

Article

GISLC: Gated-Inception Model for Skin Lesion Classification

Tamam Alsarhan ^{1,*}, Mohammad Kamal Abdulaziz ¹, Ahmad Ali ², Ayoub Alsarhan ^{3,4}, Sami Aziz Alshammari ^{5,*} ,
Rahaf R. Alshammari ⁶ , Nayef H. Alshammari ⁷  and Khalid Hamad Alnafisah ⁸ 

- ¹ King Abdullah II School of Information Technology, The University of Jordan, Amman 11942, Jordan; mhm0217426@ju.edu.jo
- ² Department of Computer Science and Technology, Hainan Bielefeld University of Applied Sciences, Hainan, 571700, China; ali.ahmad@hainan-biuh.edu.cn
- ³ Department of Data Science and Artificial Intelligence, Faculty of Information Technology, Al-Ahliyya Amman University, Amman 19111, Jordan; ayoubm@hu.edu.jo
- ⁴ Department of Information Technology, Faculty of Prince Al-Hussein Bin Abdullah, The Hashemite University, Zarqa 13133, Jordan
- ⁵ Department of Information Technology, Faculty of Computing and Information Technology, Northern Border University, Rafha 76313, Saudi Arabia
- ⁶ Department of Computer Science, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh 11671, Saudi Arabia; raralshammari@pnu.edu.sa
- ⁷ Department of Computer Science, Faculty of Computers and Information Technology, University of Tabuk, Tabuk 71491, Saudi Arabia; nh.alshammari@ut.edu.sa
- ⁸ Department of Computer Sciences, Faculty of Computing and Information Technology, Northern Border University, Rafha 76313, Saudi Arabia; khalid.alnafisah@nbu.edu.sa
- * Correspondence: t_alsarhan@ju.edu.jo (T.A.); sami.alshammari@nbu.edu.sa (S.A.A.)

Abstract

Skin-lesion recognition from clinical photographs is clinically valuable yet computationally challenging due to large intra-class variation, subtle inter-class boundaries, class imbalance, and heterogeneous acquisition conditions. To address these constraints under realistic compute budgets, we investigate Inception-family convolutional baselines and propose GISLC—a Gated-Inception model that augments a GoogLeNet/Inception-V1 backbone with a lightweight, spatial gating head inspired by ConvLSTM. Unlike static fusion (concatenation/summation) of multi-branch features, the proposed gated head performs per-location, learnable regulation of feature flow across branches, prioritizing diagnostically salient patterns while suppressing redundant activations. Experiments were conducted on the clinical-images subset of the Multimodal Augmented Skin Lesion Dataset (MASLD), an augmented derivative of HAM10000, using stratified train/validation/test splits, clinically motivated augmentation, and class-weighted optimization to mitigate skewed label frequencies. A controlled ablation study evaluates backbone choices and optimization settings and isolates the contribution of gated fusion relative to standard Inception heads. Across runs, the gated fusion strategy improves discriminative performance while remaining parameter-efficient, supporting the view that spatially adaptive regulation can enhance robustness on non-dermatoscopic clinical imagery. We further outline practical steps for calibration analysis and compression-aware deployment in clinical and edge settings.



Academic Editor: Yue Wu

Received: 11 January 2026

Revised: 5 February 2026

Accepted: 11 February 2026

Published: 18 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: skin cancer; clinical images; deep learning; inception; gated inception; convlstm; medical image analysis

1. Introduction

The skin serves as a resilient barrier, isolating internal organs and systems from the external environment [1]. It consists of three main layers: the epidermis, dermis, and hypo-

dermis. In addition to its structural complexity, the skin performs vital physiological roles, such as safeguarding against physical and microbiological damage, regulating temperature, and facilitating sensory awareness [2,3]. The typical surface area in humans is roughly 20 square feet, signifying a vital aspect of human health and serving as the body's primary protective interface with the environment.

Although resilient, the skin is susceptible to numerous pathological alterations induced by intrinsic and external causes, including aging, UV radiation, infections, and trauma. These changes frequently present as skin lesions, characterized by variations in color, texture, or morphology—such as patches, lumps, or uneven growths. Lesions are clinically categorized as benign or malignant based on their capacity to invade adjacent tissues and spread [4]. Benign lesions are limited, whereas malignant lesions, particularly melanoma and non-melanoma skin malignancies, pose considerable health hazards due to their aggressive characteristics. Therefore, prompt and precise diagnosis of malignant lesions is crucial for efficient therapy and improved patient outcomes [5,6].

Skin cancer is among the most prevalent malignancies and arises when genomic damage triggers dysregulated proliferation of cutaneous cells [7]. Prolonged exposure to ultraviolet (UV) radiation—whether from sunlight or artificial sources—induces DNA injury that can accumulate and promote tumor formation [8]. Three main forms of skin cancer are identified as basal cell carcinoma (BCC), squamous cell carcinoma (SCC), and melanoma. BCC arises in the basal cells of the lower layers of the epidermis and often manifests as lesions on the surface. SCC, on the other hand, results from an excess accumulation of squamous cells as a consequence of prolonged UV exposure [9]. Although melanoma is the least frequent type, it is also the most aggressive and life-threatening. Melanoma, also known as the “black tumor”, represents only around 1% of all skin cancer diagnoses but causes the majority of skin cancer-related deaths [10]. According to the World Health Organization's World Cancer Report, around 1.3 million new melanoma cases and more than 2.5 million non-melanoma cases are recorded worldwide each year, making skin cancer one of the most common malignancies [11]. Historically, detection has involved visual inspection and biopsy. Dermatologists often rely on direct examination of skin lesions and, in many cases, employ a dermatoscope—a handheld magnifying device—to study the structure and finer details of suspicious abnormalities [12]. While dermatoscopy enhances clinical decision-making, its interpretation remains subjective, with outcomes varying across observers and constrained by the limited availability of dermatology specialists [13].

Automated classification of clinical skin-lesion images remains challenging because lesion classes exhibit high intra-class variability and subtle inter-class differences, datasets are often imbalanced, and acquisition conditions vary across devices and clinical settings [14]. Conventional convolutional backbones may pass redundant or noisy multi-scale features to downstream classifiers, requiring large parameter budgets or extensive fine-tuning to achieve robustness. To address these limitations, large collections of clinical and dermatoscopic images with consistent annotation have been established, enabling the development of automated diagnostic systems. Publicly curated datasets such as HAM10000 [13], ISIC Archives [15,16], and their derivatives have been particularly influential, providing standardized benchmarks that allow deep learning models to be trained and validated under reproducible conditions [17,18]. In this study, we utilize the Multimodal Augmented Skin Lesion Dataset (MASLD) [19], which extends HAM10000 and integrates multiple imaging modalities. Specifically, only the clinical-images subset of MASLD was used for both training and evaluation, reflecting realistic non-dermatoscopic imaging scenarios encountered in general clinical practice.

Recent research has explored architectures that effectively capture the complexity of skin-lesion images. Models such as the Inception family strike a balance between

receptive-field diversity and computational efficiency, making them well suited for tasks where both local texture and global context matter [20,21]. Complementary to this, gating mechanisms—originally developed for sequential models such as LSTMs—offer dynamic control over information flow and have been adapted for spatial and multimodal fusion tasks in vision [22,23]. These advances, combined with deployment needs in Internet of Medical Things (IoMT) and edge-computing environments [24,25], motivate the design of models that are compact, efficient, and sensitive to clinically significant lesion classes. Inception-based architectures are effective for capturing multi-scale features; however, their branch outputs are typically fused through static operations such as concatenation or summation [26]. In contrast, gating mechanisms derived from recurrent networks provide a dynamic means of regulating information flow and have recently been adapted for spatial fusion tasks to improve attention on salient regions [27,28]. Beyond conventional RGB dermatoscopic and clinical imaging, recent advancements in medical imaging have started to investigate more powerful modalities such as hyperspectral imaging for skin cancer assessment. For example, authors in [29] proved that hyperspectral imaging integrated with machine learning can remarkably improve lesion discrimination by capturing spectral signatures that are hidden in conventional imaging. Such techniques suggest a wider shift in dermatological AI toward leveraging more informative imaging sources to improve diagnostic accuracy. Nonetheless, regardless of the imaging technique—be it traditional RGB, dermatoscopic, or hyperspectral—the efficacy of automated diagnosis ultimately hinges on the model’s ability to integrate multi-scale and geographically diverse data while mitigating redundant or noisy representations. Enhancing feature-fusion strategies in convolutional architectures is crucial, as effective fusion processes are modality-agnostic and can generalize across existing and novel imaging technologies. This encourages the investigation of adaptive gating-based fusion in Inception-style networks, as shown in this study.

This study aims to introduce a classification architecture for skin lesions that is both computationally efficient and diagnostically valuable, building on previous developments. In particular, we introduce the Gated-Inception model for Skin Lesion Classification (GISLC), a hybrid convolutional–gating architecture designed to improve discriminative performance while maintaining parameter efficiency. This model combines multiscale convolutional representations with spatial gating to prevent redundant feature propagation, emphasize diagnostically significant patterns, and maintain parameter efficiency in real-world deployment scenarios such as IoMT and edge computing [30]. To achieve this objective, the study evaluates conventional convolutional models from the Inception family and develops a hybrid architecture known as Gated-Inception. This architecture combines an Inception-style backbone with a spatial gating head inspired by ConvLSTM [22] to facilitate adaptive, per-location fusion of multiscale information.

To summarize, the significant contributions of this research are presented as follows:

- Curated an end-to-end clinical-image classification workflow for MASLD (clinical modality), including deterministic stratified splitting, clinically plausible augmentation, and imbalance-aware training.
- Established reproducible baselines using Inception-family architectures (Inception-V1/Inception-V3/Inception-V4) under controlled optimization and fine-tuning policies.
- Proposed GISLC, which couples an Inception-V1 backbone (frozen through inception5b) with a compact multi-branch head and a ConvLSTM-inspired GateCell2D module to perform spatially adaptive fusion.
- Performed systematic ablations over optimizer families, learning-rate/weight-decay settings, and backbone-freeze strategies to isolate stability and performance drivers.

- Reported complementary evaluation indicators (accuracy, precision, recall, weighted F_1 , and calibration-oriented observations) with structured experiment logging to support repeatability and auditability.

The remainder of this manuscript is organized as follows: Section 2 summarizes related work on CNN backbones, gating/attention mechanisms, and dermatology benchmarks. Section 3 details the dataset preparation pipeline and the proposed GISLC architecture, including training and ablation protocols. Section 4 reports experimental results with comparative analysis across baselines and variants. Section 5 concludes the paper and outlines directions for validation and deployment-oriented extensions.

2. Related Work

The integration of deep learning within medical imaging has expanded substantially over the last decade, driven by the need for accurate, scalable, and automated diagnostic tools [31]. Deep neural networks have enabled significant advancements in feature extraction, disease localization, and classification, providing physicians with computational support that enhances diagnostic precision [32].

