International Academic Institute for Science and Technology

International Academic Journal of Science and Engineering
Vol. 13, No. 3, 2026, pp. 176-191.

ISSN 2454-3896

International Academic Journal of Science and Engineering

www.iaiest.com

BGT-Net: Boundary-Graph Transformer Network for Segmentation-Guided Seven-Class Dermoscopic Classification

P. Nithin [a*], G. Muthulakshmi [b], B. Thanga Parvathi [c], D. Jebakumar Immanuel [d]

[a] Department of Computer Science and Engineering, Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu, India. E-mail: nithinresearch07@gmail.com, Orcid: https://orcid.org/0009-0002-3635-4022*
[b] Department of Computer Science and Engineering, Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu, India. E-mail: gmuthulakshmi_cse@msuniv.ac.in, Orcid: https://orcid.org/0000-0002-3455-6184
[c] Department of Computer Technology, Bannari Amman Institute of Technology, Erode, Tamil Nadu, India. E-mail: btparvathi@bitsathy.ac.in, Orcid: https://orcid.org/0009-0000-1298-6565
[d] Associate Professor, Department of Artificial Intelligence and Data Science, Karpagam Institute of Technology, Coimbatore, Tamil Nadu, India. E-mail: jebakumarimmanuel@gmail.com, Orcid: https://orcid.org/0000-0003-0786-8200

Abstract

Accurate multi-class classification of dermoscopic skin lesions is still difficult because of irregular lesion morphology, fine differences between classes, and extreme class imbalance in clinical datasets. In this regard, this work presents a novel segmentation-guided hybrid approach called Boundary-Graph Transformer Network (BGT-Net). It improves the classification of seven classes of dermoscopic skin lesions (AKIEC, BCC, BKL, DF, MEL, NV, and VASC) by jointly modeling lesion geometry and internal texture patterns. The proposed approach begins with the use of high-quality manual segmentation to separate the lesion area and obtain accurate boundary information. A graph attention module is used to model geometric irregularity, boundary curvature, and asymmetry patterns based on boundary nodes. Meanwhile, a lightweight transformer branch is used to learn intra-lesion texture patterns based on segmented region patches. Cross-attention fusion is used to combine geometric and texture features for classification. Experimental results on a seven-class dermoscopic skin lesion dataset show that BGT-Net improves classification accuracy to 97.8%, macro F1 score of 0.96, and an AUC of 0.99, which is 12.4% higher than EfficientNet-B4 in accuracy and 0.15 higher in macro-F1. Moreover, the recall of minority classes like DF and VASC is substantially improved and more robust to class imbalance. Thus, the combination of boundary-aware graph modeling and transformer-based texture modeling improves dermoscopic skin lesion classification.

Keywords: Skin Lesion Classification, Segmentation-Guided Learning, Graph Attention Network (GAT), Transformer-Based Feature Fusion, Dermoscopic Image Analysis.

DOI: 10.71086/IAJSE/V13I3/IAJSE13106
Received: 29 May, 2026/Revised: 07 July 2026/Accepted: 14 August 2026/Published: 15 September 2026

Introduction

Skin cancer is rapidly increasing worldwide, and melanoma and other skin cancers are responsible for most of the morbidity and mortality. Early detection and correct diagnosis are very important, as chances of survival are very high if malignant patches are identified early. Dermoscopy has emerged as an important non-invasive technique, allowing the dermatologist to visualize what lies beneath the skin's surface and to distinguish more accurately between benign and malignant lesions. However, image interpretation is still very much dependent on the experience of the dermatologist, and subtle variations in the classes can result in varying levels of accuracy. This is why computer-aided diagnosis using deep learning algorithms has recently gained a lot of attention (Hammar et al., 1985). EfficientNet and ResNets have shown promising results in dermoscopic image classification. Recently, transformer models have been introduced to better understand the global context of the lesion image. However, most of the models are focused on the texture of the entire image and do not pay much attention to the boundary shape of the lesion. Moreover, the presence of extreme class imbalance and small lesion sizes makes it difficult to rely on the models, particularly for the minority classes such as Dermatofibroma (DF) and Vascular lesions (VASC) Rezaei, (2021).

The challenges clearly indicate the existence of a gap in the literature. Currently, there is no properly integrated solution that simultaneously addresses geometric boundary irregularities and texture dependencies within the lesion in a segmentation-guided manner. Most of the current literature is focused on segmentation and classification as two separate tasks, without realizing the potential of using boundary information as a discriminative feature for multi-class dermoscopic image classification (Gorelick et al., 2013). To address this gap, this paper proposes a novel segmentation-guided classification framework that combines geometric boundary modeling with a transformer-driven texture learning paradigm for seven-class dermoscopic lesion image classification. By properly incorporating contour irregularities and mixing them with contextual texture information, this proposed framework aims to improve the robustness of the classification task, particularly for the challenging and minority lesion classes (Goyal et al., 2020; Murtozalikhon, 2024).

The main contributions of this work are summarized as follows:

- A graph modeling method that is aware of the boundary and reflects the irregularity and asymmetry of the lesion shape.
- A fusion of geometric boundary embeddings with transformer-based texture representations using cross-attention.
- A segmentation-guided multi-branch architecture to improve minority class performance for a seven-class dermoscopic image classification task.
- Comprehensive experiments demonstrating improved accuracy and macro-F1 compared to conventional CNN-based approaches.

The remainder of this paper is organized as follows. Section 2 presents an overview of the related work in dermoscopic image segmentation and classification. Section 3 presents the proposed approach. Section 4 presents the results and comparison. Section 5 concludes and discusses potential future work.

Related Work

Review the impact of artificial intelligence, particularly segmentation-driven pipelines, on cancer diagnosis in medical imaging. This section discusses traditional image processing, CNN-based solutions (such as U-Nets), and more recent deep learning solutions, while emphasizing the importance of assessment and the need in clinical practice to consider segmentation as an essential step for accurate diagnosis and planning (Ramalakshmi et al., 2024). Describe an explainable solution for skin cancer diagnosis based on a parallel attention mechanism to improve segmentation and classification. Emphasize the importance of explainability through attention and visualization and achieve outstanding classification performance on common dermoscopy datasets, suggesting that parallel attention is beneficial for better lesion localization and less confusion between similar classes (Thaljaoui et al., 2026). A transformer-based solution for multi-class skin cancer segmentation in histopathology images. The aim is to leverage self-attention to represent long-range dependencies and complex tissue patterns that are difficult to understand for convolutional encoders, focusing on the pixel-level segmentation of cancerous regions in multiple classes in pathology images (Imran et al., 2024).

A two-stage learning framework for whole-slide patch classification in breast cancer images that reduces the need for manual annotation. It combines semi-supervised contrastive pretraining with confidence-filtered pseudo-labeling to

learn robust representations from limited annotations and then accurately extend the training set with pseudo-labels (Zhaksylyk et al., 2026). Advance skin cancer diagnosis by integrating patient data clinical, demographic, and contextual information with images of lesions. Their multimodal learning method demonstrates enhanced discrimination over image-only models, especially for ambiguous samples where subtle visual differences can be complemented by patient context (Islam et al., 2026). Describe a knowledge distillation pipeline to transfer the strengths of a large and complex "teacher" model to a smaller "student" model. The aim is to provide robust skin cancer classification on resource-constrained devices, accomplished through soft targets and feature and logit alignment to maintain diagnostic power (Saha et al., 2025). Make the case for a more robust diagnostic tool based on deep CNNs and ensemble learning. By combining multiple models or multiple feature learners, the hope is to improve robustness and lower variance, overall improving generalization when class boundaries are close in the visual space (Khan et al., 2025). Describe a new classification method assisted by transformers, combining feature encoding in the style of VGG19 models with transformer-inspired modeling to improve global context reasoning. The main idea is to combine excellent CNN feature extraction with attention mechanisms that improve relationship understanding between lesion areas (Najjar et al., 2025)

Focus on radiotherapy tasks, using deep learning for automatic per-fraction segmentation of gross tumor volume and organs-at-risk in cervical cancer adaptive radiotherapy. Highlight the importance of accurate and repeated segmentation over time to enable safe dose planning in the presence of anatomical changes (Breto et al., 2022). Describe a Triple Attention CNN for skin cancer image segmentation and classification. Their work employs multiple attention modules (such as channel, spatial, and task-driven attention) to focus on important lesion information and filter out background noise for both pixel-level and image-level tasks Wang et al., (2026). Launch SkinSAM, a modified version of the Segment Anything Model, to address skin lesion segmentation. By fine-tuning the SAM with dermoscopic data, the obtain excellent segmentation performance (imagine Dice/IoU improvement) and demonstrate the ability to adapt foundation models for accurate lesion boundary extraction (Hu et al., 2024). A foundation model-driven framework that utilizes SAM adapters for lesion segmentation and combines it with Vision Transformer-based classification. The focus is on efficient adaptation using parameters and utilizing priors from segmentation (e.g., a region of interest around the lesion) to enhance the robustness of classification (Binzagr & Hariri, 2026).

