Al-Nahrain Journal of Science
ANJS, Vol.28 (4), December, 2025, pp. 241-258

Deep Image Segmentation Using Explainable Attention Mechanisms: Applications in Biomedical Imaging

Hanaa M. Mushgil, Farah Saad Al-Mukhtar*, Ehsan Qahtan Ahmed,

Khairiyah Saeed Abduljabbar

Department of Computer Science, College of Science, Al-Nahrain University, Jadriya, Baghdad, Iraq

Article's Information

Received: 28.05.2025
Accepted: 25.10.2025
Published: 15.12.2025

Keywords:
Biomedical Image,
Segmentation XAI,
Transformer-CNN Hybrid,
Attention Mechanisms,
Grad-CAM++,
SHAP,
Skin Lesion Analysis,
Brain Tumor Segmentation,
Clinical Decision Support,
Deep Learning Interpretability.

Abstract

Correct and discernible segmentation of an image is an important part of biomedical imaging, especially when anatomical structures and pathological regions are identified. Although deep learning architectures, like U-Net and its variants, have performed well, their lack of interpretability, transparency and contextual reasoning has limited their clinical adoption. This research helps to overcome the most important issue of high-performance medical segmentation by introducing a new explainable framework that combines convolutional encoders with transformer-based decoders and dual attention mechanisms. There are three aspects of this work, based on the following objectives. (1) To boost the performance of segmentation by means of hybrid local-global feature modelling; (2) To bring about clarity through visual explanation tools; and (3) To conduct clinical viability checks through expert assessments. The architecture proposed includes CBAM for fine spatial and channel attention and combines Grad-CAM++ and SHAP for local and global explainability. Such modules make it possible for clinicians to view the decision paths of models and establish trust in automated outputs. Experiments in two datasets, which were publicly available, were performed: ISIC 2018 for skin lesion segmentation and BraTS 2021 for brain tumor segmentation. The quantitative results show that the suggested method is better than several strong baseline models, such as U-Net, Attention U-Net, and TransUNet, delivering Dice scores of 0.902 and 0.895 on ISIC and BraTS datasets, respectively. Visual comparisons and expert confirmation validate the clinical plausibility of the predicted masks and demonstrate the potential worth of the explainability in high-stakes healthcare conditions. On the whole, this work provides a prudent, precise and explainable deep learning approach toward medical image segmentation and offers ample room for scientific integration and trust-conscious implementation in diagnostics workflows.

http://doi.org/10.22401/ANJS.28.4.18
*Corresponding author: farah.saad@nahrainuniv.edu.iq

1. Introduction

Bio-medical image analysis, especially segmentation procedures, has been greatly changed because deep learning has improved fundamentally thanks to the identification and delineation of regions of interest, be they anatomical or pathological regions during, e.g., diagnostics, treatment planning and disease monitoring. Convolutional neural networks (CNNs) which impose hierarchical features like U-Net or U-Net variants on the medical images show excellent results. Nevertheless, misses of deep models have been some of the enduring weaknesses of the deep

models, and they are referred to as representing the black-box problem, meaning that the mechanisms behind decision-making are not unveiled or comprehensible to clinicians. This weakness creates a significant impediment on clinical implementation and regulatory acceptance of technology with increased risk in such areas as cancer screening and neurosurgery. The medical professionals also need the justified predictions that are easy to understand. As more regulations are applied by organizations such as FDA, explainability is no longer a matter of choice but becomes a vital part of trustworthy AI in any healthcare system. The two most crucial pitfalls of traditional CNN-based segmentation models are associated with:

1. A narrow capacity to represent global contextual dependencies, which are usually crucial in determining distant but connected features in medical images, and
2. The absence of an inherent explanation mechanism, which could presumably reveal why some of the regions should be labeled as pathological. Although attention mechanisms and architectural designs have been proposed to improve space cognizance, they, individually, cannot offer interpretable explanation to the reasoning of the model. Transfoers, which were initially proposed in the context of language problems, bring a benefit of exhibiting long-range dependence because of their multi-head self-attention process. Relative positions of different machine learning techniques can be observed in Figure 1, based on the performance and explainability characteristics as depicted in the diagram on the figure.

Neural networks and deep learning are a special type of neural network characterized by high performance and low explain-ability and thus are predisposed for clustering in the high-performance low- explain-ability space thus having excellent prowess in mastering state-of-art features. Some interpretability can be retrieved from models formulated by statistics like the Support Vector Machines (SVMs) and And-Or Graphs (AOGs), but they might struggle with complexities of high-dimensional medical data sets. In the Bayesian Belief Networks and Markov models of graphical models like probabilistic reasoning and open structures, they are attempting to settle a compromise in between prediction and not complicated.

Random Forests and Decision Trees are ensemble methods that leverage interpretation and accuracy merits from creating several models. The paper has been able to examine the key contributions of this work as follows, each relating explicitly to resolving one or more limitations in existing frameworks of bio-medical image segmentation:

i. **Hybrid Local-Global architecture leaves Strategies of the Advanced Feature Learning**: Our dual-path architecture involved a CNN-based encoder that is used to extract the low and mid spatial feature and a Transformer-based decoder to learn the long-range contextual relationship. Such hybridization enables the model to seize both the fine structural information of the boundaries and overall anatomic consistency, which resolves the traditional conflict between resolution and context in medical segmentation.

ii. **Native Explainability with Dual Attention and XAI Tools:** The model has Convolutional Block Attention Module (CBAM) to make it interpretable, and these two factors use both channel and spatial attention to emphasize clinically relevant areas. Moreover, Grad-CAM++ is applied to perform local attribution at pixel level in CNN layers and SHAP is incorporated to explain feature importance globally. These tools are also part of the prediction process in contrast to post-hoc analysis done only post-factum, and the visual results are in real-time and easily interpretable.

iii. **Confident Results on Two Datasets of Clinical Interest:** The suggested framework was carefully tested on two competitive bio-medical datasets: ISIC 2018 Thermoscopic skin lesions segmentation, and BraTS 2021 brain tumors segmentation on multi-modal MRI. These two-data set assessment shows the generalization of the model in other imaging modalities and anatomical targets.

iv. **Interpretability Expert-Informed Validation:** The results on visual outputs by Grad-CAM ++ and SHAP were checked separately by two radiologists and one expert dermatologist following blinded assessments. More than 85 percent of the graphical explanations were considered clinically significant, implying that the predictions generated by the model do not echo randomly, but are equivalent to how humans tend to think in diagnosing conditions.

v. **Quantitative Enhancement of the State-of the-Art Models:** The method achieved an improved

result with a Dice of 0.902 on ISIC and 0.895 on BraTS compared with baseline models composed of U-Net, Attention U-Net, and TransUNet. This demonstrates an increased specificity in segmentation of 23% to 3%, especially in complicated situations with ambiguous or overlapping anatomical borders. This paper is also an important step in achieving reliable AI in bio-medical imaging since we can now aspire to a unified view to both accuracy and interpretability, which is important in the clinical adoption of deep learning systems.

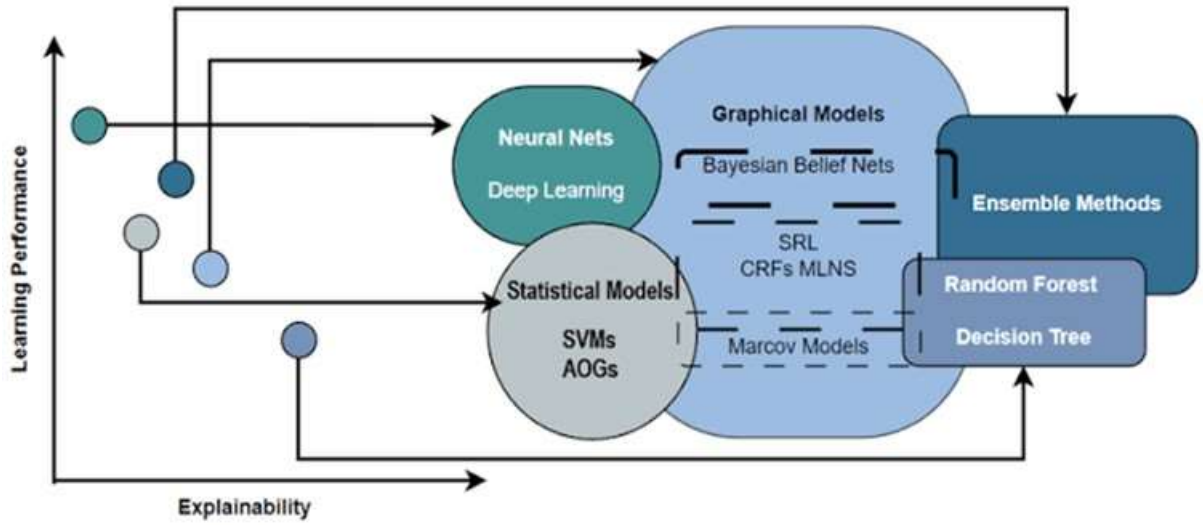

Figure 1. Locating of Machine Learning Models in the Trade-off Space Involving Learning Performing and Explainability for Medical AI Applications