2.1. Deep Learning Architectures

One of the earliest breakthroughs in computer vision was the introduction of the Inception architecture (GoogLeNet) [20], which demonstrated the benefits of multi-scale feature extraction within a single layer through parallel convolutional operations. This model significantly improved classification accuracy while reducing computational costs compared with traditional convolutional neural networks (CNNs) like AlexNet [33] and VGGNet [34]. Subsequent refinements, such as Inception-V3 [21] and Xception [35], further enhanced representational power by employing factorized convolutions and auxiliary classifiers, making them strong baselines for various computer vision applications, including medical image analysis [36]. Residual networks (ResNet) also revolutionized the field by introducing skip connections to mitigate the vanishing gradient problem, allowing for much deeper networks [37]. More recently, DenseNet [38] and EfficientNet [39] have offered alternative strategies for feature reuse and compound scaling, respectively, often yielding state-of-the-art results in lesion classification tasks [24,40].

2.2. Gating, Attention, and Recurrent Models

Beyond pure CNNs, researchers have sought to incorporate mechanisms that improve the ability of neural networks to model sequential dependencies and contextual relationships. Long Short-Term Memory (LSTM) networks [41] were developed to overcome the vanishing-gradient problem in recurrent architectures. The gating mechanisms within LSTM—input, forget, and output gates—allow dynamic control of information flow, making them highly effective for sequential data modeling. The potential of these gates extends beyond text and speech processing, as they can also refine visual features by selectively preserving or discarding information at different abstraction levels [42].

In the spatial domain, Convolutional LSTM (ConvLSTM) [22] extended this concept by replacing matrix multiplications with convolution operations, preserving spatial layout essential for image analysis. The idea of combining convolutional backbones with recurrent gating has been explored across multiple domains. For instance, fusion models integrating CNNs and LSTMs have been applied to complex tasks such as video recognition and biomedical imaging, showing that temporal or sequential feature modeling enhances discriminative power [43,44]. Recent attention mechanisms, such as Squeeze-and-Excitation (SE) blocks [28] and the Convolutional Block Attention Module (CBAM) [27], operate on similar principles of adaptive feature recalibration. These modules serve as modulators

for feature extraction pipelines, offering dynamic adjustments that compensate for noise or redundant signals in the data [45]. Such hybrid strategies have proven promising in improving robustness under variable imaging conditions [46,47].

2.3. Early Dermatology Systems and Deep CNNs

In dermatology, early computer-aided diagnostic systems relied primarily on hand-crafted features such as asymmetry, border irregularity, color variation, and diameter (ABCD rule) [48,49]. While these approaches achieved limited success, they struggled with generalization due to inter-patient variability, inconsistent imaging conditions, and dataset differences. The introduction of deep convolutional models significantly changed this landscape [50]. CNN-based frameworks achieved near-dermatologist performance in melanoma classification tasks, with architectures such as Inception-V3 and ResNet consistently outperforming traditional machine-learning baselines [36,51]. However, generalization across datasets remained a persistent challenge, particularly when transitioning from controlled dermatoscopic to real-world clinical imaging [52,53].

2.4. Datasets and Data Augmentation

To improve variability, several large, curated benchmarks were proposed, most notably the ISIC Archive and the HAM10000 dataset [13,15]. Its release enabled the development of more generalized models by providing a sufficient volume of labeled samples for deep learning. Subsequent works augmented such datasets through synthetic image generation and transformations to improve class balance and robustness [54,55]. These augmentations are essential because the distribution of lesion types is typically imbalanced, with benign conditions significantly outnumbering malignant cases [56]. Techniques such as rotation, flipping, brightness adjustment, and color shifting have been shown to enhance CNN generalization and reduce overfitting [57]. More advanced augmentation strategies, such as CutMix [58] and MixUp [59], have also been applied to dermatological data to encourage the model to learn less localized features.

In this study, the dataset used is the Multimodal Augmented Skin Lesion Dataset (MASLD) [19], an augmented extension of HAM10000. Specifically, only the clinical-images subset—comprising approximately 900 samples across nine lesion categories—was employed for both training and evaluation. This subset best reflects realistic acquisition conditions outside specialized dermatoscopic setups, ensuring applicability to non-invasive, real-world clinical diagnostics [60].

2.5. Transfer Learning in Medical Imaging

Parallel to dataset improvements, architectural innovations have continued to evolve. Inception-based models, due to their efficiency and ability to process multi-scale features, became a reliable choice for medical imaging pipelines [61,62]. Several studies emphasized that freezing and unfreezing layers of pretrained models can significantly influence performance depending on the similarity between the source dataset (e.g., ImageNet) and the target medical dataset [63,64]. Transfer learning remains a cornerstone of medical image classification, enabling large-scale pretraining for domains with limited labeled data [65]. Combined with advanced data augmentation frameworks and regularization strategies such as AdamW [66] and label smoothing [67], transfer learning facilitates robust and efficient model convergence [68].

2.6. Hybrid Gated CNN–Inception Models and Transformers

Recent explorations into gated CNN–Inception hybrids have demonstrated promising results, as gating modules can dynamically regulate which features are propagated forward, mimicking selective attention mechanisms in biological vision systems [69,70]. These

approaches reduce redundancy and emphasize clinically significant features—particularly important in skin lesion classification, where subtle texture variations may distinguish malignant from benign lesions [71]. Furthermore, the advent of Vision Transformers (ViT) [72] and hybrid Transformer–CNN models has pushed the boundary of performance further by capturing long-range dependencies in lesion images [73,74]. However, purely transformer-based models often require vast amounts of data to converge, making hybrid convolutional approaches more practical for smaller medical datasets [75].

In summary, deep learning in dermatology has transitioned from handcrafted feature extraction to sophisticated, gated hybrid architectures capable of adaptive feature selection and attention-driven learning. The proposed Gated-Inception model builds upon these advancements, combining the multiscale efficiency of Inception with a ConvLSTM-inspired gating mechanism to enhance discriminative performance while maintaining computational efficiency suitable for edge deployment on clinical devices.

3. Methodology

This section describes the methodological pipeline used to design, train, and validate the proposed Gated-Inception model for Skin Lesion Classification (GISLC). The process follows a structured workflow: dataset preprocessing and preparation; conceptual design and implementation of the hybrid model; training configuration, evaluation strategy, and ablation design; and finally, deployment considerations, limitations, and future directions. All experimental settings, seeds, and preprocessing parameters were logged to ensure full reproducibility.

3.1. Preprocessing and Data Preparation

Preprocessing converts raw clinical images into standardized, reproducible inputs suitable for model training and evaluation. In this work, we implemented a Stratified Group K-Fold Cross-Validation protocol ($K = 5$) rather than depending on a singular static train/validation/test division. This technique involved grouping images by patient identity, ensuring that all images associated with a single patient were contained within the same fold. This mitigates patient-level data leakage, a recognized cause of optimistic bias in medical imaging research. For each fold, internal training, validation, and testing subsets were created while maintaining class distribution by stratification [76]. The performance measurements presented in this work reflect the mean and standard deviation across five folds, offering resilience against biased data partitioning and ensuring dependable generalization estimates.

To guarantee reproducibility, a global random seed governed shuffling and splits, and file identities were documented for each execution. Input images were resized to match the requirements of the Inception backbones, producing feature maps of approximately $(B, 1024, 7, 7)$. During training, bicubic resizing followed by random cropping was applied, while validation and test images were preprocessed using bicubic resizing with a deterministic center crop [57].

3.2. Data Augmentation

To improve model generalization and mitigate overfitting, a series of clinically guided data augmentations were implemented during training [77]. Geometric modifications included horizontal flipping, minor rotations (± 10 – 20°), scaling/zooming, translations, and shear operations, all selected to replicate natural variations in clinical image acquisition while preserving the diagnostic characteristics of the lesion. Photometric modifications included controlled brightness, contrast, and saturation variations, accompanied by regulated hue adjustments to maintain clinically relevant color cues. These augmentations were

intended to enhance the effective training distribution while preserving fidelity to actual clinical imaging conditions. Details are shown in Figure 1.

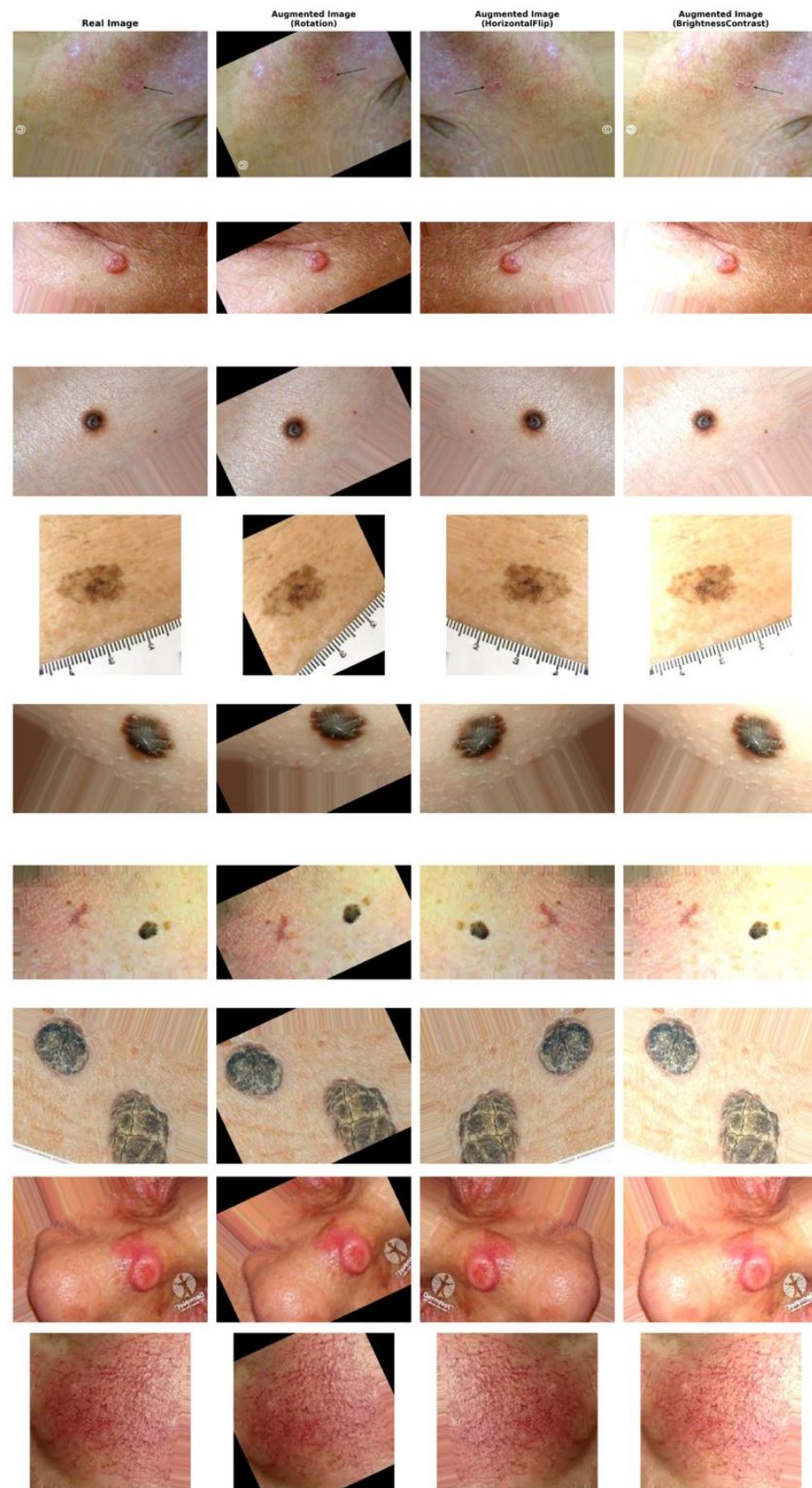


Figure 1. Data augmentation techniques.

