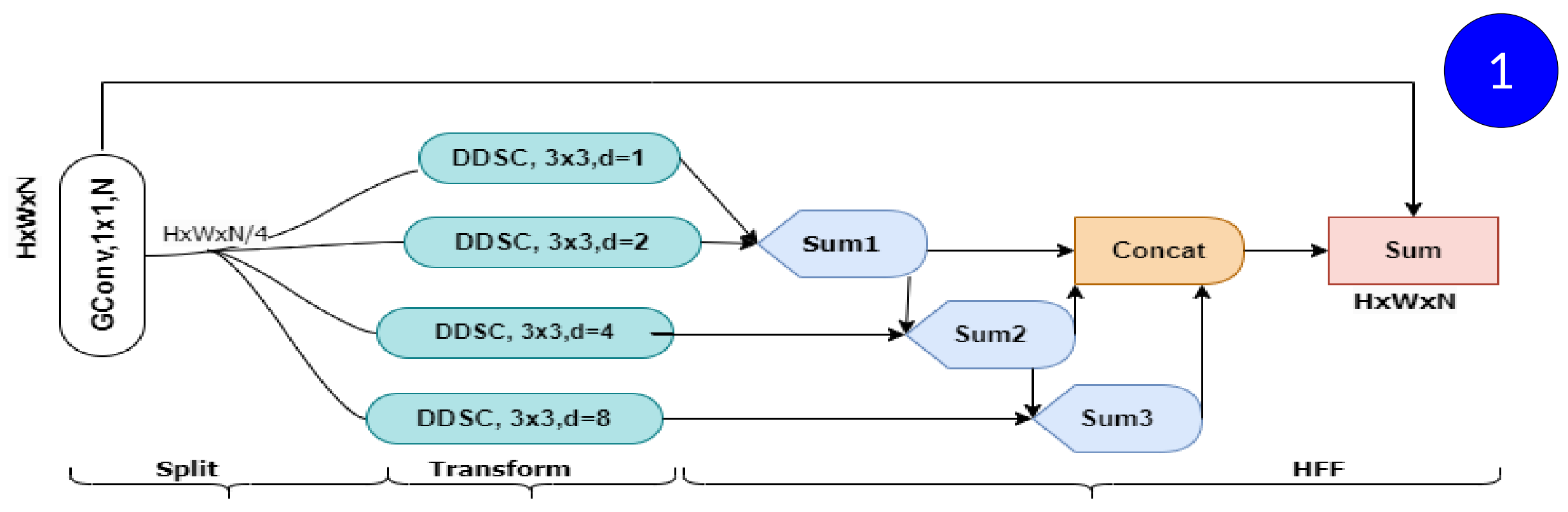

Figure 2. Hierarchical Multi-scale Feature Fusion (HMFF) module

Where $GConv, 1 \times 1$ is grouped point-wise Convolution defined as input of HMFF module, $N$: number of filters, DDSC is Dilated Depth-wise Separable Convolution, $concat$ is fusion operator concatenate.