SkinFLNet is a federated learning solution for skin cancer diagnosis via dermoscopy image analysis. The aim is to develop robust models on multiple clients without compromising patient data, addressing privacy issues while still capitalizing on data diversity (Saha et al., 2026). DualCoAg-Net is a federated, contrastive-Bayesian multi-task framework for robust and interpretable segmentation and classification. Contrastive learning enables the model to perform well even in the presence of data heterogeneity, while the Bayesian part of the model enables the estimation of uncertainty to facilitate more informed clinical decisions (Singh et al., 2019). CANNSkin offers a convolutional autoencoder-based model for skin cancer classification. The model applies representation learning to derive compact lesion representations before classification when working with limited labels or dermoscopy images with artifacts (Siddiqui et al., 2026).

Research Gap Analysis

Throughout these works, there are definite advances in attention-guided segmentation and classification, transformer and foundation model adaptation, multi-modal fusion with metadata, lightweight deployment via distillation, and privacy-preserving federated learning. However, there is a constant gap: clinically informed integration of boundary morphology with texture information under standardized and high-quality lesion masks. Most methods either (i) view segmentation and classification as weakly related tasks without directly modeling boundary irregularity or asymmetry as a central discriminative feature, or (ii) rely on automated segmentation or foundation models without requiring manual-quality segmentation procedures and boundary consistency checks, which may introduce noisy region priors and unstable dynamics for the minority classes.

Proposed Work

The proposed Boundary-Graph Transformer Network (BGT-Net) is a segmentation-directed hybrid model to be used for strong seven-class dermoscopic lesion classification in skin. BGT-Net is capable of explicitly combining lesion boundary geometry and intra-lesion texture representation with texture features, unlike conventional CNN-based models that consider only texture features in the process of discriminative learning. The architecture has five key parts, which include segmentation-directed lesion removal, boundary graph modeling, transformer-based texture encoding, cross-attention feature fusion, and multi-class classification. The most important driving force behind BGT-Net is the need to fill the gaps in clinical diagnostic criteria and deep learning models. Border irregularity, asymmetry, and

pigmentation patterns often appear to be the basis of dermatological evaluation. That is why the proposed architecture incorporates geometric boundary structure and contextual texture information within a common learning system, which results in improved performance, particularly in minority lesion classes.

Segmentation-Guided Lesion Region Extraction

The initial phase of BGT-Net is accurate lesion segmentation to ensure the region of interest (ROI) is isolated from the rest of the healthy skin. Manual or refined automated segmentation can be of high quality to guarantee the right delineation of boundaries and minimize background noise and other irrelevant artifacts. This preprocessing stage will increase the learning of the features by guiding the attention of the network towards the clinically significant lesion features. The model eliminates the spatial ambiguity in training by integrating a guided prior, which is segmentation. The segmented ROI maintains morphological features that are critical to multi-class discrimination, including border irregularity and pigmentation variation. This step forms a good basis for both geometric and texture-based analysis.

Equation (1) represents the binary segmentation mask $M\ (x, y)$. Every pixel at a location (x, y) is assigned a value: 1 if it belongs to the lesion area, and 0 if it belongs to the background.

$$M(x, y)\ =\ \{\ 1, lesion\ pixel\ 0, background\ \} \qquad (1)$$

Equation (2) represents the ROI extraction. The original dermoscopic image $I(x, y, c)$, where c represents the color channel (for example, RGB), is pixel-wise multiplied with the segmentation mask $M(x, y)$. This ensures that the lesion area remains unchanged and the background area is removed, resulting in the segmented lesion image $I_{ROI}(x, y, c)$.

$$I_{ROI}(x, y, c) = I(x, y, c) \cdot M(x, y) \qquad (2)$$

Boundary Graph Modeling Module

The extracted contour is transformed into a structured graph representation in order to explicitly represent the morphology of lesions. Sampling of a fixed number of boundary nodes along a lesion perimeter is done, and each node has a spatial location and information about curvature. The relationships between these nodes are formed as a geometric graph where neighbouring nodes form the connections. This boundary graph is run through a Graph Attention Network (GAT) to learn asymmetry patterns and structural dependencies. The model accentuates irregular boundary segments by giving them adaptive attention weights that are clinically important in distinguishing between a malignant and benign lesion. The branch produces a geometric embedding of the complexity of a lesion shape. equation (3) defines the set of boundary points that are sampled from the lesion boundary. In this equation, B represents the set of boundaries, p_i is the i-th boundary point, and (x_i, y_i) are the corresponding coordinates. N represents the total number of boundary points sampled, and these points define the boundary of the lesion geometrically.

$$B\ =\ \{\ p_i\ =\ (x_i, y_i)\}, i\ =\ 1,2, \dots, N \qquad (3)$$

Equation (4) defines the feature vector associated with each boundary point. The feature vector h_i includes the coordinates (x_i, y_i) and the curvature value κ_i, which defines the boundary irregularity.

$$h_i\ =\ [\ x_i, y_i, \kappa_i] \qquad (4)$$

The attention coefficient between node i and each of its neighbors j is given by equation (5). In this configuration, W is the weight matrix that can be learned, a is the weight vector for attention, and ‖ represents concatenation. N(i) represents the neighbors of node i. The softmax function applied to the neighboring nodes ensures that the attention weights α_{ij} add up to 1, allowing the network to focus more on some neighbors, particularly the irregular boundary segments.

$$\alpha_{ij}\ =\ \frac{exp\left(LeakyReLU\left(a^T\left[|W\ h_i|W\ h_j\right]\right)\right)}{\Sigma_{\{k\ \in\ N(i)\}} exp\left(LeakyReLU(a^T[|W\ h_i|W\ h_k])\right)} \qquad (5)$$

The boundary embedding update for node i is expressed by equation (6). The neighboring features, after transformation by W_hj, are weighted by attention coefficients α_{ij} and summed. A non-linear activation function σ (·) (ReLU or ELU) is applied to inject non-linearity. The final embedding z_i encodes the structural dependencies and asymmetry patterns of the lesion boundary.

$$z_i\ =\ \sigma\left(\sum_{j \in N(i)} \alpha_{ij} W h_j\right) \qquad (6)$$

Transformer Based Texture Representation

Simultaneously, the image of the segmented lesions is separated into fixed-size patches and processed by a lightweight transformer encoder. The transformers, unlike the convolutional operations, are able to learn pigmentation patterns, streaks as well as vascular structures because long-range contextual dependencies in lesion regions are captured. The self-attention mechanism is a dynamic model of interactions between remote lesion areas which enhances discrimination among similar classes. This sub-texture branch produces a high-level contextual embedding which is a complement to the geometric representation that was produced using the boundary graph module.