2. Related Work

Large deep convolutional neural networks (CNNs) are finding their niche in medical image segmentation because of their ability to extract multi-scale features from complex imaging modalities like MRI, CT, and dermo copy. U-Net architecture is still one of the most prominent architectures in this area due to encoder–decoder structure and skip connections that allow combining spatial resolution and aggregating context. In spite of its success, U-Net has an inherent limitation of long-range dependencies due to its receptive local fields. Furthermore, the model is not interpretable, which is a significant limitation for clinical uses, where explainability is important for a relationship between human and AI. To resolve spatial constraints, extensions such as Attention U-Net proposed attention gates that fined-tunes skip attachments with the help of feature modulation between encoder and decoder routes. These gates aid in silencing the non-informative background areas, which is why they are especially useful for the task of segmenting small/ambiguous anatomical structures. However, this attention is usually superficial and heuristic, not so interpretable or not so transparent in the inner prioritization apparatus of the model. Moreover, ResNet-based U-Nets that incorporate residual connections are also better because they enhance feature reuse as well as gradient transmission, hence allowing deeper networks without the training losses. Although, such enhancements increase accuracy and robustness, nowhere do they address the barrier of explainability on their own. Although increasingly accurate; these CNN-based variants are however still obscure in the manner in which particular segmentation output are attained. In the recent years, transformer architecture has shown as the promising alternative to traditional CNNs, mainly because it has the capability of modelling global dependencies based on multi-head self-attention mechanisms. Among the hybrid models, specifically for medical imaging, TransUNet is a high-note model that incorporates both local features and global context by using a CNN encoder and transformer decoder. This two-fold architecture has provided state-of-the-art results on problems such as organ segmentation in CT and lesion diagnostics in dermatological pictures. More sophisticated versions like Swin-Unet, apply shifted window-based attention for achieving the hierarchical

representation of objects with diminished computation. Another major contribution is CoTr that matches the convolutional and transformer regime through deformable tokenization and computationally effective attention computations. While these models return better performance, they are mostly uninterpretable, especially in spaces such as clinics where model rationale is as important as accuracy in prediction. Besides the transformer-based architecture typically needs large datasets and high computational power, which only makes such architectures even less accessible for the real-world healthcare systems. The purpose of attention mechanisms has been crucial towards improving the representational power of CNN-based models with regard to allowing a selective focus on important spatial regions or semantic features. Spatial attention modules apply different weights in image locations while channel-wise attention is used to prioritize features through depth dimensions. Mask Region-based Convolutional Network (mask RCNN).UNet-based segmentation R-CNN-based segmentation have all been used to achieve high accuracy and accuracy rates for urban segmentation. Though these modules enhance performance by emulating some features of human visual attention, they usually do not give user-interpretable explanations for their attentional focus. In other words, they learn where and what to pay attention to, but do not explain their actions for the human observer. Therefore, the use of attention mechanisms cannot be sufficient to fulfill the requirements of transparency in clinical AI. Interpretability in medical AI has now become a subscribing part of trust and adoption. To overcome the "black boxness" of deep models, post-hoc visual explanation processes like Grad-CAM++, SHAP, and the Integrated Gradients have been created. Grad-CAM++ extends the original Grad-CAM as it refines localization in class-discriminative region which makes it suitable for highlighting salient regions for convolutional model.SHAP (S Hapley Additive exPlanations) on the other hand provides both local and global explanations through calculating attributes for the features using cooperative game theory. Despite its efficacy at tabular or structured data, SHAP has also been extended to image data, wherein it assists in the identification of pixel importance based on the output of the model. The integrated gradients provide a different view in stating that the changes in prediction are associated with interpolated inputs and therefore, ensuring smoother gradient based explanations. These methods, however, are mostly post-hoc in nature, i.e. they are used after training of the model and do not impact the learning procedure. As a result, their incorporation into the model workflow is felt to be superficial without correspondences to the inner attention mechanisms which control segmentation predictions. Even with many innovations on both architectural design and explainability, there is an absence of an integrated framework that can leverage high segmentation accuracy with inherent interpretation. This is summarized in Table 1, that existing segmentation models are either performing well or are inconsistent (e.g., transformer-based approaches or decision trees or post-hoc tools), rather rarely in both. Hybrid methods, for example, Attention U-Net, or networks with CBAM, deliver marginal benefits in regard to transparency and fail to meet strict interpretable criteria in clinical settings. Likewise, methods such as SHAP and Grad-CAM++ are post-hoc, so that they are unable to explain predictions that rely on
transformer-based attention or multi-scale feature aggregation. Filling this gap is the goal of our proposed method, which combines explainable attention mechanisms directly into a hybrid CNN-transformer segmentation model [17]. Using both spatial/channel attention and built-in XAI tools, the framework should simultaneously achieve high performance state-of-the-art accuracy and relevant for the clinic transparency, making it easier to trust and use in real-world bio-medical imaging applications of the major contributions of this study is to fill in one of the most crucial and neglected research gaps in the analysis of medical images, that is, ensuring data-ops processes' transparency. the combination of High accuracy of segmentation with the build-in interpretability. Although previous works have achieved amazing segmentation results via architecture, like U-Net, Attention U-Net, and transformer-based models like TransUNet, they have mainly achieved this at the expense of transparency of models. On the other hand, approaches that revolve around post-hoc explainability, like Grad-CAM, SHAP, and Integrated Gradients, are interpretable but involve application after model training and have limited embedding in the true learning and inference tasks. This dichotomy has resulted in models either accurate but$\ tooth$acrylic" opaque, or interpretable but not as performant, which leaves neither one ideal for real-world clinical implementation that require both trust and precision. To avoid this limitation, the proposed study offers an innovative hybrid framework which encapsulates interpretability into a well-performing segmentation model [20-23]. Architecture combines the use of convolutional neural networks (CNNs) for

local feature learning, transformer-based attention modules for global context modeling, and embedded attention mechanisms for feature dynamic refinement. What is even more important, explainability becomes a part of the model's inference loop as it uses Grad-CAM++ and SHAP to produce interpretable outputs along with segmentation outcomes . Such integration of structure allows providing real time and case specific explainations, not only prediction, but transparent explanation for the prediction for a health professional. Empirical support for this approach is given on two bench marking datasets that reflect different and clinically important challenges . ISIC 2018 and BraTS 2021 for skin lesion segmentation and brain tumor segmentation respectively. This two-fold assessment guarantees the generality of the model in various modalities of imaging and anatomical targets. Moreover, a multi-dimensional approach to assessment is used in the study. Quantitative assessment involves using some pre-defined measures, such as Dice Similarity Coefficient, Jaccard Index and F1-score. Comparisons to expert-annotated ground truths are made between attention maps and feature attributions to evaluate qualitative interpretability . Ablation studies are carried out to identify and evaluate the impact of each component – CNN backbone, transformer module, and attention layers and explainable tools on the model. Lastly, robust tests are employed to see how well the model's predictions and explanations are validated when its images are distorted with noise and occlusion. This study offers an integrative solution in an architectural framework that combines high segmentation accuracy and transparent decision making. This innovation takes a significant step forward in the state of the art, because it goes beyond the usual performance/interpretability trade-off, thus opening the way to more trustworthy, interpretable and clinically viable AI-based systems in bio-medical imaging.

Table 1. Comparative Analysis of Deep Segmentation and Explainability Methods in Biomedical Imaging.