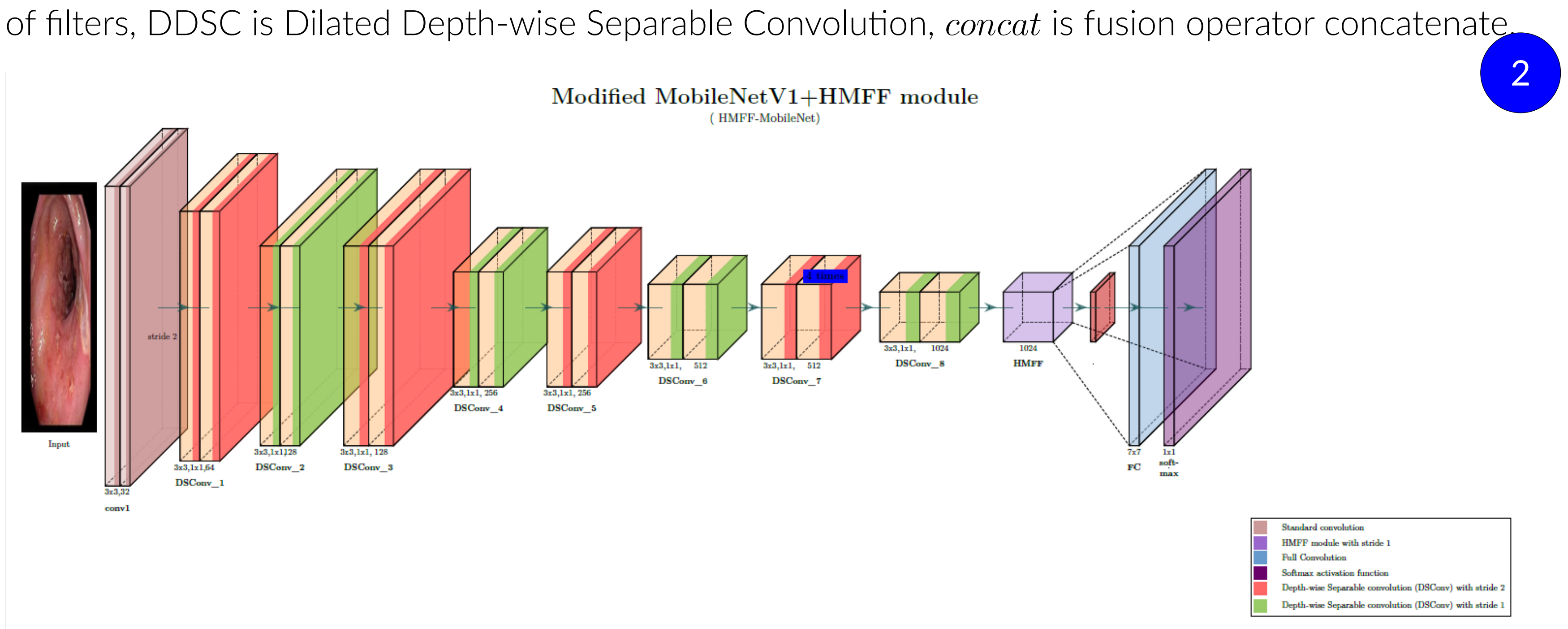

Figure 3. Combined MobileNetV1 with HMFF Module Architecture

The aim of our suggested Lightweight CNN is to balance model accuracy and complexity. We reduce the number of parameters by using a simplified architecture, which enables quick and resource-effective inference. Additionally, we make use of Hierarchical Multi-scale Features Fusion, which gives our model the ability to capture contextual data as well as fine-grained information, improving the performance of disease classification. End-to-end training of the model enables automatic feature learning and adaptability to various disease patterns.

## Results on Endoscopic Images of Gastrointestinal Disease(KvasirV1)

Table 1 presents a comparison of various HMFF-MobileNet with MobileNetV1, where the name indicates the minimum number of input channels per group, such as "HMFF-MobileNet 32ch" having a minimum of 32 input channels per group, in comparison to KvasirV1 experiments.

| Model | Trainable params | reduction | FLOPs | reduction | Test accuracy |
|---|---|---|---|---|---|
| MobilenetV1 | 3, 217, 226 | 0% | 567, 751, 710 | 0% | 71.75% |
| HMFF-Mobilenet 32Ch | 375, 882 | 88.31% | 148, 224, 350 | 73.89% | 72.88% |
| HMFF-Mobilenet 64Ch | 633, 930 | 80.29% | 238, 541, 150 | 57.98% | **74.38%** |
| HMFF-Mobilenet 128Ch | 1, 125, 194 | 65.02% | 341, 903, 582 | 39.78% | 74.25% |

Table 1. After 150 epochs, the KvasirV1 dataset showed the following results.

## Evaluation criteria: Precision and Recall

figure 4, the HMFF-Mobilenet models appear to have performed well on some classes, including "normal-cecum" and "normal-pylorus" with high precision and recall values. Moreover, it underperformed in other classes, which including "dyed-resection margins" and "polyps," with low recall and precision values.