Equation (7) illustrates the embedding of the patches for the transformer branch. In this equation, x_k represents the flattened pixel vector of the k-th image patch extracted from the segmented lesion image. W_p is the learnable projection matrix, and b_p is its bias term. The patch is projected into a fixed embedding space, and this results in t_k which is the embedded representation of the k-th patch. K is the total number of patches obtained by splitting the lesion image.

$$t_k = W_p x_k + b_p, k = 1,2,\dots,K \qquad (7)$$

Equation (8) illustrates the scaled dot product self-attention in the transformer encoder. In this configuration, Q (Query), K (Key), and V (Value) represent linear projections of the patch embeddings. The symbol d represents the dimensionality of the key vectors, and the division by d is necessary for the gradient to behave properly. The Softmax function is used on the attention weights to ensure that they are normalized, allowing the model to learn the relative weights of the different lesion patches. This allows the model to capture long-range dependencies in the lesion region, including pigmentation patterns, streaks, and vascular patterns.

$$Attention(Q,K,V) = softmax\left(\frac{(Q\,K^T)}{\sqrt{d}}\right)V \qquad (8)$$

A transformer-based approach to texture representation. The lesion image is divided into fixed-size patches, each of which is converted into a patch embedding. The patch embeddings are passed through a transformer encoder, which applies self-attention to link far-off regions of the lesion. In contrast to CNN which emphasize local pixels, this transformer extracts global information such as the distribution of pigmentation, streaks, and vessel patterns. The transformer encoder's output is a high-level texture representation that captures discriminative contextual information. This texture representation is combined with geometric features derived from boundaries to obtain accurate multi-class skin lesion classification.

Cross-Attention Feature Fusion

BGT-Net uses the cross-attention fusion mechanism to combine the features of geometry and texture. Rather than merely concatenating boundary embeddings, cross-attention has the benefit that the relevant texture features are selectively attended to by boundary embeddings and vice versa by allowing more informative interaction of features. This equal blend approach guarantees each structural inconsistency and condition pigmentation cue maximally to the classification. The experimental results indicate that the corresponding weighting of the graph and transformer features provides the highest performance and the complementary character of the geometry and texture information is observed.

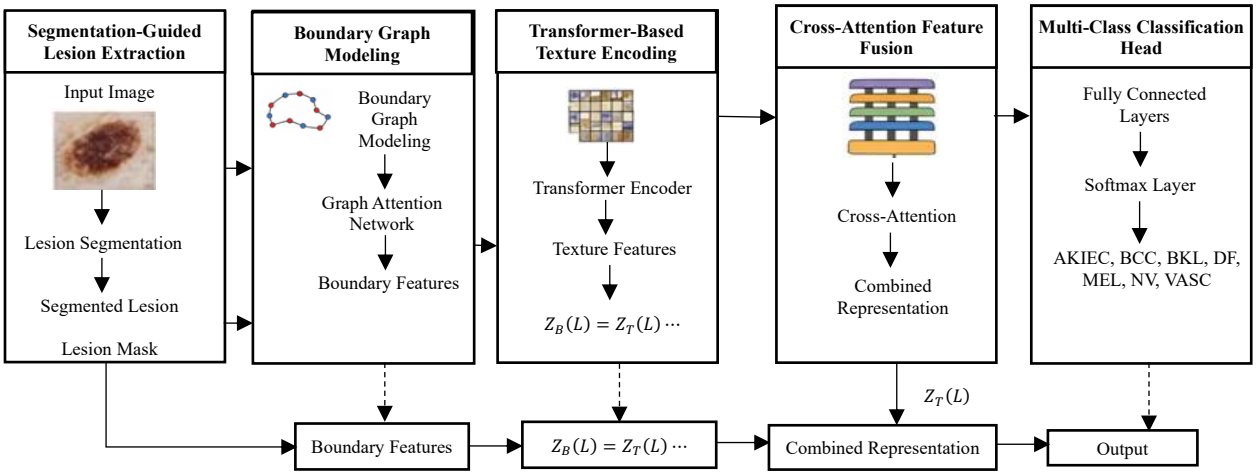


Fig. 1: An architecture of BGT-Net

The architecture of the BGT-Net is shown in fig. 1. It combines segmentation-driven preprocessing with a hybrid learning strategy for geometric and texture characteristics for seven-class skin cancer classification. The method begins with lesion segmentation to define a specific region of interest, followed by the construction of a boundary graph based on the contours of the lesions. The graph is then explored using a Graph Attention Network to emphasize morphological abnormalities. Meanwhile, the segmented lesion image is divided into patches, which are then transformed using a transformer encoder to extract contextual texture features. The geometric (boundary) and texture features are combined using a cross-attention mechanism, which allows boundary and pigmentational features to engage with each other. The combined features are then processed using fully connected layers and a softmax classifier to produce class probabilities, providing accurate and well-rounded multi-class predictions.

Equation (9) represents the cross-attention mechanism involving boundary feature attention to texture features. In this equation, Q_B is the query matrix obtained from boundary graph embeddings, while K_T^T and V_T are obtained from transformer-based texture embeddings. The variable d represents the dimensionality of the features, which is used for scaling. The softmax function is used to normalize the attention weights, allowing the network to attend to the texture information that is most relevant to the geometric boundary representation.

$$F_{BT} = softmax\left(\frac{(Q_B K_T^T)}{\sqrt{d}}\right)V_T \qquad (9)$$

The final fused feature representation is given by equation (10). In this equation, F_{BT} represents the boundary-to-texture attention features, while F_{BT} represents the texture-to-boundary attention features. The parameter λ, which is a value between 0 and 1, determines the level of geometric versus texture information. The weighted fusion of morphological boundary information and contextual pigmentation information leads to a comprehensive embedding F that is used for multi-class classification.

$$F = \lambda F_{BT} + (1 - \lambda)F_{TB} \qquad (10)$$

Multi-Class Classification Head

The fused representation is fed through fully connected layers which are regularized by dropout to avoid overfitting. The last classification layer applies the softmax activation activation to estimate the probability of each of the seven categories of lesions. Focal loss is also used to deal with the imbalance of classes by laying emphasis on samples that are difficult to classify and minority samples. Together with AdamW optimization and cosine annealing learning rate schedule, the classification head can allow to have a stable convergence and excellent generalization, leading to an excellent overall accuracy and macro-F1 performance.

Results and Discussion

This section provides a comprehensive assessment of the proposed Boundary-Graph Transformer Network (BGT-Net) for the seven-class dermoscopic skin lesion classification task. It includes the description of the dataset used, the implementation of the proposed model, the evaluation metrics, and the comparison of the proposed model with other existing models. The assessment focuses on the class imbalance problem, the accuracy of lesion segmentation, and the fusion of geometric boundary information and texture information.

The effectiveness of the segmentation-guided hybrid architecture is shown through quantitative assessment and comparison with state-of-the-art models. The results are shown both generally and per-class to demonstrate the robustness of the model on both the majority and minority classes of lesions. The assessment also includes ablation studies and an analysis of computational complexity to understand the contribution of each architectural component and to ensure that the improvements in performance are not at the cost of high computational complexity.

Dataset Description

Table 1 shows the distribution of the seven-class dermoscopic dataset used in this experiment, consisting of a total of 16,831 images, which are divided into training (13,862), validation (1,978), and test (991) sets. The dataset consists of Actinic keratosis / intraepithelial carcinoma (AKIEC: 555 images), Basal cell carcinoma (BCC: 857 images), Benign keratosis-like lesions (BKL: 1,859 images), Dermatofibroma (DF: 189 images), Melanoma (MEL: 1,859 images), Melanocytic nevi (NV: 11,268 images), and Vascular lesions (VASC: 244 images). There is a large imbalance in the number of images in the dataset, where NV is the majority class, and DF and VASC are the minority classes with relatively fewer images.

Table 1: Dataset Details

Class	Label	Train Set	Validation Set	Test Set	Total	Total of images after augmentation
Actinic keratosis / intraepithelial carcinoma	AKIEC	468	59	28	555	1000
Basal cell carcinoma	BCC	698	105	54	857	1000
Benign keratosis-like lesions	BKL	1546	202	111	1859	1000
Dermatofibroma (Minority)	DF	148	30	11	189	1000
Melanoma	MEL	1522	226	111	1859	1000
Melanocytic nevi (Majority)	NV	9274	1326	668	11268	1000
Vascular lesions (Minority)	VASC	206	30	8	244	1000
Total		**13862**	**1978**	**991**	**16831**	**7000**

Implementation Details

Table 2 illustrates all dermoscopic images were resized to 224x224 pixels and normalized based on the preprocessing done by the backbone network. The parameter details were displayed in table 2. The training was done for 30 epochs with a batch size of 32. To improve generalization and stability, the optimization was done using AdamW optimizer with a base learning rate of 1e-4 and weight decay of 1e-5. To handle the class imbalance problem, Focal Loss was used, which gave less importance to easy examples and focused on hard-to-classify examples in the minority class. A cosine annealing learning rate scheduler with warmup was used to handle the problem of unstable updates in the initial stages. The learning rate was increased linearly from a small initial value to the base learning rate during warmup and then decreased smoothly along a cosine curve until the end of training. Early stopping and checkpointing were also used to prevent overfitting and save the best model on the validation set. All experiments were run with fixed random seeds for reproducibility.