Class weights were calculated based on training frequencies and utilized in a weighted cross-entropy loss to mitigate class imbalance [78]. Calibration was further improved through dropout regularization [79] and selective label smoothing [67]. The training pipeline was designed for efficiency and reproducibility, employing multi-worker data loaders to optimize GPU utilization and automated mixed precision (AMP) [80] to enhance throughput and memory efficiency. Detailed experiment metadata, including seeds, augmentation parameters, and split identifiers, were documented in JSON logs and stored with model artifacts to ensure complete reproducibility across executions.

3.3. Gated-Inception Model

The proposed GISLC architecture consists of three main components: the Frozen GoogLeNet Backbone, the GatedInceptionHead, and the ClassifierHead, as shown in Figure 2. GoogLeNet (Inception-V1) [20] was selected as the backbone due to its proven multi-scale feature extraction capability, parameter efficiency, and suitability for medical imaging tasks, where preserving fine spatial details and texture patterns is critical under limited dataset conditions. Specifically, our model utilizes a pretrained GoogLeNet backbone (Inception-V1), which is frozen up to the inception5b block, and extends it with a custom GatedInceptionHead. This head consists of four parallel branches: a 1×1 convolution, a reduced 3×3 convolution, stacked 3×3 convolutions (effectively 5×5), and average pooling followed by a 1×1 convolution. The outputs are sequentially merged using a ConvLSTM-inspired GateCell2D. The Gated-Inception model was designed with three main objectives: to replace static concatenation with learnable, spatially aware fusion; to achieve parameter efficiency without employing deeper backbones; and to maintain lesion boundaries and textural cues during feature integration.

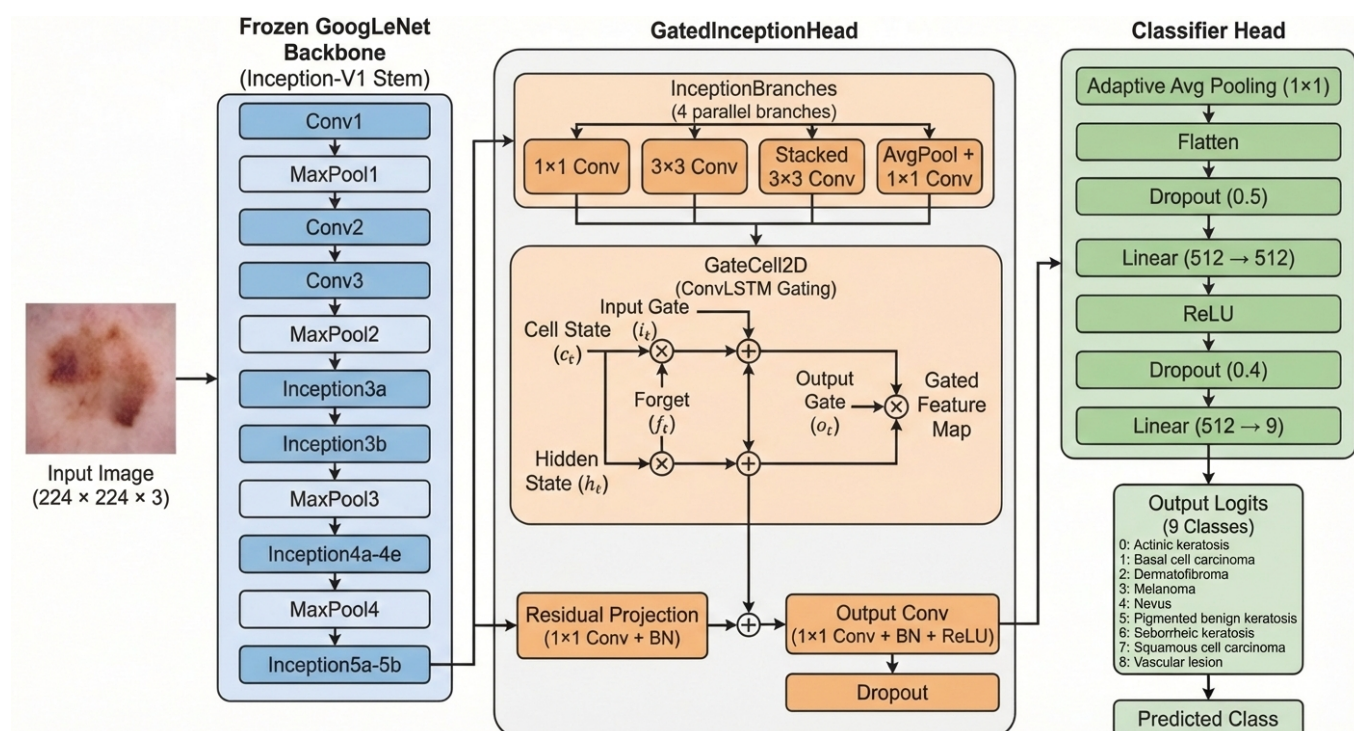


Figure 2. High-level conceptual view of Gated-Inception. Blue = Frozen Inception Backbone; Orange = Trainable Gated Head (GateCell2D Block); Green = Classifier Head.

The GateCell2D mechanism employs convolutional input, forget, output, and candidate gates across spatial dimensions, preserving hidden and cell states to perform order-aware, multi-branch fusion [22]. Within this framework, the input to the gating process

originates from the feature maps generated by the frozen GoogLeNet backbone. Let $X \in \mathbb{R}^{C \times H \times W}$ denote the feature map produced by the frozen GoogLeNet backbone. The four parallel Inception branches generate multi-scale feature representations:

$$B_i = f_i(X), \quad i \in \{1, 2, 3, 4\} \quad (1)$$

where $f_i(\cdot)$ represents the convolution or pooling operation of the i th branch. Instead of performing post-fusion attention as in SE or CBAM, GISLC performs pre-fusion sequential gating. At each step t , the GateCell2D processes one branch output B_t while conditioning on the accumulated spatial state from previous branches. The gating operations are defined as:

$$i_t = \sigma(W_i * B_t + U_i * H_{t-1}) \quad (2)$$

$$f_t = \sigma(W_f * B_t + U_f * H_{t-1}) \quad (3)$$

$$o_t = \sigma(W_o * B_t + U_o * H_{t-1}) \quad (4)$$

$$\tilde{C}_t = \tanh(W_c * B_t + U_c * H_{t-1}) \quad (5)$$

where i_t , f_t , and o_t denote the input, forget, and output gates, respectively; H_{t-1} is the hidden state from the previous branch; and $*$ denotes convolution. The internal cell and hidden states are then updated as:

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (6)$$

$$H_t = o_t \odot \tanh(C_t) \quad (7)$$

This formulation allows each branch to be spatially conditioned by the responses of preceding branches before aggregation occurs. After all four branches are processed sequentially, the final fused representation is obtained as:

$$F = H_4 \quad (8)$$

The fused feature map is passed through a compact classification head consisting of a 1×1 convolution, dropout, global adaptive pooling, and a lightweight multilayer perceptron to produce the final prediction:

$$\hat{y} = \text{Softmax}(W_{cls} \cdot \text{GAP}(F)). \quad (9)$$

The Gated-Inception architecture (schematic) is illustrated in Figure 3.

The final representation is processed by a compact classification head comprising a 1×1 convolution, dropout, adaptive pooling, and a streamlined multilayer perceptron (MLP) for nine-class prediction. This approach balances precision, efficiency, and stability, facilitating adaptive multi-scale feature fusion specifically for skin-lesion classification.

Beyond architectural efficiency, it is important to clarify how the proposed fusion strategy conceptually differs from conventional attention mechanisms. The proposed gating mechanism differs fundamentally from widely used attention modules such as Squeeze-and-Excitation (SE) and Convolutional Block Attention Module (CBAM) [28]. In SE and CBAM, attention is applied after multi-branch feature extraction and aggregation. In contrast, the GISLC architecture introduces pre-fusion sequential spatial gating, where feature refinement occurs during the fusion process rather than after it.

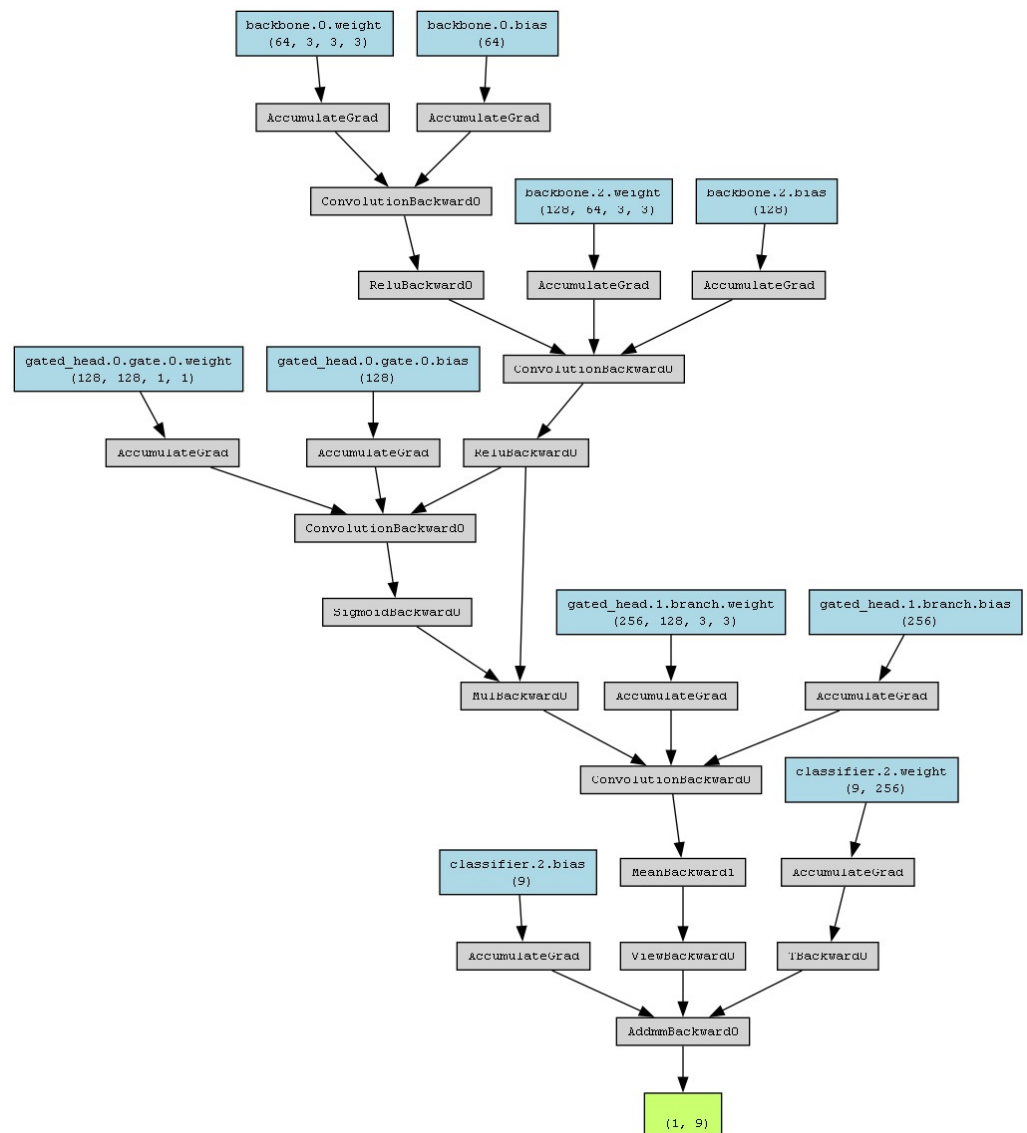


Figure 3. Gated-Inception architecture (schematic). A frozen GoogLeNet backbone (blue) produces dense feature maps, which are adaptively fused by a ConvLSTM-inspired gated head (green). The fused features pass through a 1×1 convolution, dropout, adaptive pooling, and a compact MLP to produce class logits.