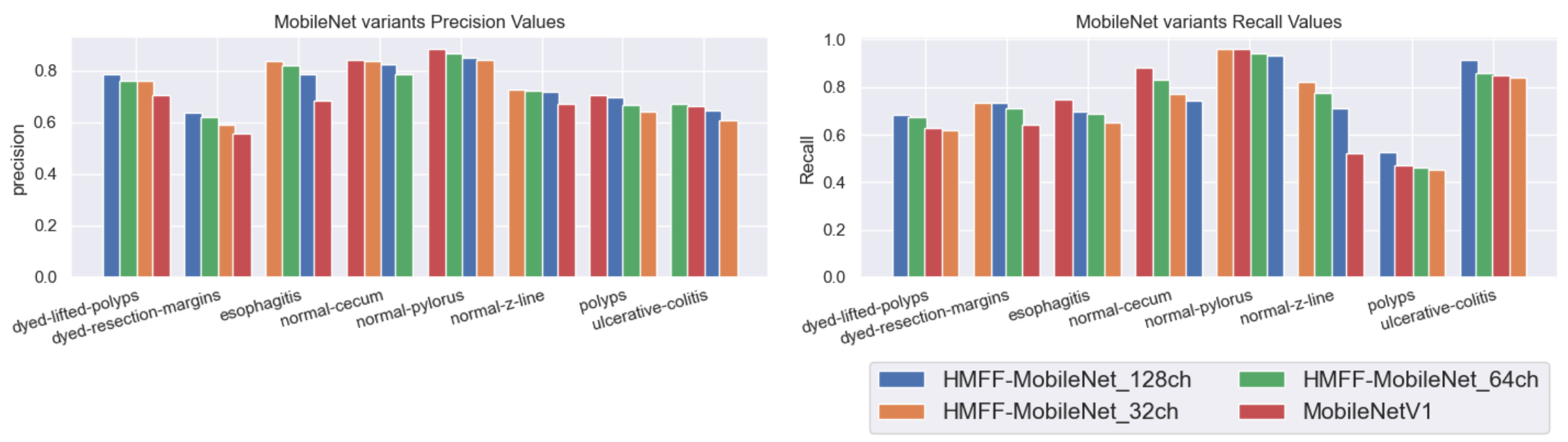

Figure 4. Performance analysis of proposed method with the MobileNet baseline

## Network Visualization Results

Grad-CAM[3] was used to visualize the heatmap of the HMFF module's layer feature maps in the various variants of network in order to more intuitively validate the effectiveness of the module. Eight images were selected from the KvasirV1 testing set.

Figure 5 makes it abundantly clear that the network's Grad-CAM mask with the HMFF module can cover the target object region better than MobileNetV1. In other words, the network connected to the HMFF module develops the ability to utilize the data from the region of the target object and combine features from it.