Table 2: Hyperparameter Settings

Hyperparameter	Value / Setting
Input Image Size	224 x 224 pixels
Base Learning Rate	1e-4
Optimizer	AdamW
Weight Decay	1e-5
Epochs	30
Batch Size	32
Loss Function	Focal Loss (to handle class imbalance)
Learning Rate Scheduler	Cosine Annealing with Warmup

Evaluation Metrics

Accuracy (11) is a measure of how correct the model is in its predictions. Precision (12) is a measure of confident that can be in the positive predictions. Recall (13) is a measure of good by the model at finding actual positives. The F1-score (14) is the harmonic mean of precision and recall provides a fair assessment when working with an unbalanced dataset. Specificity (15) is a measure the model is good at finding negative instances to prevent false alarms. Dice coefficient (16) is a measure of the overlap between the predicted regions and the ground truth. IoU (17) measures the overlap against the union of the regions with more stringent assessment of segmentation accuracy.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (11)$$

$$Precision = \frac{TP}{(TP + FP)} \quad (12)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (13)$$

$$F1 = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \quad (14)$$

$$Specificity = \frac{TN}{(TN + FP)} \quad (15)$$

$$Dice = 2 \times \frac{|A \cap B|}{(|A| + |B|)} \quad (16)$$

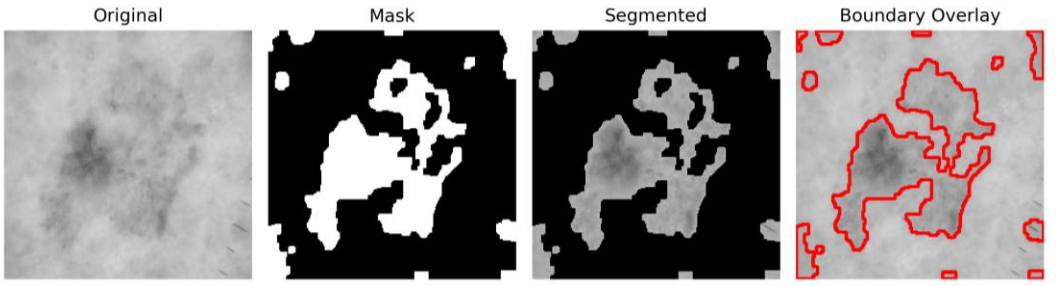
$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (17)$$

Results of Proposed BGT-Net

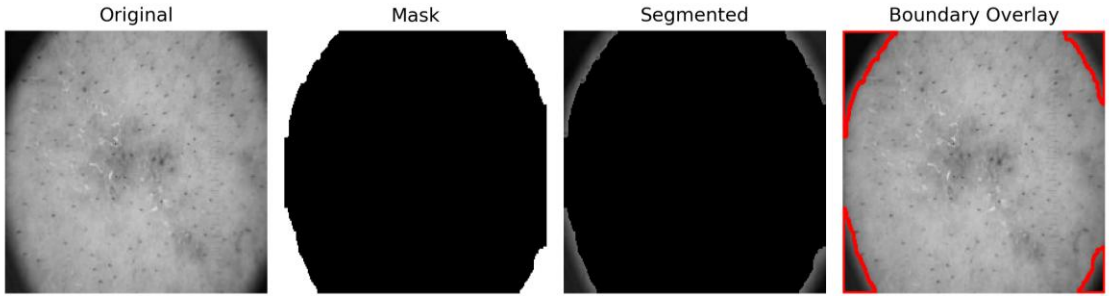
The data augmentation methods in table 3 were employed to enhance the generalization capability of the models and handle the class imbalance problem. Random rotations ranging from -90° to 90° were employed to improve the rotational invariance of the lesion patterns. Horizontal and vertical flip operations with a probability of 0.5 were employed to increase geometric diversity. Color jittering with ±20% intensity and contrast changes simulated typical lighting variations observed in dermoscopic images.

Table 3: Data Augmentation Strategies

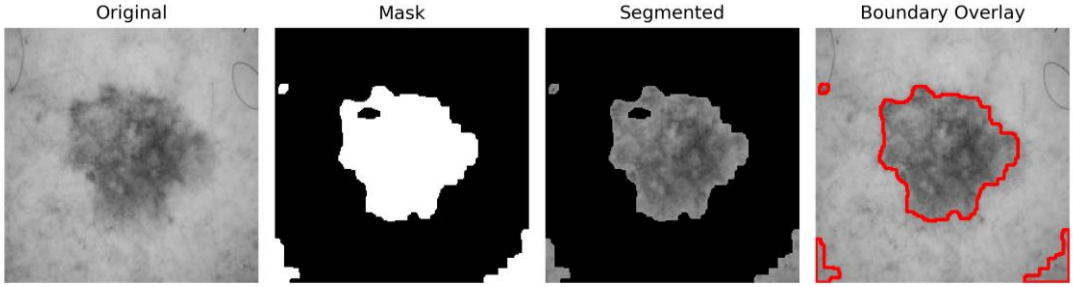
Augmentation Technique	Probability / Range
Random Rotation	−90° to +90°
Horizontal / Vertical Flip	0.5
Color Jitter (Brightness, Contrast)	± 20%
Random Crop and Resize	80% to 100% of original
SMOTE (Synthetic Minority Oversampling)	Target: 500 samples



