

ENHANCED MULTI-CANCER CLASSIFICATION USING ATTENTION-DRIVEN CNNs, CONTOUR FEATURE FUSION, AND VISION TRANSFORMERS

Waleed Nouman¹, Hysam Amran², Ahmad Raza³

Abstract—This paper extends the multi-cancer image classification work of the study by Sharma et al. [1] by incorporating several architectural enhancements. We integrate Efficient Channel Attention (ECA) and the Global Attention Mechanism (GAM) into DenseNet variants and introduce a novel hybrid framework combining attention-enhanced convolutional representations with contour-based geometric descriptors. We further evaluate segmentation-based preprocessing through an ablation study and introduce a unified multi-cancer dataset enabling cross-cancer generalization. Finally, a Vision Transformer baseline using MaxViT is included for comparison. Experimental results demonstrate consistent improvements across multiple cancer imaging datasets, validating the effectiveness of the proposed techniques.

Index Terms—multi-cancer classification, medical imaging, DenseNet, ECA, GAM, contour fusion, MaxViT, segmentation

I. INTRODUCTION

Automated cancer classification from medical imaging has gained significant attention due to its potential to support clinical decision-making. Sharma et al. [1] demonstrated the effectiveness of convolutional neural networks (CNNs) across multiple cancer types using segmentation, classical deep architectures, and a contour-based feature extraction approach. While their results were promising, several opportunities exist to enhance the representational capability and generalizability of such models.

Motivated by these gaps, this work introduces a collection of improvements to advance the original framework: (1) incorporation of lightweight and global attention mechanisms, (2) integration of contour features directly into end-to-end model training, (3) creation of a hybrid CNN–geometric model, (4) introduction of a Vision Transformer baseline, (5) segmentation-based ablation analysis, and (6) creation of a unified multi-cancer dataset supporting generalized learning.

II. RELATED WORK

Sharma et al. [1] evaluated ten CNN architectures including DenseNet201, DenseNet121, InceptionResNetV2, InceptionV3, MobileNetV2, NasNetLarge, NasNetMobile, ResNet152V2, VGG19, and Xception across multiple medical datasets. Their work also incorporated segmentation and handcrafted contour measurements to enhance classification performance.

However, their approach extracted contour features independently from the neural network. Attention-based CNN enhancements and Transformer-based models were not explored. This study addresses these limitations and expands the experimental landscape.

III. METHODOLOGY

A. Datasets

We use the same datasets as Sharma et al.: five publicly available Kaggle datasets covering leukemia, breast cancer, cervical cancer, kidney abnormalities, and lung/colon malignancies. In addition, we introduce a unified multi-cancer dataset constructed by merging all image sets and harmonizing label structures.

B. Segmentation Framework

We apply UNet-based segmentation as in the original study but extend the analysis with a formal ablation comparing raw vs. segmented images across all architectures.

C. Attention-Enhanced CNN Architectures

1) *Efficient Channel Attention (ECA)*: ECA blocks are added after each Dense block in DenseNet121, improving channel-level discrimination without significant computational cost.

2) *Global Attention Mechanism (GAM)*: GAM introduces both channel and spatial attention, allowing the model to capture global contextual relationships.

D. Contour Feature Fusion

Unlike the original study, which used contour descriptors only for evaluation, our work integrates geometric features (perimeter, area, epsilon, vertices) directly with CNN embeddings via feature concatenation.

E. Hybrid Architecture: CNN + Contours

We propose a novel hybrid model combining:

- DenseNet121 + ECA features
- Contour-based geometric descriptors

The fused representation improves classification robustness, especially where morphological structure is diagnostic.

F. Vision Transformer Baseline

We implement MaxViT as a Transformer-based baseline to compare attention-heavy CNNs with modern ViTs.

IV. EXPERIMENTS

We evaluate all models with and without segmentation, with and without contour fusion, and across individual and unified datasets. Metrics include accuracy, F1-score, ROC-AUC, and inference time.

V. RESULTS

Due to space constraints, figures and tables should be inserted later. Our preliminary tests show that:

- DenseNet121-ECA outperform baseline DenseNet121 across all datasets.
- GAM variants deliver improved global context extraction.
- Hybrid (CNN + contour fusion) achieves the best performance overall.
- MaxViT performs competitively but is computationally heavier.

VI. CONCLUSION

This work expands and improves upon the multi-cancer classification framework proposed by Sharma et al. [1]. By integrating attention mechanisms, contour fusion, hybrid architectures, segmentation ablations, and Transformer baselines, we offer a more comprehensive and robust system for medical image classification.

REFERENCES

- [1] A. Sharma, S. Kumar, and A. Chopra, "A Comprehensive Deep-Learning Approach for Multi-Cancer Classification using Contour Features, Segmentation, and Feature Extraction Techniques," *Scientific Reports*, 2024.