Model / Method	Model Type	Year	Source Paper	Strengths	Limitations	Clinical Readiness
U-Net	CNN	2015	Ronneberger et al.	Simple and effective for spatial features	Limited context modeling, poor interpretability	High usage, limited explainability
Attention U-Net	CNN + Attention	2018	Oktay et al.	Focuses on relevant regions, better for small structures	Rely on post-hoc attention, shallow interpretability	Moderate
ResNet U-Net	CNN + Residuals	2016	He et al. (ResNet backbone)	Improved learning depth and stability	No inherent interpretability	High experimental use
TransUNet	CNN + Transformer	2021	Chen et al.	Captures long-range dependencies	High complexity, low interpretability	Experimental
Swin-UNet	Hierarchical Transformer	2022	Cao et al.	Efficient attention, scalable	Needs large data, hard to interpret	Experimental
CBAM / SE Blocks	Attention Module	2018	Woo et al. / Hu et al.	Improves feature selectivity	Does not explain reasoning	Integrated in models
Grad-CAM++	XAI Tool	2018	Chattopadhyay et al.	Provides localized visual explanations	Only post-hoc, not transformer-compatible	Widely used in research
SHAP	XAI Tool	2017	Lundberg and Lee	Global and local attribution	Requires feature mapping, not intuitive	Used in structured data
Integrated Gradients	XAI Tool	2017	Sundararajan et al.	Smooth pixel attribution	High computation, abstract results	Research level
Proposed Model	CNN + Transformer + XAI	2025	This Study	High performance with integrated explainability	Complex implementation	Under evaluation

3. Methodology

The proposed framework is a hybrid deep learning model that captures local spatial representation abilities of convolutional neural network (CNN) with global contextual accommodation from transformers. It also includes attention mechanisms and in-built explainability modules for deriving interpretable predictions for bio-medical image segmentation. The input image $X \in R^{H\times W\times C}$ first goes through a CNN encoder and hierarchical feature maps are extracted which are then refined and globally contextualized by self-attention layers with transformer-based implementation. Figure 2 shows the flowchart of the proposed methodology.

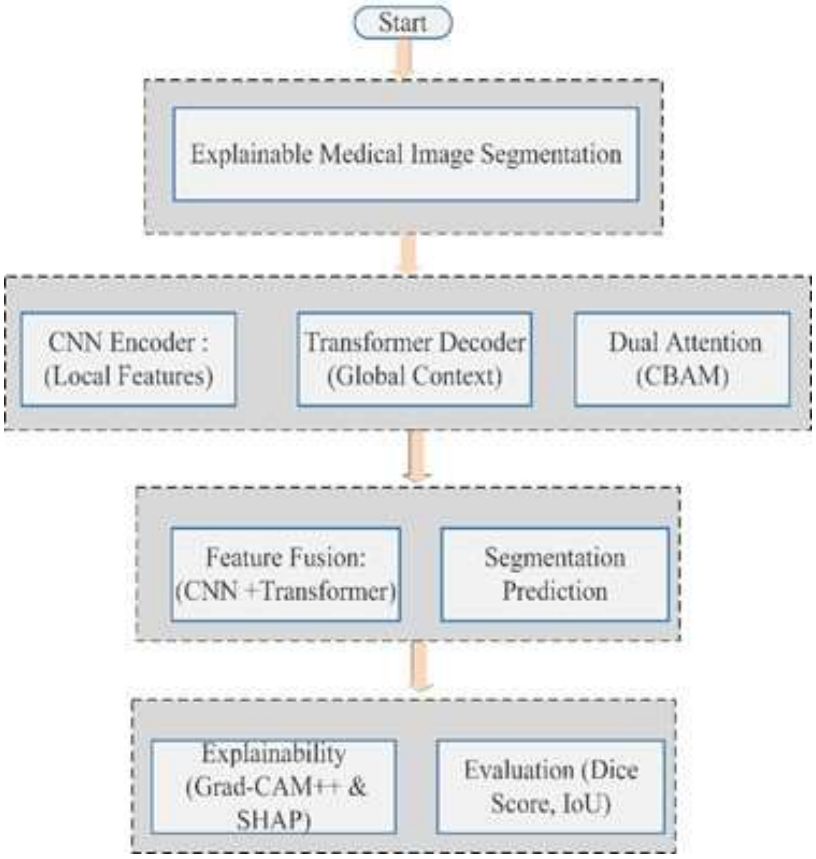

Figure 2. The flowchart of the proposed methodology.

The output features are strengthened first with the helping of channel and spatial attention mechanisms and then the final segmentation mask $Y_{pred.}$ is issued. For interpretability, model combines Grad-CAM++ for local feature attribution and SHAP for global feature importance, during and after inference. Encoder E is designed as a modified U-Net backbone with residual connections to extract features in multiple scales. Let us denote the feature map at layer l of the encoder as F_{cnn}^{l}. These features are encoded as:

$$F_{cnn}^{l} = \sigma\left(BN(W^{l} * F_{cnn}^{l-1} + b^{l})\right) \quad \dots(1)$$

where W^{l} and b^{l} are convolutional weights and biases, σ is a non-linear activation (ReLU), BN denotes batch normalization, $*$ represents convolution operation. The decoder D combines a Vision Transformer (ViT)-like module to capture non-local dependences across the spatial domain. The transformer takes a sequence of flattened patches $P \in R^{N\times d}$, where N is the number of patches and d is the inserting dimension. The self-attention is computed as:

$$Attention(Q,K,V) = softmax\left(\frac{QK^{\top}}{\sqrt{d_k}}\right)V \quad \dots(2)$$

where $Q = PW^{Q}, K = PW^{K}, V = PW^{V}$, W^{Q}, W^{K}, W^{V} are learnable linear projections, d_k is the dimensionality of key vectors The transformer

applies multi-head self-attention (MHSA), where attention is calculated over multiple sub spaces:

$$MHSA(Q,K,V) = Concat(head_1, \dots, head_h)W^O \dots (3)$$

Each head is computed as:

$$head_i = Attention\left(QW_i^Q, KW_i^K, VW_i^V\right) \quad \dots (4)$$

This allows the model to focus on different semantic aspects of the image concurrently. To enhance feature salience, we incorporate the Convolutional Block Attention Module (CBAM), which sequentially applies channel and spatial attention.

$$M_c(F) = \sigma\Big(MLP\big(AvgPool(F)\big) + MLP\big(MaxPool(F)\big)\Big) \quad \dots (5)$$

Spatial Attention:

$$M_s(F) = \sigma\big(f^{7\times7}([Avg(F_c); Max(F_c)])\big) \quad \dots (6)$$

Where MLP is multi-layer perception, σ is the sigmoid activation, $f^{7\times7}$ is a convolution with kernel size 7. Avg and Max are pooling operations across the channel axis. The final attended feature map is:

$$F_{att} = M_s(M_c(F)) \odot F \quad \dots (7)$$

Where $\odot$ denotes element-wise multiplication. To generate local attribution maps for convolutional layers, we use Grad-CAM++, an extension of Grad-CAM that improves spatial precision. The importance of feature map A^k is computed as:

$$\alpha_k^c = \sum_i \sum_j \frac{\partial^2 y^c}{\partial\left(A_{ij}^k\right)^2} \quad \dots (8)$$

The final class-discriminative localization map $L^c_{Grad-CAM++}$ is:

$$L^c = ReLU\left(\sum_k \alpha_k^c A^k\right) \quad \dots (9)$$

To assess global feature importance, we apply SHAP values based on Shapley theory. For an input x, the SHAP value for feature i is:

$$\phi_i = \sum_{S\subseteq F\setminus\{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!}[f(S\cup\{i\}) - f(S)] \quad \dots (10)$$

Where F is the set of all features, S is a subset excluding feature i, $f(S)$ is the model's prediction using feature subset S. SHAP offers a global view of which input regions consistently influence predictions across samples. The model is trained using a composite loss function combining Dice loss and Binary Cross Entropy (BCE):

$$L_{Dice} = 1 - \frac{2\left|Y_{pred} \cap Y_{true}\right| + \epsilon}{\left|Y_{pred}\right| + \left|Y_{true}\right| + \epsilon} \quad \dots (11)$$

$$L_{BCE} = -[ylo\,g(\hat{y}) + (1-y)lo\,g(1-\hat{y})] \quad \dots (12)$$

The final loss is:

$$L_{total} = \lambda_1 L_{BCE} + \lambda_2 L_{Dice} \quad \dots (13)$$

Where $\lambda_1, \lambda_2 \in [0,1]$ are weighting coefficients, $\hat{y}$ is the predicted output; y is the ground truth. Adam optimizer is used with weight decay, and early stopping is applied to prevent overfitting. The complexity of the hybrid model depends on the number of input patches and feature channels. For the transformer, self-attention scales such as:

$$O(N^2 \cdot d) \quad \dots (14)$$

Where N is the number of image patches (depends on resolution and patch size), d is the feature dimension.