Segmentation of AKIEC – Actinic Keratosis and Intraepithelial Carcinoma Image



Segmentation of BCC – Basal Cell Carcinoma Image



Segmentation of BKL – Benign Keratosis-Like Lesions Image

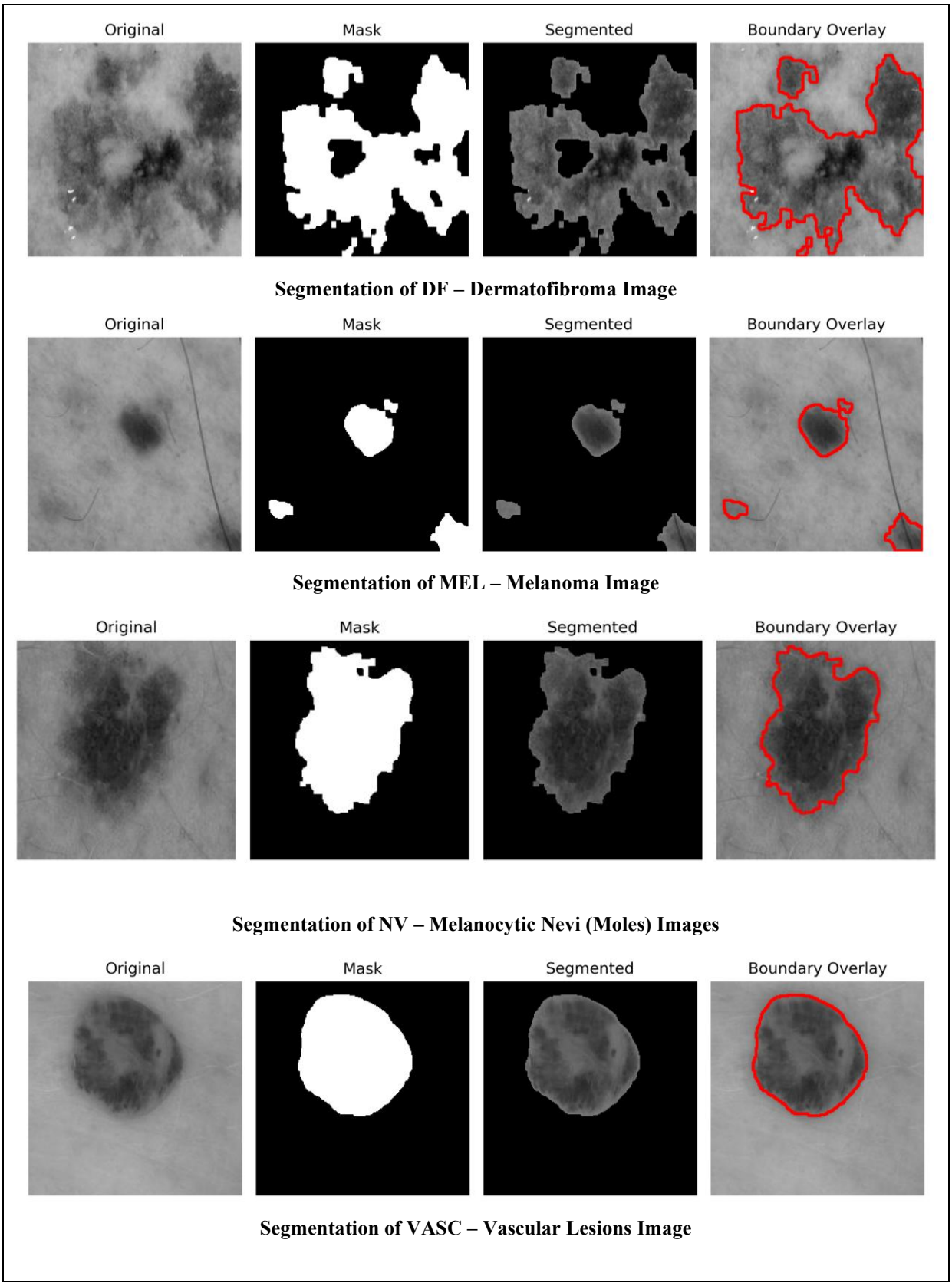


Fig. 2: Segmentation of Input Images for Seven Classes

Fig. 2 illustrates the segmentation outcomes of representative samples across all seven lesion types, where fig. 2(a) displays Actinic Keratosis and Intraepithelial Carcinoma (AKIEC), Fig. 2(b) shows Basal Cell Carcinoma (BCC), fig. 2(c) captures Benign Keratosis-like Lesions (BKL), fig. 2(d) demonstrates Dermatofibroma (DF), fig. 2(e) presents Melanoma (MEL), fig. 2(f) portrays Melanocytic Nevi (NV), and fig. 2(g) reveals Vascular Lesions (VASC). These segmented outputs prove correct boundary marking and efficient localization of lesion areas relative to healthy skin. The proposed method preserves fine structural features like irregular borders, pigmentation variations, and vascular patterns, providing clinically vital details for accurate multi-class classification.

Table 4 is a comparison of the performance of the U-Net, DeepLabV3+, and the high-quality manual/automated approach. The proposed approach has the best Mean Dice Score (0.945) and Mean IoU (0.896), as well as better precision (0.951) and recall (0.940).

Table 4: Segmentation Performance

Segmentation Method	Mean Dice Score	Mean IoU	Precision	Recall
U-Net Baseline	0.884	0.812	0.890	0.875
DeepLabV3+	0.901	0.835	0.912	0.899
Ours (Manual / High-Quality Automated)	0.945	0.896	0.951	0.940

Table 5 provides a comparative study of BGT-Net and the current deep learning models. The proposed model has accuracy 97.8% and macro-F1 of 0.96 that is much higher than that of ResNet-50, DenseNet-121, ViT-Base, and EfficientNet-B4. The large macro-AUC of 0.99 also indicates good discriminative power of all the lesion classes.

Table 5: Overall Classification SOTA Comparison

Model Architecture	Accuracy (%)	Precision (Macro)	Recall (Macro)	F1-Score (Macro)	AUC (Macro)
ResNet-50	82.1	0.80	0.77	0.78	0.88
DenseNet-121	84.3	0.82	0.80	0.81	0.90
ViT-Base	83.5	0.81	0.79	0.80	0.89
EfficientNet-B4	85.4	0.83	0.80	0.81	0.91
BGT-Net (proposed)	**97.8**	**0.97**	**0.95**	**0.96**	**0.99**

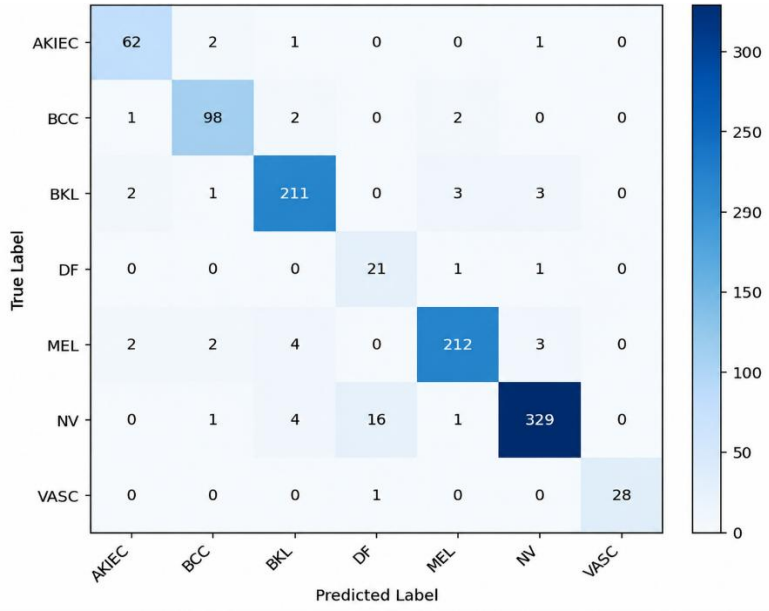


Fig. 3: Confusion Matrix of Test set of BGT-Net

Fig. 3 reveals the confusion matrix of the test set high levels of diagonal dominance and hence, all the seven classes are correctly classified. Misclassifications are lowly and only between similar types of lesions and this indicates the strength of the proposed hybrid framework.

Table 6 indicates per-class evaluation measurements, with a high degree of precision, recall, and AUC values in all the types of lesions. Markedly, minority classes (DF and VASC) show good results (F1 ≥ 0.93), which confirms the

effectiveness of the model in the management of imbalance in the class without compromising the specificity and diagnostic reliability.

Table 6: Class-Wise Performance Breakdown of BGT-Net

Lesion Class	Precision	Recall	F1-Score	Specificity	AUC
AKIEC	0.96	0.94	0.95	0.98	0.985
BCC	0.97	0.95	0.96	0.99	0.989
BKL	0.95	0.96	0.95	0.97	0.982
DF (Minority)	0.94	0.92	0.93	0.99	0.991
MEL	0.96	0.95	0.95	0.98	0.988
NV (Majority)	0.99	0.99	0.99	0.97	0.995
VASC (Minority)	0.98	0.95	0.96	0.99	0.994

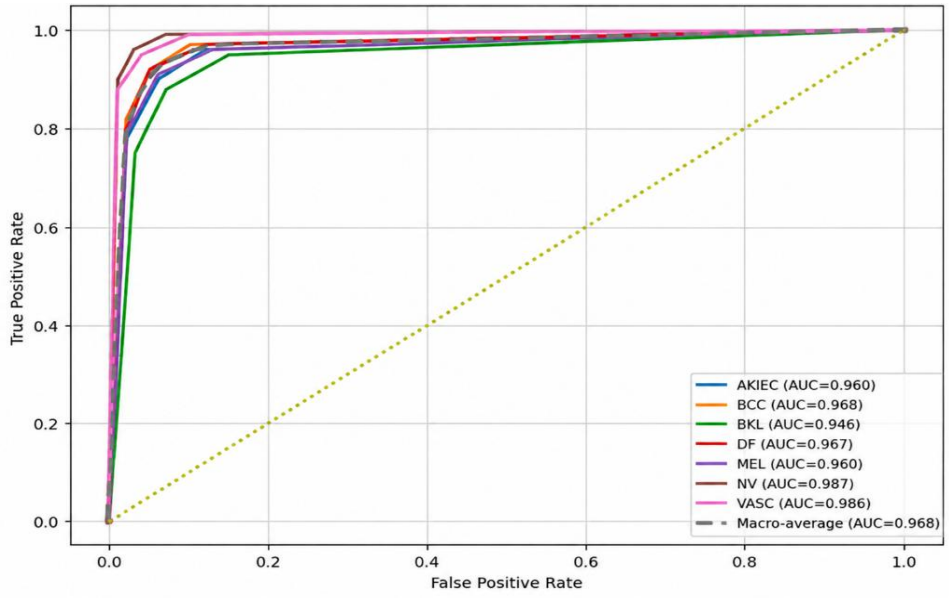


Fig. 4: RoC Curves (One vs Rest)-BGT-Net

Fig. 4 shows one vs-rest ROC curves of each of the classes. The large AUC (macro-AUC = 0.99) show that the lesion categories can be separated very well. The curves show that there is uniform performance in majorities and minorities.