3.4. Training Procedure

The model was trained using class-weighted categorical cross-entropy, with milestones determined by validation accuracy and early stopping [81]. The validation loss was optimized using the AdamW optimizer ($lr = 2 \times 10^{-4}$, weight decay = 1×10^{-4}) [66] with a ReduceLROnPlateau scheduler. Regularization included dropout within the gated head, selective label smoothing for baselines, and AMP for faster and more stable convergence [80]. To guarantee a fair comparison of architectures, all backbone models were trained using the same protocol. This encompassed identical data partitions, augmentation procedures, input resolution, batch size, optimizer settings, learning rate schedule, epoch count, and early termination conditions. A uniform approach was employed for transfer learning: initial training with the backbone layers frozen, succeeded by regulated fine-tuning of the upper layers in all models.

4. Experiments and Discussion

4.1. Dataset

The experiments utilized the Multimodal Augmented Skin Lesion Dataset (MASLD) [19], which extends HAM10000 [13] via augmentation and multimodal organization. In this study, only the clinical-images subset was used, comprising 900 color images (≈ 100 images per class) across nine lesion categories as shown in Figure 4. Images are provided as JPG/PNG files with variable spatial resolution and include accompanying metadata (e.g., age, sex, and anatomical site), enabling demographic and context-aware analysis when required [16]. Focusing on the clinical modality targets non-dermatoscopic acquisition conditions, aligning the evaluation with practical screening scenarios.

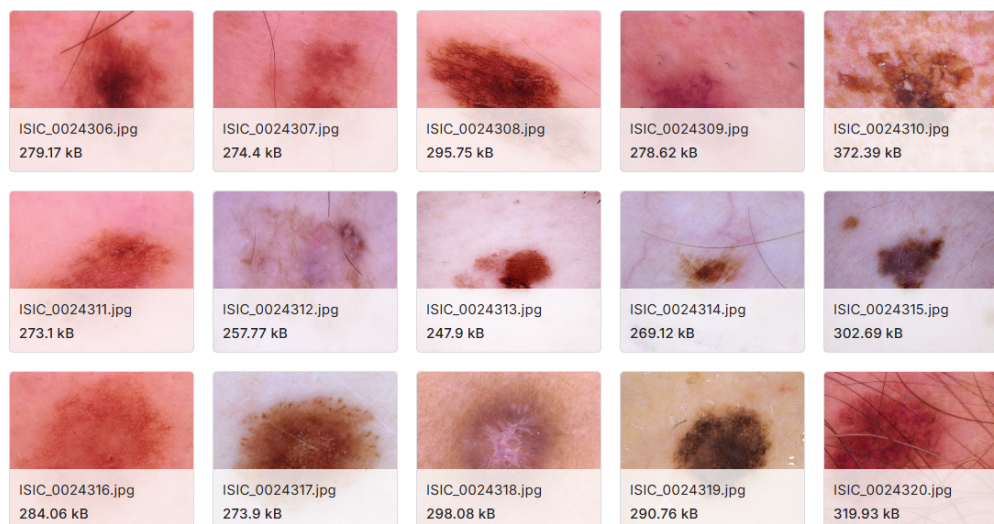


Figure 4. Samples from the clinical-images subset of the Multimodal Augmented Skin Lesion Dataset (MASLD).

4.2. Evaluation Metrics

Model performance was quantified using four widely adopted classification criteria: accuracy, precision, recall (sensitivity), and the F_1 -score [82]. Together, these measures capture overall correctness and error trade-offs that are clinically relevant, particularly under class imbalance where accuracy alone can be misleading.

Accuracy measures the proportion of correctly classified instances among all predictions:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}. \quad (10)$$

Precision quantifies the ratio of correctly predicted positive samples to all samples predicted as positive:

$$\text{Precision} = \frac{TP}{TP + FP}. \quad (11)$$

Recall (sensitivity) measures the ability of the model to correctly identify all positive samples:

$$\text{Recall} = \frac{TP}{TP + FN}. \quad (12)$$

Finally, the F_1 -score provides the harmonic mean of precision and recall:

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (13)$$

These complementary metrics provide a more reliable summary of diagnostic behavior than any single score, helping verify that improvements are not driven by over-predicting majority classes [83].

4.3. Results

The ablation study analyzed the effects of different architectural designs and optimization strategies on classification performance. The proposed Gated-Inception consistently achieved superior accuracy and robustness compared with all evaluated baselines, underscoring its enhanced representational capacity for clinical skin-lesion recognition. The comparative evaluation encompassed Inception-V1, Inception-V3, and Inception-V4 [84], as well as advanced extensions such as Transception (Inception-style convolutions with Transformer-based self-attention) and Inception-ResNetV2-CBAM (Convolutional Block Attention Module for refined spatial-channel learning) [27,84]. This combination provided a broad analysis across convolutional, hybrid, and attention-driven paradigms.

Experiments with multiple optimizers and learning rates revealed substantial effects on convergence stability and final accuracy. The most reliable and high-performing configuration paired Gated-Inception with the AdamW optimizer at a learning rate of 2×10^{-4} . These findings indicate that achieving strong results depends on both architectural innovation and a carefully tuned optimization strategy [85,86].

Crucially, Tables 1 and 2 illustrate the results of a standardized hyperparameter grid search conducted uniformly across all backbone models, ensuring a rigorous “apples-to-apples” comparison. To eliminate experimental bias, all architectures were assessed using an identical optimizer and learning rate search space with a strict “Frozen Backbone” protocol. The instability observed in deeper models (e.g., Inception-V4) indicates their high susceptibility to overfitting when adapting to limited medical datasets without extensive fine-tuning. In contrast, the stability of the Gated-Inception architecture under identical conditions demonstrates its superior inductive bias for feature modulation.

Table 1. Comparative performance of the proposed Gated-Inception architecture against standard CNN baselines and hybrid Transformer models. To ensure a fair evaluation, all backbones were frozen during training. Results represent the mean $\pm$ standard deviation across 5-fold Stratified Group Cross-Validation. The proposed model achieves state-of-the-art accuracy with significantly lower variance. Best results are in bold.

Model	State	Params (M)	Lat. (ms)	Acc. (%)	F1 (%)
Standard CNN Baselines					
Inception-V1 (Baseline)	Frozen	6.1	11.32	82.73 ± 1.59	82.49
Inception-V3	Frozen	27.7	17.04	79.85 ± 2.20	79.83
Inception-V4	Frozen	43.2	31.50	53.05 ± 3.50	53.08
Inception-ResNet-V2	Frozen	56.7	39.80	68.66 ± 2.98	68.69
Hybrid & Proposed					
Transception (ViT Hybrid)	Frozen	7.8	13.50	97.23 ± 0.50	97.22
Gated-Inception (Ours)	Frozen	12.9	13.32	98.23 ± 0.65	98.23

Table 2. Comprehensive hyperparameter sensitivity analysis detailing the performance of all architectures across the full experimental grid. All models utilized frozen backbones. Metrics are reported as mean $\pm$ standard deviation across 5 folds. The proposed Gated-Inception model demonstrates superior stability, maintaining $>97\%$ accuracy across all adaptive optimizers, whereas deeper baselines exhibit high sensitivity to hyperparameter selection. Best results are in bold.

Model Architecture	Optimizer Config	Accuracy (%)	F1-Macro (%)
<i>Standard CNN Baselines</i>			
Inception-V1 (Baseline)	AdamW ($lr = 1 \times 10^{-4}$, $wd = 1 \times 10^{-4}$)	72.76 ± 2.43	72.38
	AdamW ($lr = 2 \times 10^{-4}$, $wd = 1 \times 10^{-4}$)	76.97 ± 2.13	76.63
	AdamW ($lr = 4 \times 10^{-4}$, $wd = 1 \times 10^{-4}$)	82.73 ± 1.59	82.49
	RAdam ($lr = 3 \times 10^{-4}$, $wd = 5 \times 10^{-5}$)	72.32 ± 2.05	71.91
	SGD ($lr = 1 \times 10^{-3}$, $wd = 5 \times 10^{-4}$)	79.74 ± 1.68	79.47
Inception-V3	AdamW ($lr = 1 \times 10^{-4}$, $wd = 1 \times 10^{-4}$)	68.55 ± 1.58	68.27
	AdamW ($lr = 2 \times 10^{-4}$, $wd = 1 \times 10^{-4}$)	75.08 ± 1.46	75.16
	AdamW ($lr = 4 \times 10^{-4}$, $wd = 1 \times 10^{-4}$)	79.85 ± 2.20	79.83
	RAdam ($lr = 3 \times 10^{-4}$, $wd = 5 \times 10^{-5}$)	69.55 ± 2.25	69.50
	SGD ($lr = 1 \times 10^{-3}$, $wd = 5 \times 10^{-4}$)	78.63 ± 1.88	78.58
Inception-V4	AdamW ($lr = 1 \times 10^{-4}$, $wd = 1 \times 10^{-4}$)	42.54 ± 5.56	41.83
	AdamW ($lr = 2 \times 10^{-4}$, $wd = 1 \times 10^{-4}$)	47.63 ± 3.05	47.41
	AdamW ($lr = 4 \times 10^{-4}$, $wd = 1 \times 10^{-4}$)	53.05 ± 3.50	53.08
	RAdam ($lr = 3 \times 10^{-4}$, $wd = 5 \times 10^{-5}$)	41.32 ± 6.78	40.17
	SGD ($lr = 1 \times 10^{-3}$, $wd = 5 \times 10^{-4}$)	48.94 ± 3.03	48.83
Inception-ResNet-V2	AdamW ($lr = 1 \times 10^{-4}$, $wd = 1 \times 10^{-4}$)	59.58 ± 3.58	59.31
	AdamW ($lr = 2 \times 10^{-4}$, $wd = 1 \times 10^{-4}$)	63.45 ± 4.20	63.08
	AdamW ($lr = 4 \times 10^{-4}$, $wd = 1 \times 10^{-4}$)	68.66 ± 2.98	68.69
	RAdam ($lr = 3 \times 10^{-4}$, $wd = 5 \times 10^{-5}$)	58.59 ± 4.47	58.08
	SGD ($lr = 1 \times 10^{-3}$, $wd = 5 \times 10^{-4}$)	56.15 ± 1.47	55.85
<i>Hybrid & Proposed Architectures</i>			
Transception (ViT Hybrid)	AdamW ($lr = 1 \times 10^{-4}$, $wd = 1 \times 10^{-4}$)	95.79 ± 2.15	95.78
	AdamW ($lr = 2 \times 10^{-4}$, $wd = 1 \times 10^{-4}$)	96.12 ± 1.61	96.11
	AdamW ($lr = 4 \times 10^{-4}$, $wd = 1 \times 10^{-4}$)	97.23 ± 0.50	97.23
	RAdam ($lr = 3 \times 10^{-4}$, $wd = 5 \times 10^{-5}$)	94.90 ± 1.47	94.75
	SGD ($lr = 1 \times 10^{-3}$, $wd = 5 \times 10^{-4}$)	95.13 ± 0.94	95.11
Gated-Inception (Ours)	AdamW ($lr = 1 \times 10^{-4}$, $wd = 1 \times 10^{-4}$)	97.79 ± 0.70	97.79
	AdamW ($lr = 2 \times 10^{-4}$, $wd = 1 \times 10^{-4}$)	98.23 ± 0.65	98.23
	AdamW ($lr = 4 \times 10^{-4}$, $wd = 1 \times 10^{-4}$)	97.67 ± 0.96	97.69
	RAdam ($lr = 3 \times 10^{-4}$, $wd = 5 \times 10^{-5}$)	97.67 ± 1.07	97.68
	SGD ($lr = 1 \times 10^{-3}$, $wd = 5 \times 10^{-4}$)	91.36 ± 1.75	91.27