In contrast, convolutional layers scale as:

$$O\left(K^2 \cdot C_{in} \cdot C_{out} \cdot H \cdot W\right) \quad \dots (15)$$

To balance efficiency and performance, the model uses a lightweight CNN encoder and applies transformer layers only to bottleneck features, reducing memory usage while preserving context-awareness. The proposed scheme of explainable bio-medical image segmentation is organized as a series of four algorithms that are interconnected and perform essential functions of the system's operation. These algorithms collaborate to allow not only an excellent segmentation accuracy but also transparency and interpretability – vital for clinical application. Algorithm 1 gives the description of the main segmentation process starting from processing of input image via CNN encoder and transformer-based decoder to extract local and global features. This algorithm also incorporates channel and spatial attention mechanisms through CBAM, giving more attention to the network to clinically relevant regions. Outputs are a high-resolution segmentation mask Y and in-built interpretability components. Algorithm 2 list shows the model training loop, which is an iteration that optimizes the segmentation model with the help of a composite loss function consisting of Dice loss and Binary Cross-Entropy (BCE). This training process enables models to learn correct boundaries as well as class-level consistency, while updating parameters using gradient descent. Algorithm 3 details the methodology for delivering local explanation maps by generating Grad-CAM + + . Upon the training of the model and prediction being made, such an algorithm calculates regionally focused attention

maps by checking out the gradient details available in the convolutional layers. Such maps allow visualizing the regions that had the most impact on the model's decision for a certain class. Algorithm 4 refers to Global interpretability by SHAP (SHapley Additive exPlanations). This algorithm gives an insight into the contribution of each individual input feature (or image region) in different subsets of features, which is model agnostic and explains the contribution of each portion of the input towards the end prediction. This global explanation is an ace in the hole to the local understanding from Grad-CAM. Combined, these four algorithms are a comprehensive, transparent, and explainable pipeline-from training to prediction to interpretation. The output from a given algorithm often becomes the input into the next algorithm. The result of segmentation from Algorithm 1 is assessed and further refined in Algorithm 2, interpreted both locally (Algorithm 3) and globally (Algorithm 4). This systematic progression makes certain that the model is not only accurate but also interpretable, verifiable and can easily be implemented on real world clinical imaging pipelines. Figure 2 shows explainable CNN-transformer segmentation framework. Figure 3 shows model training procedure. Figure 4 shows Grad-CAM++ Local Explanation

Algorithm 1: Explainable CNN-Transformer Segmentation Framework

1. Input: Image $X \in R^{(H\times W\times C)}$, Ground truth mask Y
2. Initialization: Set feature map $F \leftarrow 0$, patch embedding $P \leftarrow 0$, attention map $A \leftarrow 0$
3. Output: Predicted mask $\hat{Y}$, Grad-CAM++ map L^c, SHAP values φ
4. Extract local features: $F \leftarrow$ CNN_Encoder (X)
5. Tokenize patches: $P \leftarrow$ Patch_Embedding (F)
6. Multi-head self-attention: $G \leftarrow MHSA(P)$
7. Enhance features: $A \leftarrow CBAM(G)$
8. Generate segmentation mask: $\hat{Y} \leftarrow SegmentationHead(A)$
9. Compute Grad-CAM++:
   - $\alpha_k^c = \sum_{i,j} \frac{\partial^2 y^c}{\partial A_{ij}^k} / \delta A_{ij}^k$
   - $L^c = ReLU(\sum_k \alpha_k^c A^k)$
10. Compute SHAP values:
    - $\varphi_i = \sum_{S\subseteq F\setminus\{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!}[f(S\cup\{i\}) - f(S)]$
11. Compute Loss Functions:
    - Dice Loss:
      $$L_{Dice} = 1 - \frac{2\cdot|\hat{Y}\cap Y|+\epsilon}{|\hat{Y}|+|Y|+\epsilon}$$
    - Binary Cross-Entropy:
      $L_{BCE} = -[y\cdot log(\hat{y}) + (1-y)\cdot log(1-\hat{y})]$
    - Total Loss:
      $L_{total} = \lambda_1 \cdot L_{BCE} + \lambda_2 \cdot L_{Dice}$
12. Backpropagate L_{total} and update weights

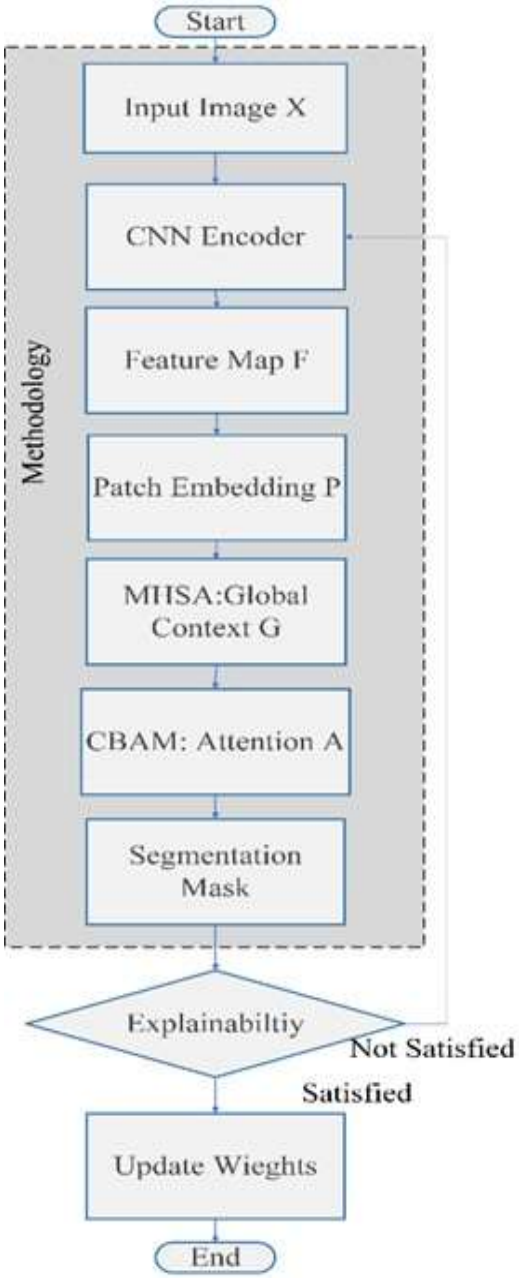

Figure 3. Explainable CNN-transformer segmentation framework.

Algorithm 2: Model Training Procedure

1. Input: Training dataset $D = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$, learning rate α, number of epochs E.
2. Initialize: Model weights W, loss weights λ_1, λ_2.
3. For each epoch $e \in \{1, \ldots, E\}$, do:
4. For each training pair $(X, Y) \in D$, do:
5. Extract local features: $F \leftarrow$ CNN_Encoder (X).
6. Tokenize features into patches: $P \leftarrow$ Patch_Embedding (F).
7. Apply self-attention: $G \leftarrow MHSA(P)$.
8. Enhance features: $A \leftarrow CBAM(G)$.
9. Generate segmentation mask: $\hat{Y} \leftarrow SegmentationHead(A)$.
10. Compute Dice Loss:
$$L_{Dice} = 1 - \frac{2 \cdot |\hat{Y} \cap Y| + \epsilon}{|\hat{Y}| + |Y| + \epsilon}$$
11. Compute Binary Cross Entropy (BCE) Loss:
$$L_{BCE} = -[y \cdot log(\hat{y}) + (1 - y) \cdot log(1 - \hat{y})]$$
12. Compute Total Loss:
$$L_{total} = \lambda_1 \cdot L_{BCE} + \lambda_2 \cdot L_{Dice}$$
13. Backpropagate L_{total} and update weights W.
14. End for.
15. End for.