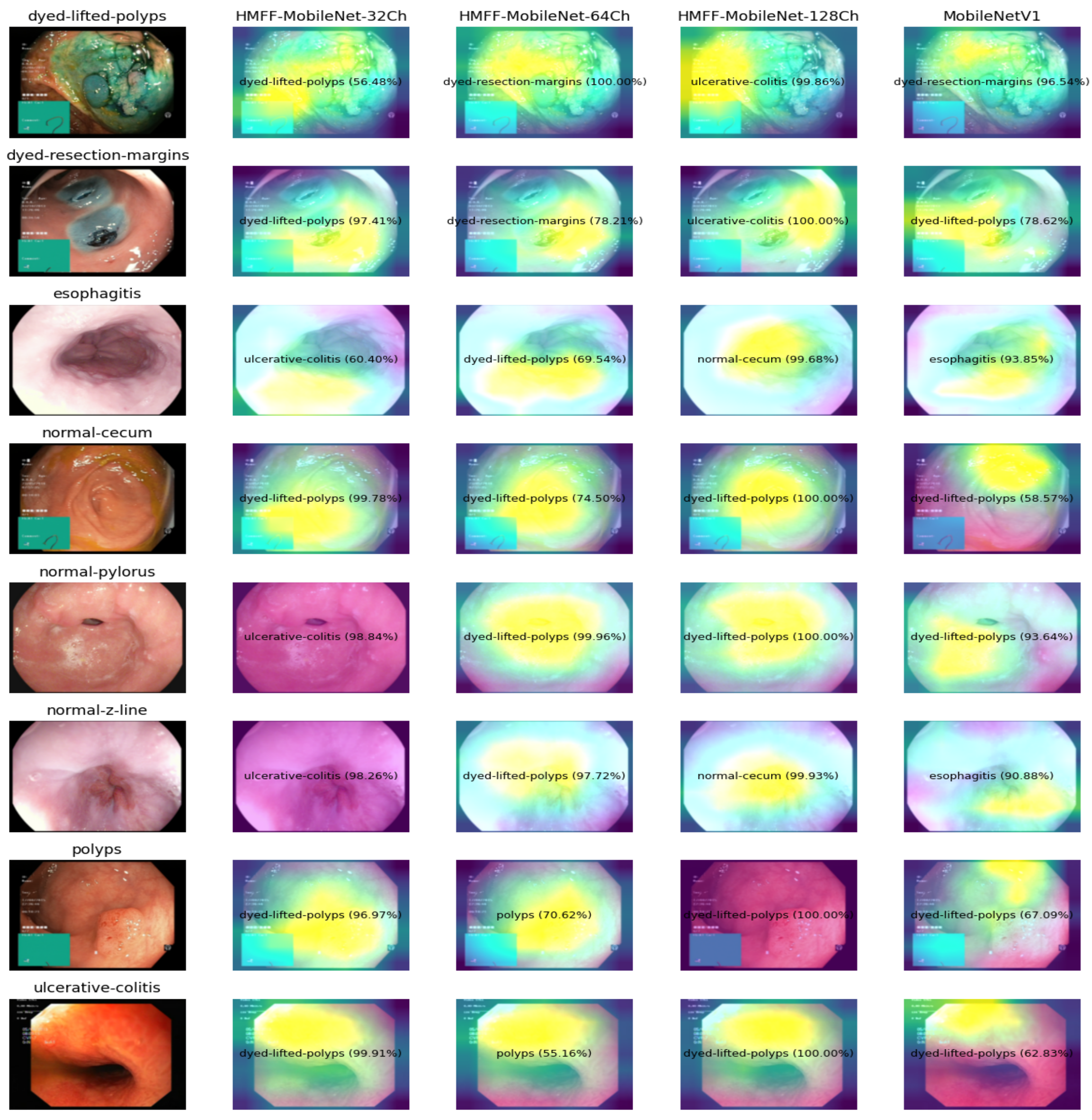
dyed-lifted-polyps
HMFF-MobileNet-32Ch
HMFF-MobileNet-64Ch
HMFF-MobileNet-128Ch
MobileNetV1
dyed-lifted-polyps (56.48%)
dyed-resection-margins (100.00%)
ulcerative-colitis (99.86%)
dyed-resection-margins (96.54%)
dyed-resection-margins
dyed-lifted-polyps (97.41%)
dyed-resection-margins (78.21%)
ulcerative-colitis (100.00%)
dyed-lifted-polyps (78.62%)
esophagitis
ulcerative-colitis (60.40%)
dyed-lifted-polyps (69.54%)
normal-cecum (99.68%)
esophagitis (93.85%)
normal-cecum
dyed-lifted-polyps (99.78%)
dyed-lifted-polyps (74.50%)
dyed-lifted-polyps (100.00%)
dyed-lifted-polyps (58.57%)
normal-pylorus
ulcerative-colitis (98.84%)
dyed-lifted-polyps (99.96%)
dyed-lifted-polyps (100.00%)
dyed-lifted-polyps (93.64%)
normal-z-line
ulcerative-colitis (98.26%)
dyed-lifted-polyps (97.72%)
normal-cecum (99.93%)
esophagitis (90.88%)
polyps
dyed-lifted-polyps (96.97%)
polyps (70.62%)
dyed-lifted-polyps (100.00%)
dyed-lifted-polyps (67.09%)
ulcerative-colitis
dyed-lifted-polyps (99.91%)
polyps (55.16%)
dyed-lifted-polyps (100.00%)
dyed-lifted-polyps (62.83%)

Figure 5. Sample visualization on Kvasir testing split set generated by Grad-CAM. All target layers selected are "HMFF's layer".