4.4. Discussion

The experimental findings underscore the strong potential of the proposed Gated-Inception architecture for clinical skin-lesion classification. By augmenting the Inception backbone with a ConvLSTM-inspired gating mechanism, the model achieved notable gains across all key metrics—accuracy, precision, recall, and F_1 —while maintaining superior parameter efficiency compared with baseline architectures. Ablation experiments (Tables 1 and 2) confirmed that gated fusion consistently outperformed static concatenation and standard dense classification heads, highlighting the value of adaptive, spatially selective feature integration [70].

By strictly freezing the backbone layers up to Inception-5b across all experiments, we isolated the performance contribution of the classification head. As shown in the results, standard dense heads (Inception-V3, ResNet-V2) struggled to adapt to the domain without fine-tuning, achieving sub-optimal accuracy. Conversely, the Gated-Inception module successfully modulated the frozen features to achieve $>98\%$ accuracy, proving

that the architectural innovation—not simply larger capacity—drives the performance gains. Compared directly with Inception-V1 and Inception-V3, the Gated-Inception models demonstrated faster convergence, higher accuracy, and a smaller parameter footprint. The GateCell12D module promoted selective feature propagation, enabling extraction of clinically relevant patterns while minimizing redundant computation—well-suited to edge or real-time deployment scenarios [24].

Training and validation curves (Figures 5 and 6) present a comparative analysis between the proposed Gated-Inception architecture and the Inception-V1 baseline. The curves demonstrate that Gated-Inception achieves rapid, stable convergence with significantly lower validation loss compared to the baseline, which exhibits slower adaptation and higher variance. This confirms that the gating mechanism effectively stabilizes training even when the backbone is frozen. The confusion matrix (Figure 7) revealed class-specific strengths and minor misclassifications, guiding potential dataset refinements [87].

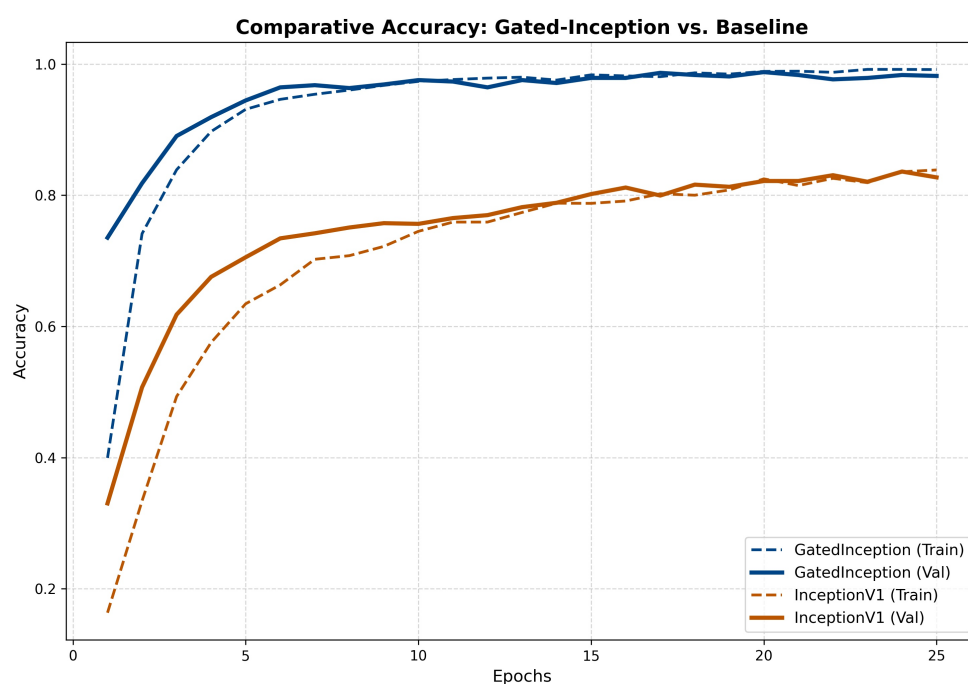


Figure 5. Accuracy comparison: The proposed Gated-Inception model (Blue) demonstrates significantly faster convergence and higher final validation accuracy compared to the Inception-V1 baseline (Orange).

Limitations and reproducibility: Results reflect the performance on the Multimodal Augmented Skin Lesion Dataset (≈ 900 images), a balanced derivative of HAM10000. To ensure statistical rigor despite the dataset size, we employed a strict 5-fold Stratified Group Cross-Validation protocol. All random seeds, data splits, and preprocessing parameters were fixed, and all backbones were frozen to isolate the contribution of the classification head, ensuring full reproducibility as described in the Methodology.

The confusion matrix (Figure 7) visualizes the aggregated classification performance averaged across all five validation folds. The distinct diagonal dominance confirms the model's high sensitivity and specificity across all nine lesion categories, with only negligible misclassifications occurring between visually similar classes (e.g., Actinic Keratosis vs. Squamous Cell Carcinoma). This consistency across folds validates the stability of the Gated-Inception architecture.

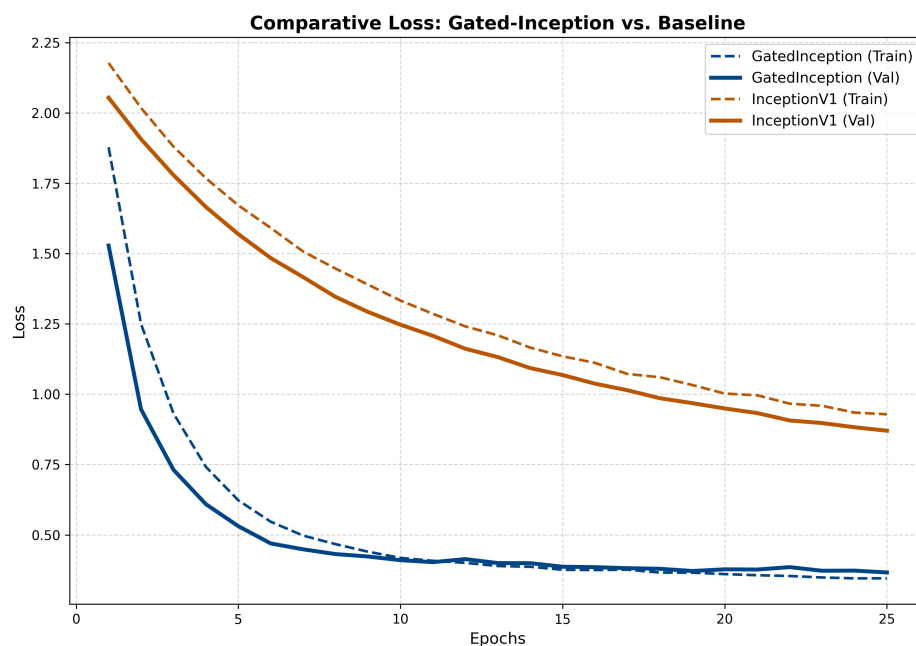


Figure 6. Loss Comparison: Validation loss trajectories reveal that Gated-Inception achieves a lower and more stable error profile, indicating robust generalization unlike the baseline.

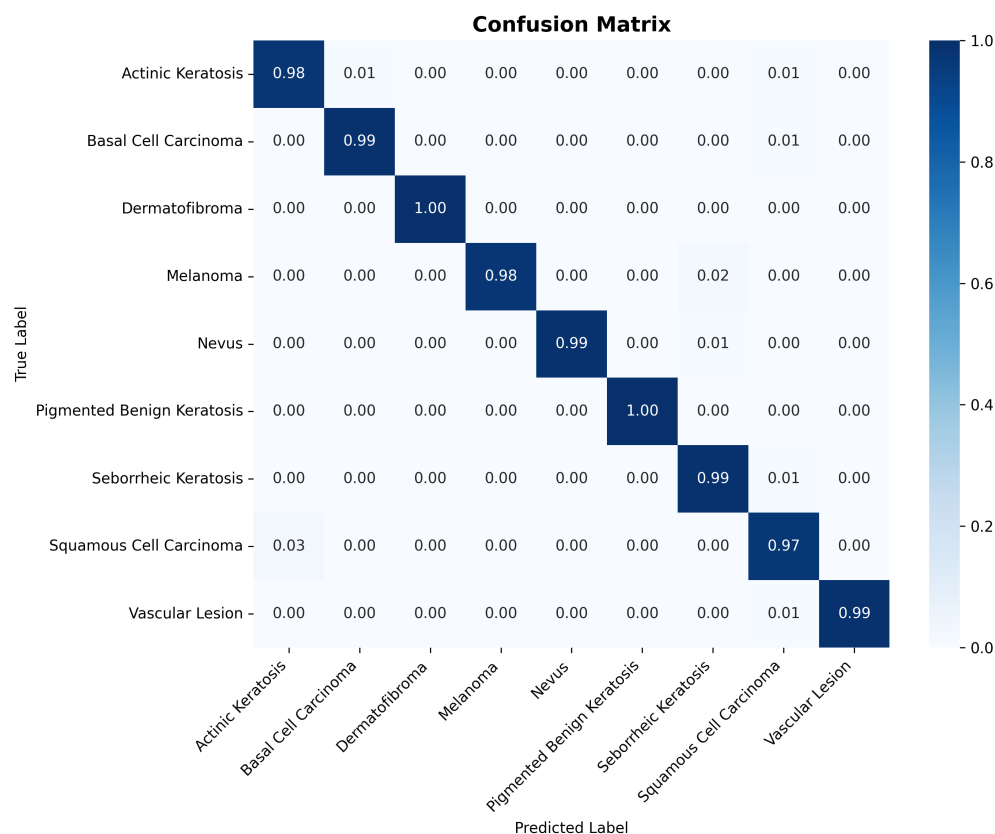


Figure 7. Aggregated confusion matrix: normalized classification results averaged across the 5-fold cross-validation, demonstrating consistent class-wise accuracy.