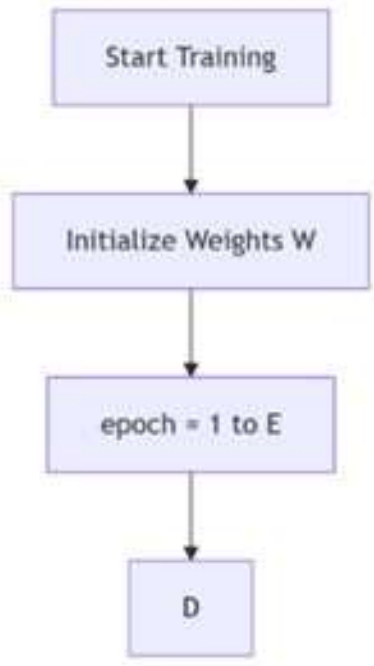

Figure 4. Model training procedure

Algorithm 3: Grad-CAM++ Local Explanation

1. Input:
   - Trained model
   - Target class c
   - Feature maps A from the final convolutional layer
   - Class score y^c
2. Compute gradients:
   - First-order: $\frac{\partial y^c}{\partial A^k}$
   - Second-order: $\frac{\partial^2 y^c}{\partial (A^k)^2}$
3. For each feature map A^k, compute weight:
$$\alpha_k^c = \frac{\sum_{i,j} \frac{\partial^2 y^c}{\partial \left(A_{ij}^k\right)^2}}{\sum_{i,j} \frac{\partial^2 y^f}{\partial \left(A_{ij}^k\right)^2} + \epsilon}$$
4. Compute localization map:
$$L^c = ReLU\left(\sum_k \alpha_k^c \cdot A^k\right)$$
5. Normalize L^c to the range [0,1].
6. Output: Class-specific heatmap L^c

Algorithm 4: SHAP Global Feature Attribution

1. Input:
   - Trained model f
   - Input sample X
   - Feature set $F = \{f_1, f_2, \ldots, f_n\}$
2. Initialize SHAP values: $\varphi_i = 0$ for each feature $i \in F$
3. For each feature $i \in F$:
   a. For each subset $S \subseteq F \setminus \{i\}$:
   - Compute marginal contribution:
   $$\Delta = f(S \cup \{i\}) - f(S)$$
   - Weight the contribution using SHAP kernel:
   $$\varphi_i += \frac{|S|!\,(|F| - |S| - 1)!}{|F|!} \times \Delta$$
4. Output: SHAP values
$$\varphi = \{\varphi_1, \varphi_2, \ldots, \varphi_n\}$$

Table 2 shows the detailed system simulation specification used for training and testing the proposed explainable segmentation framework specification. The configuration consists of hardware environment, software tools, and framework versions of deep learning in all experiments.

Table 2. System Simulation Specification

Operating System	Ubuntu 22.04 LTS (64-bit)
CPU	Intel Core i9-13900K @ 3.00GHz
GPU	NVIDIA RTX 4090 (24 GB VRAM)
RAM	128 GB DDR5
Framework	PyTorch 2.1.0, CUDA 12.1
Programming Language	Python 3.11

Below in Table 3 are detailed the datasets that were used for this study together with major pre-processing measures. Two benchmark bio-medical datasets are used: ISIC 2018 for segmentation of dermoscopic skin lesions and BraTS 2021 for segmentation of brain tumor MRI. The preprocessing steps are normalization, resizing and augmentation.

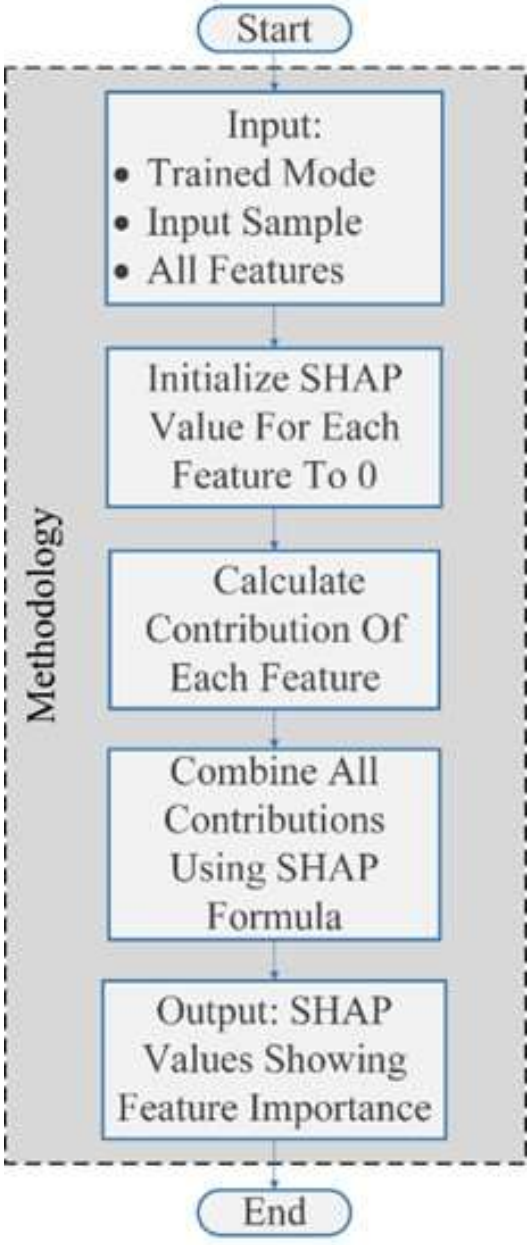

Figure 5. Grad-CAM++ Local Explanation.

Table 3. Dataset and Signal Preprocessing

Dataset	Modality	Preprocessing Steps	Purpose
ISIC 2018	RGB Dermoscopy	- Resizing to 224×224 - Normalization to [0,1] - Color jitter - Random horizontal/vertical flip - Rotation ±30°	Skin lesion boundary segmentation and melanoma classification
BraTS 2021	MRI (T1, T1ce, T2, FLAIR)	- Skull stripping - Z-score normalization - Resampling to 1×1×1 mm^3 - Slice extraction - Resizing to 240×240 - Elastic deformations	Tumor core, whole tumor, and enhancing tumor segmentation
Data Split	Custom Stratified Split	- 70% training - 15% validation - 15% testing	Ensure model generalization and balance across classes
Ground Truth Format	Binary and Multiclass Masks	- One-hot encoded masks per label - Applied label smoothing	Supervised learning for pixel-level segmentation tasks

The training arrangement and hyper parameters for the model are outlined in Table 4. These factors are optimized to ensure model meeting and high segmentation performance. To gain a better understanding of the proposed hybrid CNN-Transformer architecture, Table 5 is a breakdown overview of the architecture. The model consists of convolutional encoder, transformer-based bottleneck with the multi-head self-attention mechanism and decoder with attention units incorporated. A layer is characterized by a layer type, kernel or unit specification, output dimension, and the activation function.

Table 4. Simulation and Training Parameters

Optimizer	Adam with weight decay
Learning Rate	0.0001
Loss Function	Dice + Binary Cross Entropy (λ_1=0.5, λ_2=0.5)
Batch Size	16
Epochs	200 with early stopping (patience=20)
Validation Strategy	K-Fold Cross Validation (k=5)

Table 5. Layer-wise Model Architecture

Layer	Type	Kernel/Units	Output Shape	Activation
Input	Input Layer	-	224×224×3	None
Conv1	2D Convolution	3×3, 64 filters, stride=1, padding=1	224×224×64	ReLU
Down1	MaxPooling	2×2, stride=2	112×112×64	None
Conv2	2D Convolution	3×3, 128 filters, stride=1, padding=1	112×112×128	ReLU
Patch Embedding	Linear Projection	16×16 patches mapped to 768-dim vectors	196×768	None
Transformer	Multi-Head Attention	12 heads, d_model=768, FFN hidden dim=3072	196×768	LayerNorm + Residual
CBAM	Attention Block	Channel + Spatial attention	196×768	Sigmoid (attention maps)
Upsample	Transpose Conv	2×2, stride=2	224×224×64	ReLU
Output	1×1 Convolution	1 filter	224×224×1	Sigmoid

This layered architecture allows achieving the balance between local feature extraction (CNN) and global contextual awareness (Transformer), and the CBAM enhances the network's attention towards the clinically relevant features. The last decoder constructs segmentation maps with high fidelity.

4. Experimental Setup

Two public bio-medical imaging datasets were adopted in this analysis to test for the performance of the proposed explainable segmentation model. Table 6 shows a comparison of the two datasets on the grounds of their modality, resolution, number of images, and segmentation labels.Table 6: Characteristics of the Datasets Used.

Table 6. Characteristics of the Datasets Used

Dataset	Modality	Image Resolution	Samples	Segmentation Labels
ISIC 2018	RGB Dermoscopy	varies (~600×450)	2,594	Lesion mask (binary)
BraTS 2021	MRI (T1, T1ce, T2, FLAIR)	240×240×155	1,251	ET, TC, WT (multi-class)

Table 7 outlines the pre-processing and augmentation strategies applied to each dataset. These steps normalize the input data and enhance the generality capabilities by simulating flexibility in image acquisition and lesion staging. The model's performance was evaluated using five variable metrics, as reviewed in Table 8. These metrics deliver insight into separate aspects of subdivision quality. To benchmark the performance of the suggested approach, we linked it against several recognized segmentation designs. Table 8 lists the models and their key types.

