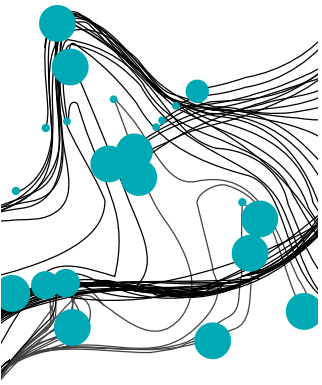
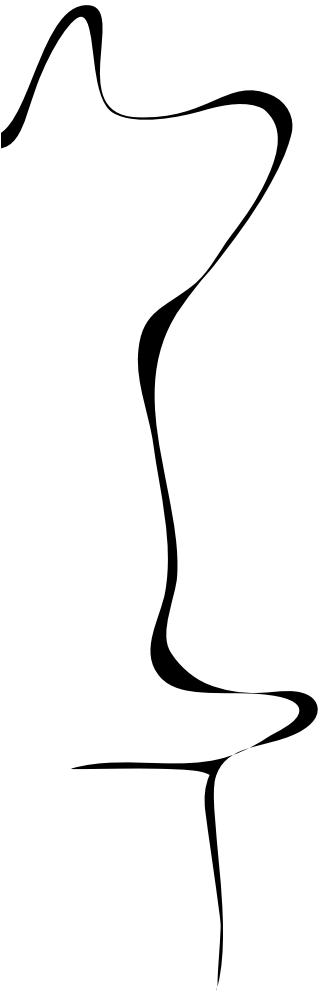


.6569

DMB

DATA MANAGEMENT AND BIOMETRICS



SYMMETRY ENHANCED SKIN LESION CLASSIFICATION NETWORK: A CASCADED MULTI-TASK LEARNING APPROACH

Akash Ramakrishnan

MASTER THESIS ASSIGNMENT

Committee:
dr. E. Talavera Martínez MSc
dr.ir. L.J. Spreeuwers
dr. D.V. Le Viet Duc

July, 2024

2024DMB0006
Data Management and Biometrics
EEMathCS
University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands

UNIVERSITY OF TWENTE. | DIGITAL SOCIETY INSTITUTE

Symmetry Enhanced Skin Lesion Classification Network: A Cascaded Multi-Task Learning Approach

Master's Thesis
University of Twente

Akash Ramakrishnan
University of Twente
Netherlands
a.ramakrishnan@student.utwente.nl

Abstract

Skin lesions represent a category of dermatological conditions where timely and accurate analysis is crucial for preventing malignancy. Assessing the symmetry of a lesion is a critical factor in determining its malignancy. This report presents a Multi-Task Learning (MTL) approach to skin lesion analysis that leverages lesion symmetry as a key feature for skin lesion classification. The study proposes a cascaded MTL architecture for comprehensive skin lesion analysis, incorporating three distinct tasks: skin lesion segmentation, lesion symmetry classification, and skin lesion classification. The proposed Symmetry Enhanced Lesion Classification Network (SE-LCN) leverages segmentation masks to refine the performance of two classification networks, enhancing the overall diagnostic accuracy. Class Activation Maps (CAMs) from the symmetry classification network are employed to augment the lesion classification network, aiming to provide more targeted and effective predictions. An extensive ablation study is conducted to analyze the impact of using various masking strategies on both classification networks and the efficiency of CAM transfer between the networks. This study not only analyzes the impact of these features on the predictive performance of these networks, but also the localization accuracy of the CAMs generated by these networks. The results demonstrate that integrating CAMs improves the predictive performance, and localization accuracy of the lesion classification network, validating the effectiveness of the proposed cascaded MTL architecture.

1 INTRODUCTION

Skin lesions are abnormal changes in the skin's appearance, texture, color, or structure, attributed to various factors such as acne, chickenpox, injuries, or allergies. However, lesions are also developed due to the presence of skin cancer, which is one of the most widespread diseases in the world (around 40% of all cancers) [1]. The most dangerous type of skin cancer is melanoma, developed from pigment-producing cells known as melanocytes [2]. One of the main causes of melanoma is exposure to ultraviolet light [3]. Despite being incurable, the early detection of melanoma can lead to successful interventions in up to 90% of the cases [4].