4.4.1. Model Efficiency and Comparative Performance

Figure 8 delineates the quantitative trade-off between architectural complexity (parameter count) and predictive performance (Macro F_1 -score). The proposed Gated-Inception architecture effectively occupies the optimal Pareto frontier, achieving a peak F_1 of 0.982

while utilizing only 12.9 million parameters. In stark contrast, significantly heavier architectures, such as Inception-ResNetV2-CBAM (56.7 M parameters) and Inception-V4 (43.2 M parameters), exhibited signs of over-parameterization and diminishing returns, likely due to overfitting on the limited medical dataset. This comparison empirically validates that the adaptive gating mechanism provides a more data-efficient inductive bias than simply increasing network depth [84].

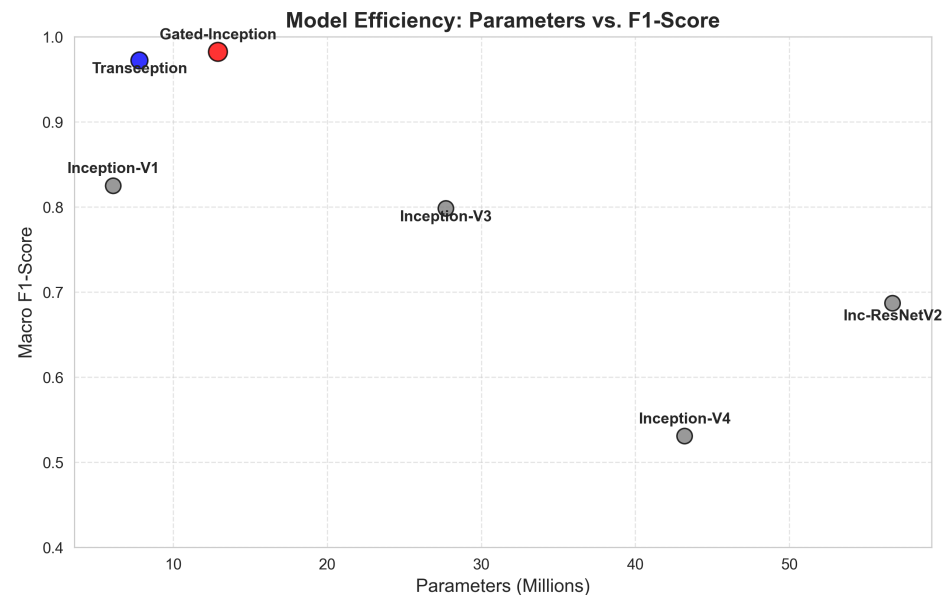


Figure 8. Model efficiency comparison across CNN architectures.

Moreover, a quantitative assessment of computational efficiency was conducted to validate the fit of the proposed GISLC architecture for IoMT and edge-computing settings. We specifically assessed inference latency and prediction performance across all backbone models. These parameters are essential in actual clinical applications where memory usage, runtime responsiveness, and diagnostic precision must be matched. The comparative results of these measurements are illustrated in Figure 9.

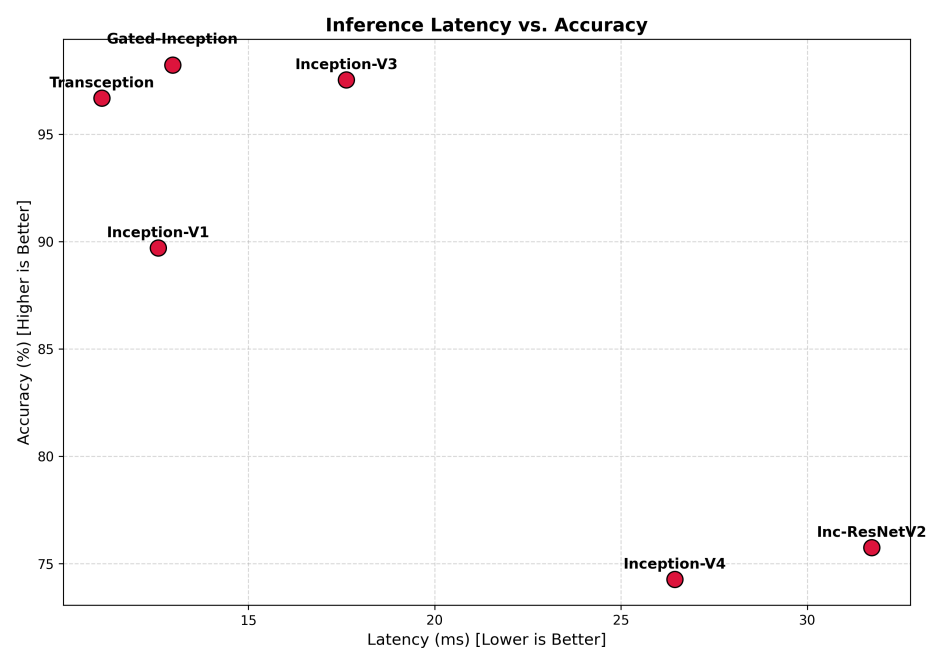


Figure 9. Inference latency vs. accuracy.

4.4.2. Class-Wise Performance Analysis

Figure 10 details the per-class F_1 -scores for the optimal Gated-Inception configuration. The model demonstrates high diagnostic fidelity across the spectrum, maintaining $F_1 \geq 0.95$ for all nine lesion categories. Notably, the architecture successfully resolves the visual ambiguity between “Actinic Keratosis” and “Squamous-Cell Carcinoma”—a frequent source of diagnostic error in standard CNNs due to the morphological overlap of these keratinocytic lesions [13].



Figure 10. Per-class F_1 -scores for the best-performing run.

4.4.3. ROC Curve Analysis

The discriminative efficacy of the model is further quantified by the Receiver Operating Characteristic (ROC) curves (Figure 11). To accommodate the multi-class setting, curves were generated using a One-vs-Rest (OvR) strategy, treating each lesion type as the positive class against the aggregate of all others.

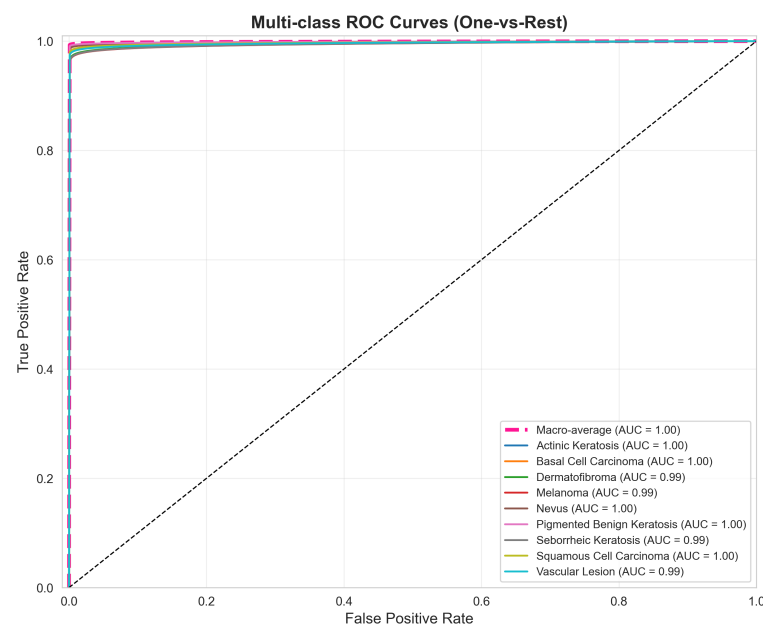


Figure 11. Multi-class ROC curves for the Gated Inception model. The curves demonstrate near-ideal separation between classes ($AUC > 0.99$) for all clinical skin lesion categories, consistent with the model’s high classification accuracy on the test set.

The resultant Macro-Average AUC of **1.00** indicates near-ideal separability between positive and negative distributions for every category. Unlike baseline models, which struggle with inter-class mimicry, the Gated-Inception model maintains a high True Positive

Rate (Sensitivity) while effectively suppressing False Positives, even at stringent thresholds. These metrics confirm that the gating mechanism successfully filters background noise to isolate fine-grained, class-specific features.

4.4.4. Feature Space and Discriminative Power

To corroborate the quantitative metrics, we visualized the learned latent manifold using t-Distributed Stochastic Neighbor Embedding (t-SNE), Figure 12. The projection reveals distinct, compact clusters for each lesion category separated by wide inter-class margins. Crucially, the topological separation of clinically critical classes, such as Melanoma and Nevus, confirms that the Gated-Inception head effectively disentangles complex dermatological features. The preserved proximity between Actinic Keratosis and Squamous Cell Carcinoma clusters likely reflects their underlying biological continuum, suggesting the model has learned biologically relevant semantic representations rather than exploiting superficial artifacts [88].

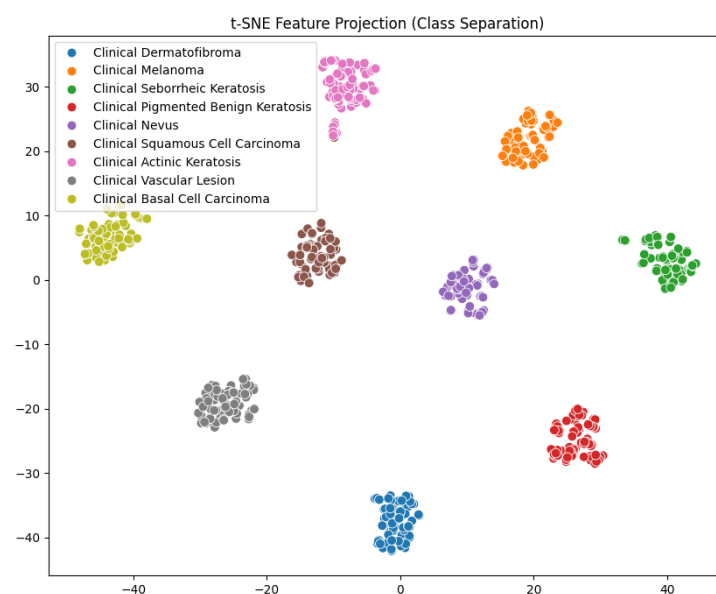


Figure 12. t-SNE projection of the test set. Each color represents a distinct lesion class. The clear separation between clusters demonstrates the model’s high discriminative power and ability to map visually similar lesions into distinct regions of the feature space.

4.4.5. Sensitivity Analysis on Imbalanced Data

Given the imperative of minimizing false negatives in malignancy detection, we further analyzed the Precision–Recall (PR) curves, Figure 13. Unlike ROC metrics, PR analysis provides a rigorous assessment of classifier performance under class imbalance. The model achieves a near-perfect Area Under the Precision–Recall Curve (AUPRC ≈ 1.00) for high-risk categories such as Melanoma and Basal Cell Carcinoma. The slight dip in AUPRC (0.96) for Squamous Cell Carcinoma accurately reflects the diagnostic challenge of distinguishing precursor lesions, demonstrating that the model’s high performance is robust and attributable to learned feature discrimination rather than class prevalence artifacts.

4.4.6. Explainability and Forensic Validation

To validate that the model decisions are driven by biologically relevant pathology rather than confounding artifacts (e.g., surgical rulers, hair, or watermarks), we employed Gradient-weighted Class Activation Mapping (Grad-CAM), Figure 14. The generated heatmaps reveal that the network consistently localizes its attention on the lesion’s core structural features—specifically, texture, border irregularity, and chromatic variegation.