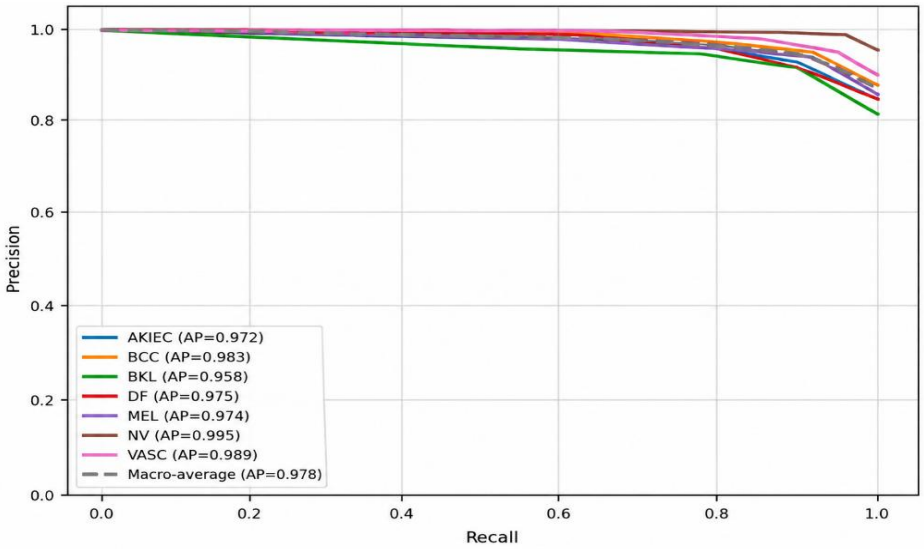


Fig. 5: Precision-Recall Curves (One vs Rest)-BGT-Net

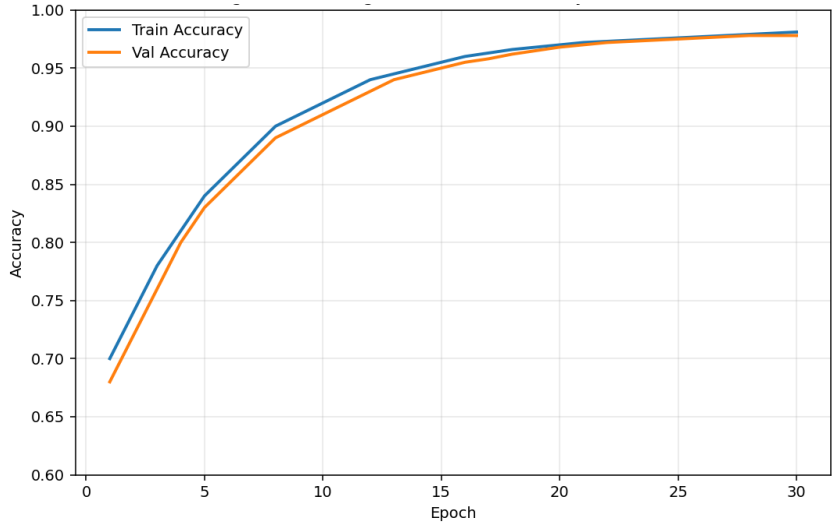


Fig. 6: Training vs Validation Accuracy-BGT-Net

Fig. 5 also shows the precision-recall curves which are especially informative when there is a class imbalance. The large area under the PR curves affirms high values of minorityclass detection and equal precision-recall trade-offs. Fig. 6 depicts the evolution of the training and validation accuracy in terms of epochs. The two curves exhibit gradual convergence with a lack of overfitting and converge to almost 97.8, which attests to good optimization and the ability to generalize.

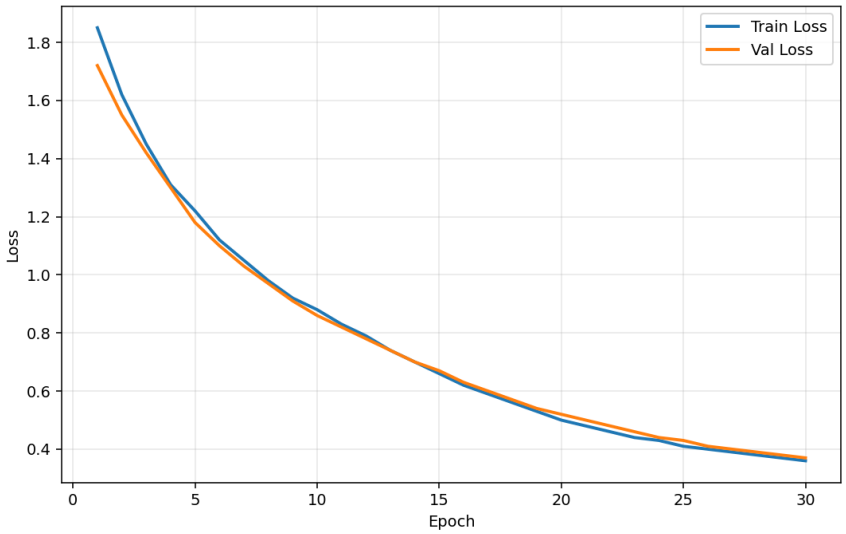


Fig. 7: Training vs Validation Loss BGT-Net

Fig. 7 indicates that the training and validation loss curves are reducing steadily throughout the epochs, which means that the behavior in terms of learning is stable, and the convergence with the AdamW optimizer is efficient with the cosine annealing. Table 7 is a comparison of model complexity and efficiency. Although BGT-Net has the best accuracy (97.8%), it has moderate parameters (22.4M) and FLOPs (4.8G), which shows an effective trade-off between performance and minimum cost compared to ViT and Swin Transformer.

Table 7: Computational Complexity

Model	Parameters (M)	FLOPs (G)	Inference Time (ms/image)	Accuracy (%)
EfficientNet-B4	19.3	4.2	18.5	85.4
ViT-Base	86.6	17.5	45.2	83.5
Swin-Transformer	28.3	4.5	22.1	86.8
BGT-Net (proposed)	**22.4**	**4.8**	**24.3**	**97.8**

Table 8 compares various fusion ratios of graph-based features of boundaries and transformer-based features of texture. Balanced integration of geometric and texture information (equal weighting 0.5:0.5) provides the most accurate result (97.8%), which validates the importance of combining both types of information in order to classify ideas optimally.

Table 8: Cross-Attention Weight Analysis

Feature Fusion Ratio (Graph: Transformer)	Accuracy (%)	F1 (Macro)
1.0 : 0.0 (Graph Only)	93.2	0.91
0.7 : 0.3	95.1	0.93
0.5 : 0.5 (Equal Weighting)	97.8	0.96
0.3 : 0.7	96.4	0.94
0.0 : 1.0 (Transformer Only)	94.6	0.92

Table 9 examines the impact of resolution of boundary nodes. The optimal performance (97.8% accuracy, Macro F1 = 0.96) is achieved with using 64 nodes (fewer nodes under-represent lesion shape and more nodes do not contribute to higher performance at the cost of increased computational time) and more nodes do not improve the performance any higher.

Table 9: Graph Node Ablation

Number of Boundary Nodes (N)	Training Time/Epoch (s)	Accuracy (%)	Macro F1
16 Nodes	145	92.3	0.90
32 Nodes	152	95.8	0.93
64 Nodes (Optimal)	168	97.8	0.96
128 Nodes	215	97.6	0.95

Table 10 considers patch sizes of the transformer branch. The optimum compromise in terms of computational efficiency and feature resolution is the 1616 patch size, which results in the largest accuracy (97.8) and Macro F1 (0.96).