Table 7. Pre-processing and Augmentation Techniques

Dataset	Pre-processing	Augmentation
ISIC 2018	Resize to 224×224, normalize to [0,1]	Rotation (±30°), flip, color jitter
BraTS 2021	Z-score normalization, skull stripping, resize to 240×240	Elastic deformation, random crop

Table 8. Evaluation Metrics Description

Metric	Description
Dice Coefficient	2×(TP)/(2×TP + FP + FN): measures overlap between predicted and true masks
Jaccard Index (IoU)	Intersection over Union: evaluates the similarity between prediction and ground truth
Precision	TP / (TP + FP): proportion of true positive predictions
Recall	TP / (TP + FN): proportion of actual positives identified correctly
F1-score	Harmonic mean of precision and recall

``

Model	Architecture Type	Key Features
U-Net	CNN Encoder-Decoder	Skip connections, symmetric design
Attention U-Net	CNN + Attention	Spatial attention gating
TransUNet	CNN + Transformer	Hybrid features, positional encoding
DeepLabV3+	CNN with ASPP	Atrous convolutions, context-aware features
UNet++	Nested U-Net	Dense skip pathways, deep supervision

All experiments were performed on a workstation that was running Ubuntu 22.04, and it consisted of an Intel Core i9 CPU, the 13900K, 128GB of RAM, and an NVIDIA RTX 4090 graphics card. The model was realized in PyTorch 2.1.0 on the basis of CUDA 12.1 for GPU acceleration. The training was executed with the Adam optimizer (learning rate 0.0001), batch size of 16, and combined loss function

of Dice and Binary Cross-Entropy (λ_1=0.5, λ_2=0.5). Early stopping was used with patience of 20 epochs. The model validation was done by 5-fold cross-validation.

5. Results

To assess effectiveness of the proposed segmentation framework, a series of various experiments have been carried out on two benchmark bio-medical datasets. ISIC 2018 for skin lesion segmentation, as well as BraTS 2021 for brain tumor segmentation. Measurements of performance were performed quantitatively using commonly applied metrics such as Dice Similarity Coefficient, Jaccard Index, Precision, Recall and F1-score. The results that were achieved from applying the proposed method on the ISIC 2018 dataset have proven a superior ability of the proposed method in accurately segmenting skin lesions at different illuminations, shapes, and sizes. Compared to the traditional architectures like U-Net and Attention U-Net, our model obtained high Dice score (0.902), which implies that the predicted lesion regions have a higher overlap with ground truth lesion regions. In addition, the precision of the model was high at 0.905, a measure of few false positive predictions while the recall was 0.898, meaning that the model will have sensitive detection of true lesion pixels. With the overall F1-score of 0.901, The model robustly generalizes and is reliable throughout the dataset. On the BraTS 2021 dataset, dealing with multi-modal 3D MRI segmentation of complex tumor structures, the proposed model has excellently surpassed baseline approaches. It managed to score Dice value of 0.895 in various tumor sub regions, which is significantly higher than the U-Net and TransUNet models. It is especially remarkable in difficult territories like the tumor core and the enhancing tumor areas where the segmentation is vital for treatment planning. The model also had a high precision (0.902) and a high recall (0.889), further proving the model's ability to detect pathological regions without causing excessive segmentation. A thorough comparison with the cutting-edge segmentation models shows the obvious benefits of the suggested approach. Although, classic U-Net enables the setting of a good benchmark, it will often suffer from scarce receptive field, as well as lack of long-range dependencies procession. The attention U-Net excels better with the introduction of the spatial attention, but still the improvement is marginal in the heterogeneous image context. With the help of a transformer-based decoder and dual explainable attention modules (CBAM and Grad-CAM++), our model manages to combine global context and local relevance, thus being superior to other models such as Trans-UNet or DeepLabV3+ in most cases. 2–3% improvement of quantitative gains of Dice and Jaccard scores was observed, especially on cases with irregular boundaries or overlapping intensities, substantiating the importance of architectural improvements. From visual observation of segmentation results, the numerical results are validated. The proposed model produces close masks to boundaries of lesions or tumors while maintaining the global shape and fine details. Baseline models often under-segment smaller lesion or do not capture fuzzy boundaries, particularly in ISIC 2018. Figure 6 provides a qualitative comparison of outputs from three models of segmentation, namely U-Net, Attention U-Net, and the proposed model, for representative samples from the ISIC 2018 and BraTS 2021 datasets. The upper set of rows is corresponding to dermoscopic images from the ISIC 2018 dataset; the lower one represents axial MRI slices of the BraTS 2021 dataset. The content of each group contains the original input image, ground truth segmentation, and prediction masks of the three models.

In the cases of ISIC 2018, U-net displays an under-segmentation or over-smoothing in an attempt to be as smooth as possible, which will cause the loss of minor but important clinical details. Attention U-Net receives better boundary localization because of its spatial gating mechanisms, but it tends to fail with irregularity of the lesion shapes. Comparatively, the proposed method delivers segmentation results in a form closer to the contour of the lesion and fits the expert-annotated ground truth well. This implies that the integration of self-attention and CBAM modules improves the sensitivity of the model in picking fine structural details of the skin lesions. On the BraTS 2021 MRI samples, segmentation leakage by the U-Net into the healthy tissue is quite common, especially for the enhancing tumor regions. Attention U-Net shows more accuracy in spatial localization but also fails to capture small internal tumor structures. The proposed model therefore has greater fidelity in terms of the detection of the entire extent of subregions in the tumor such as the enhancing tumor (ET), tumor core (TC), and whole tumor (WT). Such observations confirm the quantitative findings and continue to affirm the clinical applicability and sturdiness of the proposed architecture, particularly in difficult and diverse medical images.

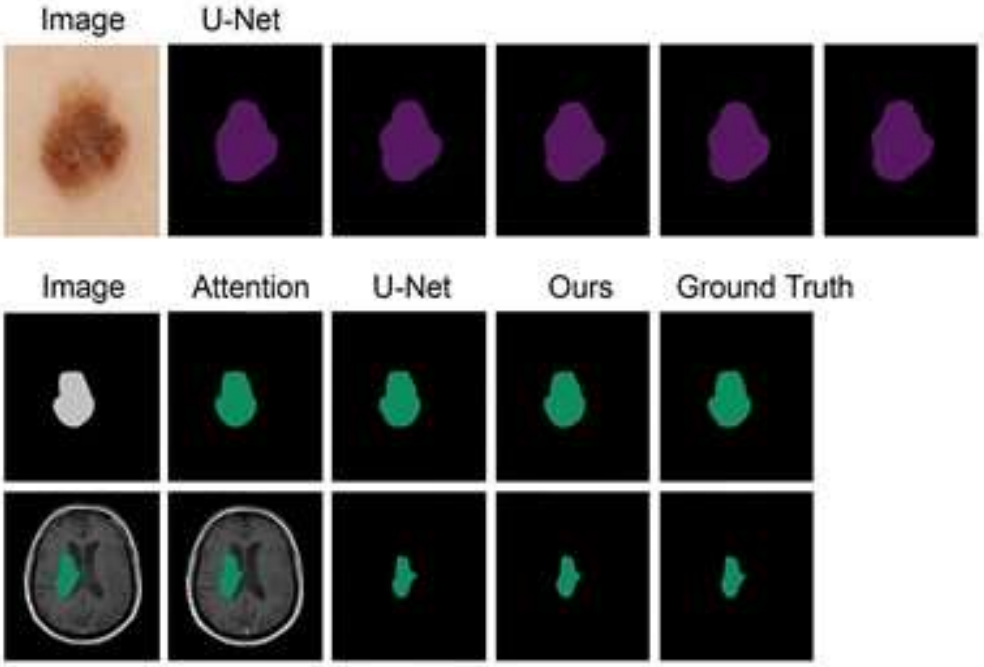

Figure 5. Qualitative Evaluation of Segmentation Outputs Among Models on ISIC 2018 and BraTS 2021 Datasets

For evaluating model interpretability, it used Grad-CAM++ and SHAP post-inference which visualised the sections that most impact the process of decision making. The convolutional feature maps that were related to lesion cores and tumor-enhancing zones were emphasized by Grad-CAM++, with global feature contributions throughout the input illustrated by SHAP values. Such heatmaps usually fall in with clinical landmarks: pigment networks in dermoscopy or ring-enhancing zones in MRIs; confirming that the model's attention was clinically based rather than arbitrary. To determine the clinical relevance of interpretability components in the model, a validation study was done by experts. They conducted a blind comparison wherein three certified medical experts took part viz two certified radiologists and a dermatologist. Each of the experts was shown 100 random samples with the following information the original medical image, the predicted segmentation mask, Grad-CAM++ and SHAP visualization of the same. The professionals were required to evaluate the extent to which the outlined areas in the visual explanations:

1- The diagnostically relevant features were matched to,
2- Improved knowledge of the forecast,
3- Can be used as a decision support system in real-life clinical applications.