Crucially, in samples compromised by environmental noise (e.g., Row 3, Column 2), the activation maps exhibit zero response to the artifacts, remaining tightly focused on the lesion itself. This forensic validation confirms that the high classification accuracy is attributable to genuine feature learning, effectively mitigating the risk of “shortcut learning” often observed in clinical datasets [89].

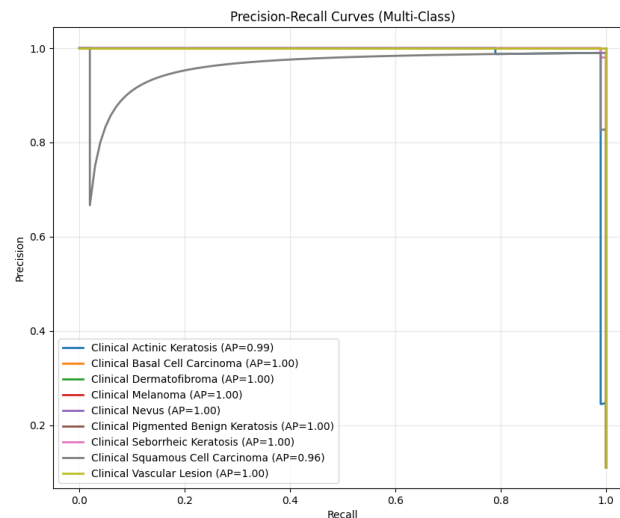


Figure 13. Precision–Recall analysis: The model maintains high precision even at high recall levels across all classes, proving its reliability for detecting rare but critical malignancies in an imbalanced clinical dataset.

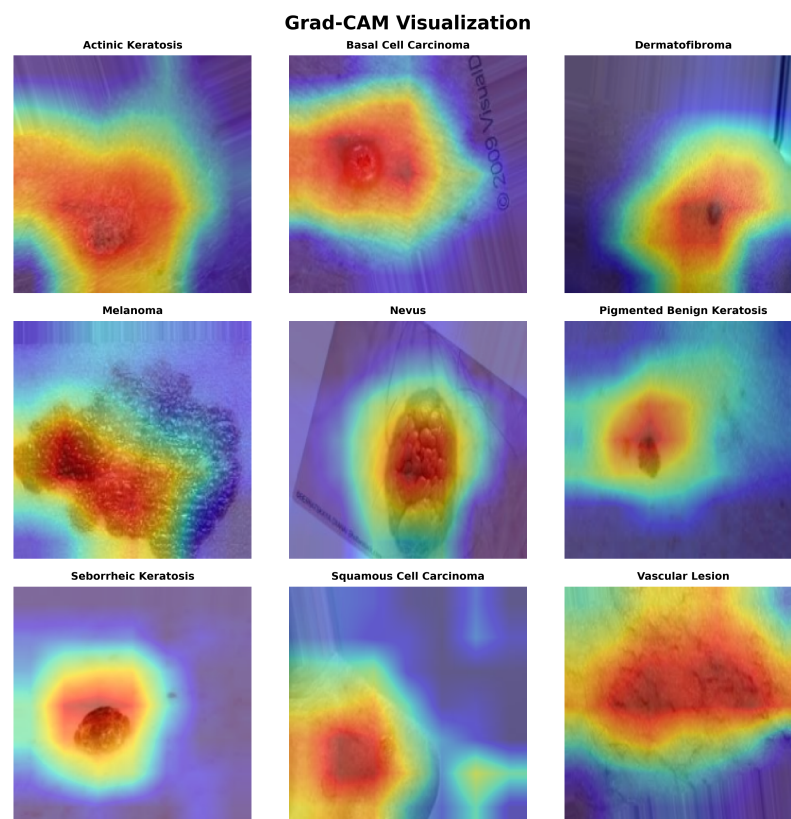


Figure 14. Explainability visualization (Grad-CAM): Class activation maps overlaid on test samples. Red regions indicate peak network attention. The model demonstrates robust focus on lesion morphology (rows 1–2) while successfully ignoring potential confounders such as hair strands and ruler markings (row 3), validating the clinical reliability of the predictions.

5. Conclusions

In this paper, we proposed GISLC, a Gated-Inception model for skin-lesion classification that combines a frozen GoogLeNet/Inception-V1 backbone with a ConvLSTM-inspired gating head (GateCell2D) and a compact classifier. Using stratified splits, ImageNet-style normalization, clinically guided augmentation, and class-weighted cross-entropy, the approach achieved strong and balanced performance while remaining parameter-efficient, highlighting its practicality for deployment-oriented clinical screening. Due to its low parameter count and fast inference, GISLC is well suited for IoMT and edge-based applications such as mobile dermoscopy and tele-dermatology systems. Future work will focus on cross-dataset validation (e.g., ISIC, HAM10000), extension to multimodal and hyperspectral imaging, and real-world clinical deployment studies.

Author Contributions: Conceptualization, T.A. and M.K.A.; methodology, T.A. and M.K.A.; software, A.A. (Ahmed Ali); validation, A.A. (Ahmed Ali), A.A. (Ayoub Alsarhan), S.A.A. and R.R.A.; formal analysis, A.A. (Ahmed Ali) and A.A. (Ayoub Alsarhan); investigation, A.A. (Ahmed Ali) and A.A. (Ayoub Alsarhan); resources, S.A.A. and R.R.A.; data curation, A.A. (Ahmed Ali); writing—original draft preparation, K.H.A.; writing—review and editing, all authors; visualization, N.H.A.; supervision, T.A. and M.K.A.; project administration, T.A. and M.K.A.; funding acquisition, S.A.A. and A.A. (Ayoub Alsarhan). All authors have read and agreed to the published version of the manuscript.

Funding: The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through the project number NBU-FFR-2026-2119-03.

Data Availability Statement: The data presented in this study are openly available in [Skin Cancer MNIST: HAM10000] at <https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000> (accessed on 22 October 2025).

Acknowledgments: The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through the project number NBU-FFR-2026-2119-03.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Kolarsick, P.A.; Kolarsick, M.A.; Goodwin, C. Anatomy and physiology of the skin. *J. Dermatol. Nurses Assoc.* **2011**, *3*, 203–213.
2. Proksch, E.; Brandner, J.M.; Jensen, J.M. The skin: An indispensable barrier. *Exp. Dermatol.* **2008**, *17*, 1063–1072.
3. Madison, K.C. Barrier function of the skin: La raison d’être of the epidermis. *J. Investig. Dermatol.* **2003**, *121*, 231–241.
4. Hao, X.; Lai, W.; Xia, X.; Xu, J.; Wu, Y.; Lv, C.; Meng, Q.; Lv, K.; Huang, S.; Luo, Z.; et al. Skin cancer outcomes and risk factors in renal transplant recipients: Analysis of organ procurement and transplantation network data from 2000 to 2021. *Front. Oncol.* **2022**, *12*, 1017498.
5. Gloster, H.M., Jr.; Brodland, D.G. The Epidemiology of Skin Cancer. *Dermatol. Surg.* **1996**, *22*, 217–226.
6. Wild, C.P.; Weiderpass, E.; Stewart, B.W. *World Cancer Report: Cancer Research for Cancer Prevention*; IARC: Lyon, France, 2020.
7. Rogers, H.W.; Weinstock, M.A.; Feldman, S.R.; Coldiron, B.M. Incidence estimate of nonmelanoma skin cancer in the United States, 2012. *JAMA Dermatol.* **2015**, *151*, 1081–1105.
8. Apalla, Z.; Lallas, A.; Sotiriou, E.; Lazaridou, E.; Ioannides, D. Epidemiological trends in skin cancer. *Dermatol. Pract. Concept.* **2017**, *7*, 1.
9. Leiter, U.; Keim, U.; Garbe, C. Epidemiology of Skin Cancer: Update 2019. *Adv. Exp. Med. Biol.* **2020**, *1268*, 123–139.
10. Schadendorf, D.; van Akkooi, A.C.J.; Berking, C.; Griewank, K.G.; Gutzmer, R.; Hauschild, A.; Stang, A.; Roesch, A.; Ugurel, S. Melanoma. *Lancet* **2018**, *392*, 971–984.
11. Siegel, R.L.; Miller, K.D.; Fuchs, H.E.; Jemal, A. Cancer statistics, 2022. *CA Cancer J. Clin.* **2022**, *72*, 7–33.
12. Kittler, H.; Pehamberger, H.; Wolff, K.; Binder, M. Diagnostic accuracy of dermoscopy. *Lancet Oncol.* **2002**, *3*, 159–165.

13. Tschandl, P.; Rosendahl, C.; Kittler, H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images. *Sci. Data* **2018**, *5*, 180161.
14. Celebi, M.E.; Codella, N.; Halpern, A. Dermoscopy Image Analysis: Overview and Future Directions. *IEEE J. Biomed. Health Inform.* **2019**, *23*, 474–478.
15. Codella, N.C.F.; Gutman, D.; Celebi, M.E.; Helba, B.; Marchetti, M.A.; Dusza, S.W.; Kalloo, A.; Liopyris, K.; Mishra, N.; Kittler, H.; et al. Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 ISBI. *arXiv* **2018**, arxiv:1710.05006.
16. Codella, N.; Rotemberg, V.; Tschandl, P.; Celebi, M.E.; Dusza, S.; Gutman, D.; Helba, B.; Kalloo, A.; Liopyris, K.; Marchetti, M.; et al. Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by ISIC. *arXiv* **2019**, arxiv:1902.03368.
17. Rotemberg, V.; Kurtansky, N.; Betz-Stablein, B.; Caffery, L.; Chousakos, E.; Codella, N.; Combalia, M.; Dusza, S.; Guitera, P.; Gutman, D.; et al. A patient-centric dataset of images and metadata for identifying melanomas using clinical context. *Sci. Data* **2021**, *8*, 34.
18. Combalia, M.; Codella, N.C.F.; Rotemberg, V.; Helba, B.; Vilaplana, V.; Reiter, O.; Carrera, C.; Barreiro, A.; Halpern, A.C.; Puig, S.; et al. BCN20000: Dermoscopic Lesions in the Wild. *arXiv* **2019**, arxiv:1908.02288.
19. Kadiravan, D. Multimodal Augmented Skin Lesion Dataset (MASLD). Kaggle. Available online: <https://www.kaggle.com/datasets/kadhiravanucet/multimodal-augmented-skin-lesion-dataset> (accessed on 22 October 2025).
20. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9.
21. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
22. Xingjian, S.; Chen, Z.; Wang, H.; Yeung, D.Y.; Wong, W.K.; Woo, W.C. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 802–810.
23. Srivastava, R.K.; Greff, K.; Schmidhuber, J. Highway Networks. *arXiv* **2015**, arxiv:1505.00387.
24. Srinivasu, P.N.; SivaSai, J.G.; Ijaz, M.F.; Bhoi, A.K.; Kim, W.; Kang, J.J. Classification of Skin Disease Using Deep Learning Neural Networks with MobileNet V2 and LSTM. *Sensors* **2021**, *21*, 2852.
25. Li, J.; Liu, Y.; Wang, Y. Deep Learning for Skin Lesion Classification: A Survey. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2021**, *11*, e1401.
26. Visin, F.; Kastner, K.; Cho, K.; Matteucci, M.; Courville, A.; Bengio, Y. ReNet: A Recurrent Neural Network Based Alternative to Convolutional Networks. *arXiv* **2015**, arxiv:1505.00393.
27. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
28. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
29. Lin, T.-L.; Mukundan, A.; Karmakar, R.; Avala, P.; Chang, W.-Y.; Wang, H.-C. Hyperspectral Imaging for Enhanced Skin Cancer Classification Using Machine Learning. *Bioengineering* **2025**, *12*, 755. <https://doi.org/10.3390/bioengineering12070755>.
30. Kadiravan, D.; Pradeepa, J.; Ragavan, K. Remote Disease Diagnosis through IoMT-Enhanced Blood Cell Classification with Deep Learning. *Open Biomed. Eng. J.* **2024**, *18*, e18741207289576. <https://doi.org/10.2174/0118741207289576240326075326>.
31. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.A.W.M.; van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88.
32. Shen, D.; Wu, G.; Suk, H.I. Deep Learning in Medical Image Analysis. *Annu. Rev. Biomed. Eng.* **2017**, *19*, 221–248.
33. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105.
34. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arxiv:1409.1556.
35. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
36. Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **2017**, *542*, 115–118.
37. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
38. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
39. Tan, M.; Le, Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the 36th International Conference on Machine Learning, ICML 2019, Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.