Figure 6 shows that HMFF-MobileNet-128Ch has the lowest loss compared to the baseline MobileNetV1 model with fewer parameters and float point operation. This indicates that HMFF-MobileNet-128Ch is not only more efficient in terms of computational resources but also achieves better performance in terms of minimizing loss. These results suggest that the proposed HMFF-MobileNet-128Ch model can be a promising choice for applications where low loss is a critical factor.

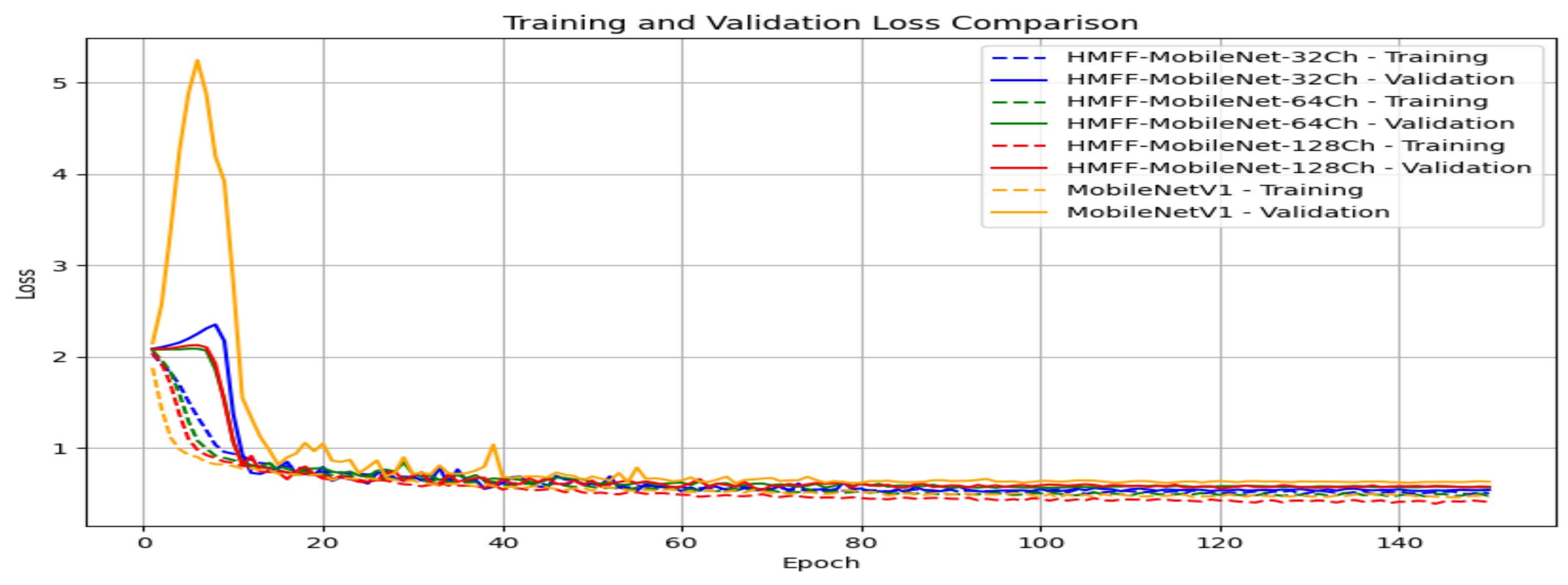

Figure 6. Comparison of Training and Validation Loss on Endoscopic Images of Gastrointestinal Disease for Different Models

## Conclusion

Our Lightweight Convolutional Neural Network with Multi-scale Hierarchical Features Fusion is a potentially effective method for correctly classifying gastrointestinal diseases in endoscopic images. This technology has the potential to significantly improve patient outcomes and advance the practice of gastroenterology by offering real-time decision support and effective inference.

## References

[1] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

[2] Gaetano Cristian Morreale, Emanuele Sinagra, Alessandro Vitello, Endrit Shahini, Erjon Shahini, and Marcello Maida. Emerging artificial intelligence applications in gastroenterology: A review of the literature. *Artificial Intelligence in Gastrointestinal Endoscopy*, 1(1):6–18, 2020.

[3] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017.