Dermoscopic imaging is a valuable tool in dermatology that is used in inspecting skin lesions. It is a technique that involves a handheld device called a dermatoscope which is equipped with specialized magnifying lenses and lighting systems. This facilitates an extensive analysis of the skin's surface for structures and patterns. Currently, the standard method of melanoma diagnosis is a visual analysis by a specialist [2]. This method, however, is cumbersome and often leads to incorrect analysis because of the complexity of skin lesions. Such kind of analysis is also subjective and can depend upon the experience of the specialist due to which only 84% of the visually inspected cases are accurate [5]. Consequently, there is a pressing need for computer-aided diagnosis (CAD) systems to enhance early-stage melanoma detection accuracy.

Various computer vision techniques have been employed for skin lesion analysis and melanoma detection. One crucial step in skin lesion analysis is segmentation, an efficient technique to delineate the foreground from the background, yielding a region of interest [6]. This helps diagnostic systems focus more on the interior of the lesion and improves the robustness of classification. Recently, with the research advancements in the field of deep learning, many Deep Convolutional Neural Network (DCNN) [7] based approaches have been implemented for skin lesion segmentation [8].

The segmented images also feed into rule-based diagnostic systems which estimate features like asymmetry, border irregularity, lesion size, and color or texture smoothness to detect melanoma. One of the most popular rule-based algorithms is the ABCD algorithm (Asymmetry, Border, Color, Diameter) [9] which is based on the lesion. This algorithm was later improved by ABCDE (ABCD and Evolution of lesion area) [10] and ABCDEF (ABCDE and the "ugly duckling" sign) [11]. Similarly, the 7-point checklist [12] and the Menzies method [13] are popular rule-based algorithms to classify lesion malignancy. Recently, state-of-the-art results on lesion classification have been achieved through deep learning models. DCNNs excel in extracting complex spatial features and pixel patterns that outperform traditional feature-based classification methods.

Lesion symmetry is considered a crucial feature while assessing the malignancy of a lesion, especially in rule-based diagnostic systems. The classification of symmetry, however, can be altered by the judgment of the individual inspecting the lesion [14]. Recently, there has been some research on the automatic classification of symmetry of skin lesions, using classical methods [15] and deep learning approaches [16, 17]. These methods classify a lesion into three categories – fully asymmetric, symmetric through one axis, and symmetric through both axes.

Most deep learning models are typically trained for specific tasks, such as segmentation or classification, each requiring separate models. However, in the context of skin lesion analysis, segmentation and classification are intricately related. Segmentation helps remove distractions (like hair and other artifacts) from the background for better feature extraction and classification by the model. Classification, on the other hand, extracts useful class-specific features that could enhance segmentation accuracy [18]. The approach of training a network to do multiple tasks is called Multi-Task Learning (MTL) [19] offering a compelling framework capable of leveraging information from various tasks relevant to skin lesion analysis. In 2020, Xie et al. [18] proposed a mutual bootstrapping approach utilizing a coarse segmentation network to generate a mask, which was subsequently used to train a mask-guided classification network. The Class Activation Maps (CAMs) from this network were then used as features for an enhanced segmentation network, refining the segmentation masks. While exploring the impact of symmetry classification in an MTL framework, Talavera-Martinez et al. [17] mention that the model is coherent while classifying malignancy and symmetry. 72.5% of the "malignant" lesions are classified as "asymmetric", and 93.4% of the "benign" lesions are classified as "symmetric". This means that the region of neurons activated during symmetry classification could be useful in determining the malignancy of a lesion. Hence, the CAMs extracted from a symmetry classification model could aid in classifying the type of skin lesion. The goal of this research is to address the following research question:

(1) How does the inclusion of Class Activation Maps (CAMs) from the symmetry classification network impact the predictive performance of the skin lesion classification network in a Multi-Task Learning (MTL) framework?
(2) How does the inclusion of Class Activation Maps (CAMs) from the symmetry classification network impact the localization accuracy of CAMs generated from the skin lesion classification network in a Multi-Task Learning (MTL) framework?