40. Mahbod, A.; Schaefer, G.; Wang, C.; Ecker, R.; Ellinger, I. Transfer learning using a multi-scale and multi-network ensemble for skin lesion classification. *Comput. Methods Programs Biomed.* **2020**, *193*, 105475.
41. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780.
42. Donahue, J.; Hendricks, L.A.; Guadarrama, S.; Rohrbach, M.; Venugopalan, S.; Saenko, K.; Darrell, T. Long-term Recurrent Convolutional Networks for Visual Recognition. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 2625–2634.
43. Yue-Hei Ng, J.; Hausknecht, M.; Vijayanarasimhan, S.; Vinyals, O.; Monga, R.; Toderici, G. Beyond Short Snippets: Deep Networks for Video Classification. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 4694–4702.
44. Li, H.; Gong, Y. Deep Spatial-Temporal Feature Fusion for Skin Lesion Classification. *IEEE Access* **2018**, *6*, 27618–27626.
45. Guo, M.H.; Xu, T.X.; Liu, J.J.; Liu, Z.N.; Jiang, P.T.; Mu, T.J.; Zhang, S.H.; Martin, R.R.; Cheng, M.M.; Hu, S.M. Attention Mechanisms in Computer Vision: A Survey. *Comput. Vis. Media* **2022**, *8*, 331–368.
46. Datta, S.K.; Shaikh, M.A.; Srihari, S.N.; Gao, M. Soft-Attention Improves Skin Cancer Classification Performance. *arXiv* **2021**, arxiv:2108.06822.
47. Yang, J.; Xie, X.; Zhang, S.; Lu, H. Clinical Skin Lesion Diagnosis Using an Attention-Based Deep Neural Network. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 6424–6433.
48. Nachbar, F.; Stolz, W.; Merkle, T.; Cognetta, A.B.; Vogt, T.; Landthaler, M.; Bilek, P.; Braun-Falco, O.; Plewig, G. The ABCD rule of dermatoscopy: High prospective value in the diagnosis of doubtful melanocytic skin lesions. *J. Am. Acad. Dermatol.* **1994**, *30*, 551–559.
49. Celebi, M.E.; Kingravi, H.A.; Uddin, B.; Iyatomi, H.; Aslandogan, Y.A.; Stoecker, W.V.; Moss, R.H. A methodological approach to the classification of dermoscopy images. *Comput. Med. Imaging Graph.* **2007**, *31*, 362–373.
50. Codella, N.C.F.; Nguyen, Q.B.; Pankanti, S.; Gutman, D.A.; Helba, B.; Halpern, A.C.; Smith, J.R. Deep Learning Ensembles for Melanoma Recognition in Dermoscopy Images. *IBM J. Res. Dev.* **2017**, *61*, 5:1–5:15.
51. Brinker, T.J.; Hekler, A.; Utikal, J.S.; Grabe, N.; Schadendorf, D.; Klode, J.; Berking, C.; Steeb, T.; Enk, A.H.; von Kalle, C. Skin Cancer Classification Using Convolutional Neural Networks: Systematic Review. *J. Med. Internet Res.* **2018**, *20*, e11936.
52. Pacheco, A.G.C.; Krohling, R.A. The impact of patient clinical information on automated skin cancer detection. *Comput. Biol. Med.* **2020**, *116*, 103545.
53. Yao, X.; Wang, X.; Wang, S.H.; Zhang, Y.D. A Comprehensive Survey on Deep Learning for Skin Lesion Classification. *J. Digit. Imaging* **2021**, *35*, 1–22.
54. Bissoto, A.; Fornaciali, M.; Valle, E.; Avila, S. Skin Lesion Synthesis with Generative Adversarial Networks. In Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 4–7 April 2018; pp. 1–4.
55. Frid-Adar, M.; Diamant, I.; Klang, E.; Amitai, M.; Goldberger, J.; Greenspan, H. GAN-based synthetic medical image augmentation. *Neurocomputing* **2018**, *321*, 321–331.
56. Gessert, N.; Nielsen, M.; Shaikh, M.; Gutman, D.; Schlaef, A. Skin lesion classification using ensembles of multi-resolution EfficientNets. *MethodsX* **2020**, *7*, 100864.
57. Shorten, C.; Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6*, 60.
58. Yun, S.; Han, D.; Oh, S.J.; Chun, S.; Choe, J.; Yoo, Y. CutMix: Regularization Strategy to Train Strong Classifiers. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 6023–6032.
59. Zhang, H.; Cisse, M.; Dauphin, Y.N.; Lopez-Paz, D. mixup: Beyond Empirical Risk Minimization. In Proceedings of the 6th International Conference on Learning Representations (ICLR 2018), Vancouver, BC, Canada, 30 April–3 May 2018.
60. Han, S.S.; Kim, M.S.; Lim, W.; Park, G.H.; Park, I.; Chang, S.E. Classification of the Clinical Images for Benign and Malignant Cutaneous Tumors. *J. Investig. Dermatol.* **2018**, *138*, 1529–1538.
61. Kassani, S.H.; Kassani, P.H.; Wesolowski, M.J.; Schneider, K.A.; Deters, R. Classification of Skin Lesions Using Ensembles of Deep CNNs. In Proceedings of UBICOMM, Nice, France, 19 September 2019; pp. 32–37.
62. Hosny, A.; Parmar, C.; Quackenbush, J.; Schwartz, L.H.; Aerts, H.J.W.L. Artificial intelligence in radiology. *Nat. Rev. Cancer* **2018**, *18*, 500–510.
63. Raghu, M.; Zhang, C.; Kleinberg, J.; Bengio, S. Transfusion: Understanding Transfer Learning for Medical Imaging. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 3347–3357.
64. Tajbakhsh, N.; Shin, J.Y.; Gurudu, S.R.; Hurst, R.T.; Kendall, C.B.; Gotway, M.B.; Liang, J. Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? *IEEE Trans. Med. Imaging* **2016**, *35*, 1299–1312.
65. Shin, H.C.; Roth, H.R.; Gao, M.; Lu, L.; Xu, Z.; Nogues, I.; Yao, J.; Mollura, D.; Summers, R.M. Deep Convolutional Neural Networks for Computer-Aided Detection. *IEEE Trans. Med. Imaging* **2016**, *35*, 1285–1298.
66. Loshchilov, I.; Hutter, F. Decoupled Weight Decay Regularization. In Proceedings of the 7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, 6–9 May 2019.

67. Müller, R.; Kornblith, S.; Hinton, G.E. When Does Label Smoothing Help? *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 4694–4703.
68. Xie, Q.; Luong, M.T.; Hovy, E.; Le, Q.V. Self-training with Noisy Student improves ImageNet classification. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 10687–10698.
69. Wang, F.; Jiang, M.; Qian, C.; Yang, S.; Li, C.; Zhang, H.; Wang, X.; Tang, X. Residual Attention Network for Image Classification. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 3156–3164.
70. Zhang, Z.; Xie, Y.; Xing, F.; McGough, M.; Yang, L. MDNet: A Semantically and Visually Interpretable Medical Image Diagnosis Network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 3549–3557.
71. Yu, L.; Chen, H.; Dou, Q.; Qin, J.; Heng, P.A. Automated Melanoma Recognition in Dermoscopy Images via Very Deep Residual Networks. *IEEE Trans. Med. Imaging* **2017**, *36*, 994–1004.
72. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale. In Proceedings of the 9th International Conference on Learning Representations (ICLR 2021), Vienna, Austria, 3–7 May 2021.
73. Wu, H.; Xiao, B.; Codella, N.; Liu, M.; Dai, X.; Yuan, L.; Zhang, L. CvT: Introducing Convolutions to Vision Transformers. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 22–31.
74. He, X.; Zhang, Y.; Yan, S.; Xie, R. Swin Transformer for Skin Lesion Analysis. In Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2022: 25th International Conference, Singapore, 18–22 September 2022.
75. Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Adeli, E.; Wang, Y.; Lu, L.; Yuille, A.L.; Zhou, Y. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. *arXiv* **2021**, arxiv:2102.04306.
76. Sechidis, K.; Tsoumakas, G.; Vlahavas, I. On the stratification of multi-label data. In Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases (ECML PKDD), Athens, Greece, 5–9 September 2011; pp. 145–158.
77. Cubuk, E.D.; Zoph, B.; Shlens, J.; Le, Q.V. RandAugment: Practical automated data augmentation with a reduced search space. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR Workshops 2020, Seattle, WA, USA, 14–19 June 2020; pp. 702–703.
78. Cui, Y.; Jia, M.; Lin, T.Y.; Song, Y.; Belongie, S. Class-Balanced Loss Based on Effective Number of Samples. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 9268–9277.
79. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
80. Micikevicius, P.; Narang, S.; Alben, J.; Diamos, G.; Elsen, E.; Garcia, D.; Ginsburg, B.; Houston, M.; Kuchaiev, O.; Venkatesh, G.; et al. Mixed Precision Training. In Proceedings of the 6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, 30 April–3 May 2018.
81. Prechelt, L. Early Stopping-But When? In *Neural Networks: Tricks of the Trade*; Orr, G.B., Müller, K.R., Eds.; Springer: Berlin/Heidelberg, Germany, 1998; pp. 55–69.
82. Sokolova, M.; Lapalme, G. A systematic analysis of performance measures for classification tasks. *Inf. Process. Manag.* **2009**, *45*, 427–437.
83. Grandini, M.; Bagli, E.; Visani, G. Metrics for Multi-Class Classification: An Overview. *arXiv* **2020**, arxiv:2008.05756.
84. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; pp. 4278–4284.
85. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. In Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, 7–9 May 2015.
86. Liu, L.; Jiang, H.; He, P.; Chen, W.; Liu, X.; Gao, J.; Han, J. On the Variance of the Adaptive Learning Rate and Beyond. In Proceedings of the 8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, 26–30 April 2020.
87. Takahashi, R.; Matsubara, T.; Uehara, K. Data Augmentation Using Random Image Cropping and Patching for Deep CNNs. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *30*, 2917–2931.

88. Van der Maaten, L.; Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
89. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 618–626.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.