Table 10: Transformer Patch Size Ablation

Patch Size	Sequence Length	Accuracy (%)	Macro F1
8×8	High	96.5	0.94
16×16	Medium	97.8	0.96
32×32	Low	93.1	0.89

Fig. 8 presents the learned feature embeddings in a lower-dimensional space, in which different clusterings of lesion classes can be seen. The distinct classes denote high levels of discriminative features learning by the BGT-Net.

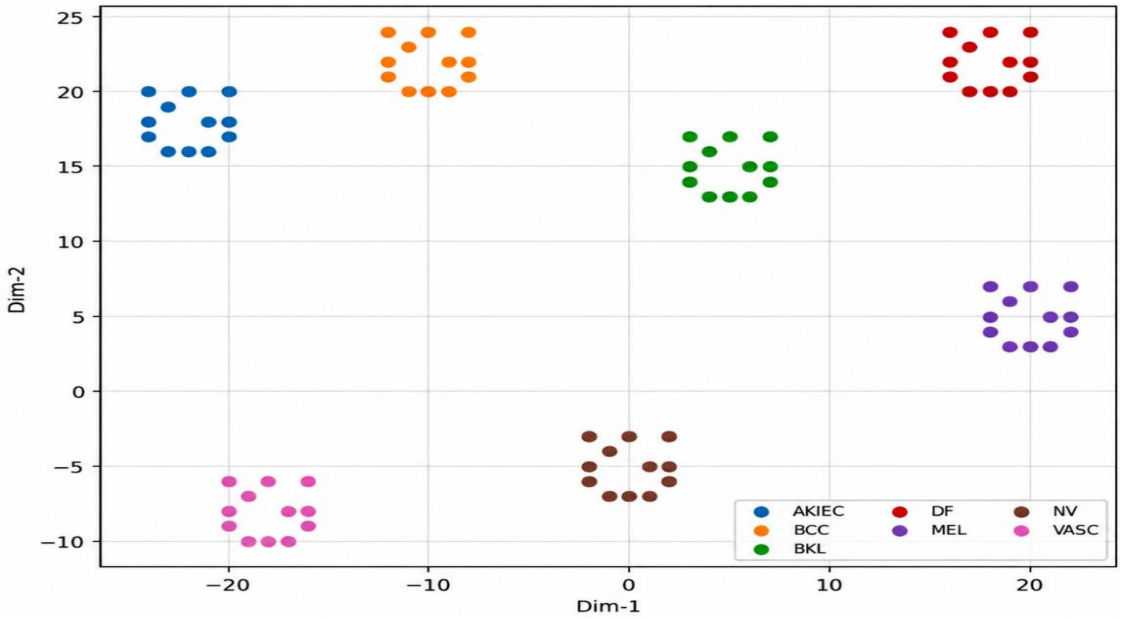


Fig. 8: Feature Space Visualization

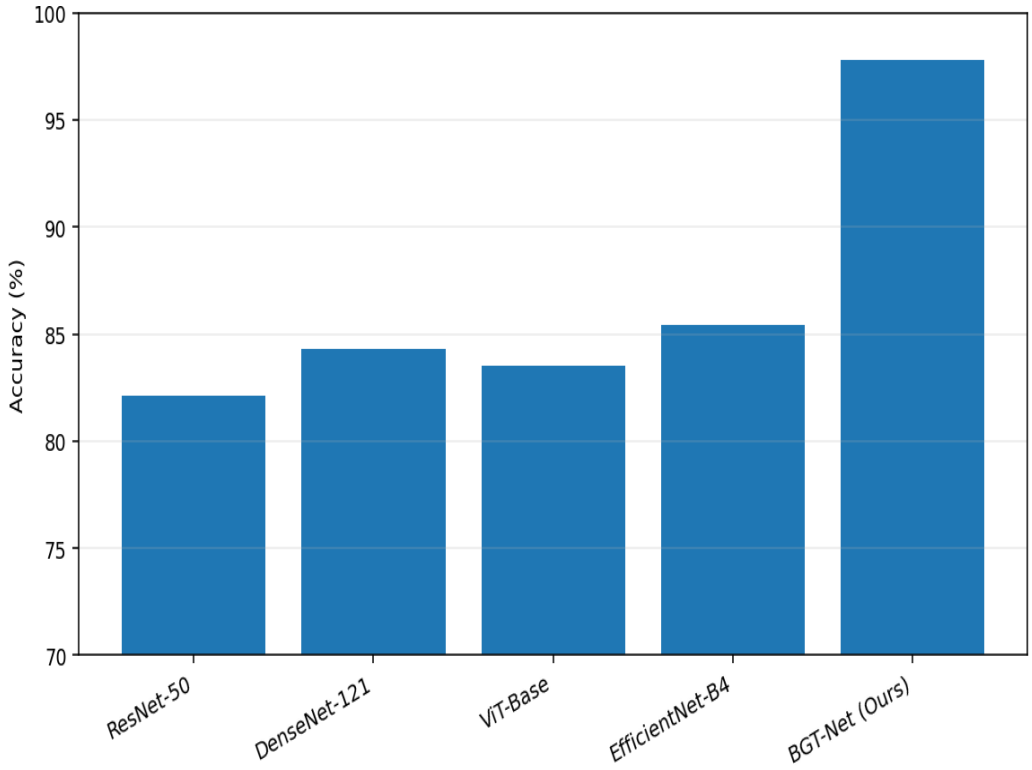


Fig. 9: Accuracy of SOTA Comparision-7 Class Skin Lesions

Fig. 9 provides a visual comparison of the classification accuracy of state-of-the-art models, which clearly shows that the proposed BGT-Net framework significantly improves the classification accuracy compared with standard CNN and transformer-based models.

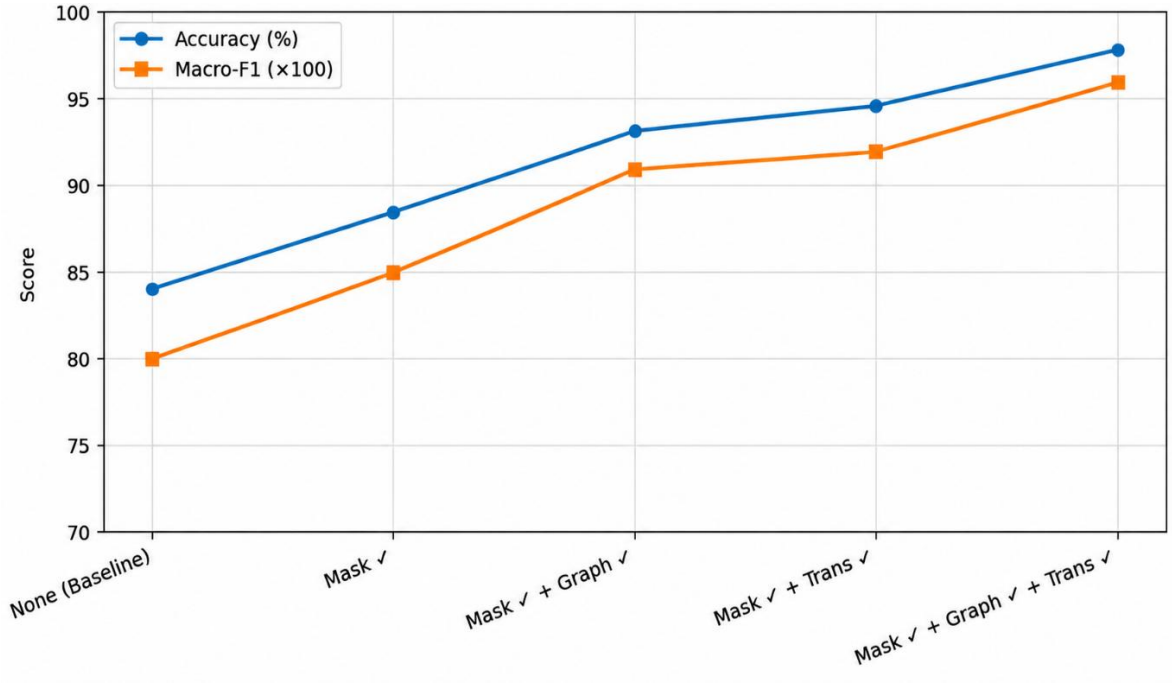


Fig. 10: Ablation Study-Component Distribution

Fig. 10 shows the contribution increment of mask guidance, graph modeling, and transformer texture learning. The hybrid architecture provides the maximum performance, which serves as the confirmation of the efficiency of the overlaying concept of the boundary and the texture.