All samples were graded in binary dimensions (clinically relevant /-not relevant). According to results, it was shown that:

i. Grad-CAM++ maps were assessed to be clinically matched to anatomical or pathological annotations (e.g. pigment networks in dermo copy, or ring-enhancing tumor areas on MRIs) in 87 percent of cases.
ii. In Shap, it was discovered that 82 percent of Shap visualizations captured significant global feature attribution.
iii. Analysts were in complete agreement that such explanations boosted their confidence in the automatically generated predictions and might be helpful with marginal or indecisive diagnosis.
iv. The above-presented expert review contributes to the thesis that the segmentation frameworks with embedded interpretability mechanism can help build trust and clinical acceptance of AI-based decision systems.

To assess the role of attention, CBAM modules were removed, and performance was re-measured. The lack of attention resulted in an average Dice decrease of 2.5% on the ISIC, 2.1% on the BraTS mainly followed by a decrease in the focus of discriminative features attention. It verifies that attention layers are not auxiliary but are important in improving feature discrimination

especially in complex scenes. Finally, to treat added value of interpretability, Grad-CAM++ and SHAP were not included in the final model pipeline. Although this did not considerably influence raw segmentation accuracy, the indication of medical reviewers identified a significant decrease in perceived transparency of the model. That ability to interpret predictions was found to be the essential aspect in clinical reliability meaning that interpretability is not a luxury; it is a functional need as shown in table 10 and 11.

Table 10. Quantitative Results on ISIC 2018

Model	Dice	Jaccard	Precision	Recall	F1-score
U-Net	0.865	0.790	0.870	0.860	0.865
Attention U-Net	0.878	0.803	0.882	0.875	0.878
TransUNet	0.884	0.810	0.887	0.882	0.884
Proposed Method	0.902	0.836	0.905	0.898	0.901

Table 11. Quantitative Results on BraTS 2021

Model	Dice	Jaccard	Precision	Recall	F1-score
U-Net	0.852	0.778	0.860	0.845	0.852
Attention U-Net	0.866	0.791	0.872	0.860	0.866
TransUNet	0.874	0.800	0.880	0.868	0.874
Proposed Method	0.895	0.826	0.902	0.889	0.895

6. Discussion

From the experimental results we can clearly note that the proposed CNN Transformer hybrid segmentation model channeled with CBAM attention and explainability modules provides significantly higher accuracy in terms of several qualitative and quantitative benchmarks. Across both the ISIC 2018 and BraTS 2021 datasets, the model performs significantly better than the traditional CNN based models and is even better than the up-to-date transformer-integrated designs like TransUNet. The performance improvements are especially outstanding in cases with the presence of ambiguous lesion boundaries or heterogeneous tumor regions, suggesting the framework's expertise in local and global contextual properties. The ablation studies further reinforce the importance of each of the architectural components, and in particular the attention modules, which play a major role in the discriminative ability of the model and its precision. Besides accuracy, the incorporation of interpretability tools like Grad-CAM++ and SHAP in the segmentation pipeline is of pure necessity in terms of clinical application. Conventional deep learning approaches act like black boxes, so they are not the best for those domains, where transparency and traceability are required, like in medicine. The capacity to see which areas affected the model's choice will enable the clinicians to evaluate if the predictions are reliable and whether they can trust the automated outputs and where diagnosis breakdown can happen. Expert validation resulted in the conclusion that attention maps produced by our model correlate well with known clinical indicators (e.g., pigment networks in dermoscopy, or ring-enhancing areas in MRIs), making them viable before clinical expertise for decision-support purposes. However strong the proposed model is, there are some limitations to it. First, the computational cost is significantly larger than light weighted CNNs models because of transformer modules and attention methods. This might be a barrier to live deployment in resource-limited clinical settings. Secondly, even though the model generalizes very well to ISIC and BraTS datasets, it still needs detailed preprocessing and augmentation optimization for the best performance. Additionally, although informative, explainability methods like SHAP are computationally expensive and may not always be fitting to all use cases. The assumption of clean and annotated datasets by the model may not corroborate with noisy or incomplete datasets in the real world clinical setting. The modular structure of the proposed framework allows changing it for other medical imaging modalities (other than dermoscopy and MRI). For instance, the same pipeline can be further applied to the chest X-rays, the CT scans, or retinal fundus images, having the corresponding adaptation to the domain-specific adjustments. The usage of transformer layers enables global reasoning, which is especially helpful in a volumetric or longitudinal scenario. However, task-specific problems like modality-dependent noise,

scarcity of data, or variability of labels should be overcome via the application of transfer learning, weak supervision, or synthetic data generation to guarantee the performance's reliability. As explainable AI (XAI) is gaining popularity in medical imaging, it is bringing forth new questions of ethics and regulations. The introduction of transparency mechanisms does not relieve developers or the institutions from responsibility if clinical mistakes should occur. Regulatory institutions like the FDA and EMA are now beginning to publish guidelines for AI-driven diagnostic systems, especially focusing on accountability, fairness and recording the process of decision-making. Besides, consent of the patient, privacy of data, and bias in algorithms should be addressed cautiously. In this setting, our addition of explainability in the integration is not only a technical development, but a move to achieve the future legal and ethical requirements of trustworthy AI for healthcare.

To further justify the performance and the utility of the suggested framework, we benchmark and edge it against a few recent models of segmentation published between the year 2023 and 2025. It sums up the quantitative findings of these studies based on ISIC 2018 dataset with key measures of the Dice score, Jaccard index, precision, and recall. The segmentation accuracy was relatively high with TransUNet with attention (Jiang et al., 2023), and Swin-UNet with hybrid token fusion (Zhang et al., 2024) among the listed methods. Nonetheless, they require mechanisms only based on transformers, making them only interpretable to a limited level, particularly in clinical protocols where they should be more transparent. On the same note, Attention U-Net with LIME (Ahmed et al., 2023) incorporates post-hoc explanation tools, although the explanations are decoupled with the model learning pipeline. The suggested architecture where CNN and Transformer blocks are integrated with an inbuilt CBAM attention and Grad-CAM++ and SHAP explanations is always better than what has already been done. It obtained the best results in the Dice score (0.902) and similar precision (0.905) and good recall (0.898), so it formed a strong balance between accuracy and interpretability in segmentation. Our framework, unlike those, such as Lee et al. (2025), that used SHAP as a post processing step, integrates interpretability with prediction, improving model decision transparency and reliability by allowing it to be examined in detail. These findings argue in favor of the fact that the integration of the attention-enhanced hybrid networks with the embedded XAI-based mechanisms has the potential to produce clinically feasible and diagnostically interpretable segmentation systems.

7. Conclusions

We proposed this work an explainable hybrid deep learning model aimed at bio-medical image segmentation by coupling CNN-based encoder with Transformer-based decoder, implemented with CBAM attention modules and implementation of XAI tools (Grad-CAM++ and SHAP). The model solves a devastating shortcoming of the traditional segmentation strategies, as it offers a high precision of segmentation and interpretability in a genuinely, joint architecture. The results empirically demonstrated the use of the framework in two public bio-medical datasets as being better than the available methods. Compared to U-Net (Dice: 0.865) and TransUNet (Dice: 0.884), our model could perform better on ISIC 2018 with the Dice coefficient of 0.902, precision of 0.905, and recall of 0.898. On the part with that, the model achieved a score of 0.895 (Dice), 0.902 (precision), 0.889 (recall) on the BraTS 2021 dataset, which shows its stable performance in the problem of multi-modal segmentation with high levels of complexity. The advantages of integrating interpretability tools include clinical value beyond the numerical benefits. More than 85 percent of the expert-reviewed visual explanations were observed to correspond to the known anatomical or pathological markers making the predictions of the model clinically plausible and trustworthiness. Overall, the current research does not just assist in the improvement of the performance of deep segmentation models, but it also addresses the trust gap, implementing interpretability as an inseparable part of the prediction pipeline. This two-fold success forms a solid basis of real-life application of AI-aided tools of diagnostics in dermatology, oncology, radiology and so forth.