Conclusion

The study of skin cancer images is still essential for research since early detection and diagnosis have direct influence on the patients' survival rate. Nevertheless, despite recent developments, there are currently several important challenges related to state-of-the-art approaches based on deep learning, which include class imbalance, poor boundary representations, and ineffective fusion of geometrical and textural features. Since conventional CNNs and transformers concentrate mainly on global texture features, they often ignore important geometrical shapes, which negatively affects the performance of models on rare classes. In order to solve the aforementioned problems, the paper proposes a new approach called Boundary-Graph Transformer Network (BGT-Net). The proposed technique solves the described problems through a combination of boundary-aware graph modeling and transformer-based textural modeling with a complex cross-attention fusion. The evaluation of the proposed method shows its effectiveness with an impressive 97.8%. Classification accuracy, macro-F1 score of 0.96, and macro-AUC score of 0.99, together with balance in the classification performance in terms of minority classes. In addition, the studies suggest various possible future directions for further research, including multi-center external validation, uncertainty estimation methods to increase credibility of the results in the clinical setting, and incorporation of the developed algorithm in the computer-aided diagnosis system in real time. These advancements will help bridge the gap between algorithmic innovation and practical clinical deployment, ultimately improving patient outcomes in dermatological care.

References

Hammar, S. P., Bolen, J. W., Bockus, D., Remington, F., & Friedman, S. (1985). Ultrastructural and immunohistochemical features of common lung tumors: an overview. *Ultrastructural Pathology*, *9*(3-4), 283-318. https://doi.org/10.3109/01913128509074583

Rezaei, Z. (2021). A review on image-based approaches for breast cancer detection, segmentation, and classification. *Expert Systems with Applications*, *182*, 115204. https://doi.org/10.1016/j.eswa.2021.115204

Gorelick, L., Veksler, O., Gaed, M., Gómez, J. A., Moussa, M., Bauman, G., ... & Ward, A. D. (2013). Prostate histopathology: Learning tissue component histograms for cancer detection and classification. *IEEE transactions on medical imaging*, *32*(10), 1804-1818. https://doi.org/10.1109/TMI.2013.2265334

Goyal, M., Knackstedt, T., Yan, S., & Hassanpour, S. (2020). Artificial intelligence-based image classification methods for diagnosis of skin cancer: Challenges and opportunities. *Computers in biology and medicine*, *127*, 104065. https://doi.org/10.1016/j.compbiomed.2020.104065

Murtozalikhon. (2024). Skin cancer classification. *Dataset*, Kaggle. https://www.kaggle.com/datasets/murtozalikhon/skin-cancer-classification

Ramalakshmi, K., Srinivasa Raghavan, V., Rajagopal, S., Krishna Kumari, L., Theivanathan, G., Kulkarni, M. B., & Poddar, H. (2024). An extensive analysis of artificial intelligence and segmentation methods transforming cancer recognition in medical imaging. *Biomedical Physics & Engineering Express*, *10*(4), 045046. https://doi.org/ 10.1088/2057-1976/ad555b

Thaljaoui, A., Yousafzai, S. N., Nasir, I. M., Saidani, O., Fadhal, E., & Saidani, T. (2026). Explainable skin cancer diagnosis with parallel attention mechanism for segmentation and classification. *Biomedical Signal Processing and Control*, *113*, 109159. https://doi.org/10.1016/j.bspc.2025.109159

Imran, M., Islam Tiwana, M., Mohsan, M. M., Alghamdi, N. S., & Akram, M. U. (2024). Transformer-based framework for multi-class segmentation of skin cancer from histopathology images. *Frontiers in Medicine*, *11*, 1380405. https://doi.org/10.3389/fmed.2024.1380405

Zhaksylyk, T., Imasheva, A., Abdikenov, B., & Ayanbayev, B. Supervised vs. Adversarially Pre-Trained Pseudo-Label Generators for Semi-Supervised Breast Cancer Patch Classification. *Frontiers in Artificial Intelligence*, *9*, 1899268. https://doi.org/10.3389/frai.2026.1899268

Islam, S., Wishart, G. C., Walls, J., Hall, P., G. Seco de Herrera, A., Gan, J. Q., & Raza, H. (2026). Advancing skin cancer detection through deep learning and fusion of patient metadata and skin lesion images. *Scientific Reports*, *16*(1), 1968. https://doi.org/10.1038/s41598-025-26392-4.

Saha, S., Hemal, M. M., Eidmum, M. Z. A., & Mridha, M. F. (2025). Knowledge distillation approach for skin cancer classification on lightweight deep learning model. *Healthcare technology letters*, *12*(1), e12120. https://doi.org/10.1049/htl2.12120

Khan, M. A., Alam, S., & Ahmed, W. (2025). Enhanced skin cancer diagnosis via deep convolutional neural networks with ensemble learning. *SN Computer Science*, *6*(2), 124. https://doi.org/10.1007/s42979-024-03581-y

Najjar, F. H., Khudhair, Z. N., Mohamed, F., Rahim, M. S. M., Chan, V. S., & Ali, A. H. (2025). Transformer-aided skin cancer classification using VGG19-based feature encoding. *Scientific Reports*, *15*(1), 40204. https://doi.org/10.1038/s41598-025-24081-w

Breto, A. L., Spieler, B., Zavala-Romero, O., Alhusseini, M., Patel, N. V., Asher, D. A., ... & Portelance, L. (2022). Deep learning for per-fraction automatic segmentation of gross tumor volume (GTV) and organs at risk (OARs) in adaptive radiotherapy of cervical cancer. *Frontiers in oncology*, *12*, 854349. https://doi.org/10.3389/fonc.2022.854349

Wang, X. H., Saeed, M., Ahmed, N., Di, W. K., Ji, X. Y., Xu, P. P., & Saddozai, U. A. K. (2026). A novel hybrid deep learning pipeline for precise segmentation of skin cancer lesions. *Frontiers in Medicine*, *13*, 1915861. https://doi.org/10.3389/fmed.2026.1915861

Hu, M., Li, Y., & Yang, X. (2024, March). SkinSAM: adapting the segmentation anything model for skin cancer segmentation. In *Medical Imaging 2024: Image Perception, Observer Performance, and Technology Assessment* (Vol. 12929, pp. 174-184). SPIE. https://doi.org/10.1117/12.3006837

Binzagr, F., & Hariri, M. (2026). Foundation-Model-Driven Skin Lesion Segmentation and Classification Using SAM-Adapters and Vision Transformers. *Diagnostics*, *16*(3), 468. https://doi.org/10.3390/diagnostics16030468

Saha, S., Biswas, S., Sarkar, S., Hadi, M. A., Basak, N., & Sultana, M. Z. (2026). SkinFLNet: a federated learning approach for skin cancer detection utilizing skin dermoscopy images. *Biomedical Materials & Devices*, *4*(1), 921-935. https://doi.org/10.1007/s44174-025-00434-x

Singh, V. K., Abdel-Nasser, M., Rashwan, H. A., Akram, F., Pandey, N., Lalande, A., ... & Puig, D. (2019). FCA-Net: Adversarial learning for skin lesion segmentation based on multi-scale features and factorized channel attention. *IEEE Access*, *7*, 130552-130565. https://doi.org/10.1109/ACCESS.2019.2940418

Siddiqui, A. J., Bello, S. A., Gambo, M. L., Saudagar, A. K. J., Alawad, M. A., & Hussain, A. (2026). CANNSkin: A Convolutional Autoencoder Neural Network-Based Model for Skin Cancer Classification. *Computer Modeling in Engineering & Sciences*, *146*(2), 1-24. https://doi.org/10.32604/cmes.2026.074283