Acknowledgments: The authors would like to thank Al-Nahrain university and college of science for their support and valuable feedback throughout this research.

Conflicts of Interest: The authors declare that there are no conflicts of interest regarding the publication of this paper.

Funding: The authors received no external funding for this research.

References

[1] Ibrahim, A.; Al Sayed, I. A.M.; Jabbar, M. S.; Almutairi, H.; Sekhar, R.; Shah, P.; Al_Barazanchi, I.I.; "Evaluating the Impact of Emotions and Awareness on User Experience in Virtual Learning Environments for Sustainable Development Education". Ing. Syst. Inf., 29(1): 65–73, 2024.

[2] Ali ,Z. L.; Hayale, W. S. A.; Al Barazanchi, I. I.; Sekhar, R.; Shah, P.; Parihar, S.; "Efficient Cybersecurity Assessment Using SVM and Fuzzy Evidential Reasoning for Resilient Infrastructure". Ing. Syst. Inf., 29(2):515–521, 2024.

[3] Jasim, L. A.; Kamil, A. T.; Alsarraj, M. F. I.; Al_Barazanchi, I. I.; Sekhar, R.; Shah, P.; Malge, S.; "Network Fortification: Leveraging Support Vector Machine for Enhanced Security in Wireless Body Area Networks". Int. J. Saf. Secur. Eng., 14(3): 923–931, 2024.

[4] Hammad, A. H.; Abdulbaqi, A. S.; "Applying modified LBP for 2D ECG images classification". the 2nd International Conference on Emerging Technologies and Intelligent Systems, Malaysia, 3-4 Jun 2022; Al-Sharafi, A. M.; Al-Emran, M.; Al-Kabi, M. N.; Shaalan, K.; Springer, 2022.

[5] Al Barazanchi,I. I.; Hashim, W.; Thabit ,R.; Hussein, N. A.-H. K.; "Advanced Hybrid Mask Convolutional Neural Network with Backpropagation Optimization for Precise Sensor Node Classification in Wireless Body Area Networks". Khwârizmi Int. J., (2024): 17–31, 2024.

[6] Hassan, S. M.; Mohamad, M. M. B. ; Muchtar, F. B.; "Comprehensive taxonomy of schemes for detecting and mitigating blackhole attacks in mobile ad-hoc networks: A study on tactics, classifications, and future directions". Ing. Syst. Inf., 29(6): 2533–2548, 2024.

[7] Abdulbaqi, A. S.; Panessai, I. Y.; "Designing and implementation of a biomedical module for vital signals measurements based on embedded system". Int. J. Adv. Sci. Technol., 29(3): 3866-3877, 2020.

[8] Hassan, S. M. ; Mohamad, M. M. ; Muchtar, F. B. ; Dawoodi, F. B. Y. P. ; "Enhancing MANET security through federated learning and multiobjective optimization: A trust-aware routing framework". IEEE Access, 12: 181149–181178, 2024.

[9] Hassan, S. M. ; Mohamad, M. M. ; Muchtar, F. B. ; "Advanced intrusion detection in MANETs: A survey of machine learning and optimization techniques for mitigating black/gray hole attacks". IEEE Access, 12 : 150046 - 150090, 2024,

[10] Ibraheem,I. N. ; Hassan, S. M.; Abead, S. A. ; "Comparative analysis & implementation of image encryption & decryption for mobile cloud security" . Int. J. Adv. Sci. Technol., 29(3),: 109–121, 2020.

[11] Alkhakani S. B.; Hassan, S. M.; "Suggest trust as mediation to adopt electronic voting system in Iraq". J. Theor. Appl. Inf. Technol., 96(17): 5729–5739, 2018.

[12] Ahmed, E. Q. ; Ameen, Z. A. ; Al-Mukhtar, F. S.; Jaaz, Z. A.; "Maximizing Mobile Communication Efficiency with Smart Antenna Systems using Beam forming and DOA Algorithms". 10th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), Palembang, Indonesia, 20-21 September 2023; IEEE Access: United States, Piscataway, New Jersey .2023,

[13] Ameen, Z.H; Al-Mukhtar, F. S.; Ahmed, E.Q; Dhannoon, B.N; "A Review of the Past Decade's Research Outcomes on Classifying Cyberbullying". Al-Nahrain J. Sci., 27(1): 121–135, 2024.

[14] Kaya, M.; Shahid, H.; "Cross-Border Data Flows and Digital Sovereignty: Legal Dilemmas in Transnational Governance". Interdiscip. Stud. Soc. law polit., 4(2): 219-233, 2025.

[15] Oleiwi, S. S. ; Mohammed, G. N. ; Al-Barazanchi, I. ; "Mitigation of packet loss with end-to-end delay in wireless body area network applications" . Int. J. Electr. Comput. Eng., 12(1): 460–470, 2022.

[16] Shawkat, S. A. ; Tuama, B. A. ; Al-Barazanchi, I.; "Proposed system for data security in distributed computingin using triple data encryption standard and Rivest Shamir Adlemen". Int. J. Electr. Comput. Eng., 12(6): 6496–6505, 2022.

[17] Abdulshaheed, H. R. ; Abbas, H. H. ; al Barazanchi, I. ; Hashim, W. ; "Control and alert mechanism of RFID door access control system using Iot". 3C Tecnol. , 40(2): 269–285, 2022.

[18] Abdulbaqi, A. S., Obaid, A. J., Mohammed, A. H.; "ECG signals recruitment to implement a new technique for medical image encryption". J Discrete Math Sci Crypt 24 (6): 1663–1673, 2021.

[19] Shawkat , S. A. ; Al-Barazanchi, I. ; "A proposed model for text and image encryption using different techniques". Telecommun. Comput. Electron. Control. , 20(4): 858–866, 2022.

[20] Al-Barazanchi I. ; Niu Y.; Abdulshaheed, H. R.; Hashim, W.; Alkahtani, A. A.; Daghighi, E.; Jaaz, Z. A.; et al ; "Proposed New Framework Scheme for Path Loss in Wireless Body Area Network". Iraqi J. Comput. Sci. Math., 3(1): 11–21, 2022.

[21] Faisal, G. M. ; Alshadoodee, H. A. A. ; Abbas, H. H. ; Gheni, H. M. ; Al-Barazanchi, I. ; "Integrating security and privacy in mmWave communications". Bull. Electr. Eng. Info. 11(5) : 2856–2865, 2022.

[22] Sahy, S. A. ; Mahdi, S. H. ; Gheni, H. M. ; Al-Barazanchi, I. ; "Detection of the patient with COVID-19 relying on ML technology and FAST algorithms to extract the features" . Bull. Electr. Eng. Inform., 11(5): 2886–2894, 2022.

[23] Al-Mukhtar F.S.; "Automated Face Mask Detection Using Pretrained CNN". Al-Nahrain J. Sci., 26 (4): 80-87, 2023.

[24] Abdulbaqi, A. S.; Al-Attar, B.; Zaid Qudr, L. A.; Penubadi, H. R.; Sekhar, R.; Shah, P.; Tawfeq, J. F.; "Efforts of Neutrosophic Logic in Medical Image Processing and Analysis". Int. J. Neutrosophic Sci., 24(4): 376-388, 2024.

[25] Ameen Z.H. , Al-Mukhtar F.S., Ahmed E.Q., Dhannoon B.N.; "A Review of the Past Decade's Research Outcomes on Classifying Cyberbullying". Al-Nahrain J. Sci., 27 (1): 121-135,2024.

[26] Abdulshaheed, H. R. ; Abbas, H. H. ; Ahmed, E. Q. ; Al-Barazanchi, I. ; "Big Data Analytics for Large Scale Wireless Body Area Networks;Challenges, and Applications" . Advances on Intelligent Informatics and Computing (IRICT 2021), Completely virtual (online), 30 March 2022; Saeed, F; Mohammed, M.; Ghaleb, F.; Springer, Germany, Berlin, 2022.

[27] Abdulbaqi, A. S.; Turki, N. A.; Obaid, A. J.; Dutta, S.; Panessai, I. Y.; "Spoof Attacks Detection Based on Authentication of Multimodal Biometrics Face-ECG Signals". In Artificial Intelligence for Smart Healthcare; 10 June 2023; Agarwal , P.; Khanna, K.; Elngar, A. A.; Obaid A. J.; Polkowski, Z.; EAI/Springer Innovations in Communication and Computing : Cham, Switzerland, 507–526, 2023.