We also aim to answer the following subquestions:

(1) How does using skin lesion segmentation to mask the inputs impact the predictive performance of symmetry classification and skin lesion classification networks?
(2) How does using skin lesion segmentation to mask the inputs improve the localization accuracy of CAMs generated from the symmetry classification and skin lesion classification networks?

To address these questions, we propose the Symmetry Enhanced Lesion Classification Network (SE-LCN), a cascaded MTL architecture that includes three tasks: skin lesion segmentation, symmetry classification, and skin lesion classification. In the implemented MTL architecture, Class Activation Maps (CAMs) from the symmetry classification encoder are used to enhance the performance of classifying dermoscopic skin lesion images into 7 classes provided in the HAM10000 dataset [20]. To merge the CAMs with features of the lesion classification network we use the E-Layer [18]. Additionally, a segmentation network is trained to mask the images that are used as inputs for both classification networks. An ablation study is conducted to compare the impact of two distinct masking strategies on both classification networks. The evaluation metrics focus on the predictive performance of the networks, and the localization accuracy of the CAMs, providing a comprehensive overview of the proposed methods. The major contributions of this work are as follows:

(1) Introduction of a novel MTL architecture that utilizes CAMs from the symmetry classification network to enhance skin lesion classification.
(2) Analysis of the impact of two masking strategies on both symmetry classification and skin lesion classification.
(3) Introduction of a novel method of utilizing segmentation masks to measure the localization of CAMs using three metrics - Intersection over Union, Percentage Overlap, and Pointing Game score.

The report structure entails a section for scientific background (Section 2), containing the technical background of all the relevant terminologies, which is followed by a section that discusses related works (Section 3) in skin lesion analysis and multi-task learning approaches in this field. This section also provides a detailed overview of the available datasets for skin lesion analysis. A methodology section follows outlining the proposed approach (Section 4). Next, the experimental framework for this research is presented (Section 5), including the results obtained. Subsequently, this report contains a detailed discussion analyzing the findings (Section 6), after which the report concludes by outlining the conclusions and future scope of this research (Section 7).

2 SCIENTIFIC BACKGROUND

This section presents a detailed description of the technical background of the methodologies used in this work.

2.1 Multi-Task Learning

Multi-Task learning (MTL) is an approach that was first introduced in the field of machine learning in 1992 [19]. The paper states that it is an approach of inductive transfer that improves generalization by using the domain information contained in the training signals of related tasks as an inductive bias.

There are several advantages of employing an MTL architecture [21]:

- **Implicit Data Augmentation:** Training a common model on multiple tasks with different noise patterns brings robustness to various types of noise, reducing the chances of overfitting.
- **Attention Focusing within Tasks:** Auxillary tasks can provide evidence for the relevance of specific features.
- **Cross-Task Feature Learning:** MTL helps certain tasks to eavesdrop and learn useful features that are easier to learn from other tasks.
- **Representational Bias:** MTL encourages the learning of representations favored by all tasks, thus promoting generalization.

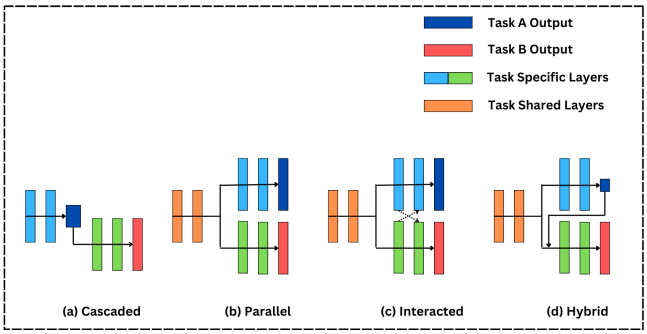


Figure 1: The four types of multi-task deep learning architectures include (a) cascaded, (b) parallel (c) interacted, and (d) hybrid.

.

- **Regularization** MTL acts as a form of regularization, preventing model overfitting.

MTL architectures can be broadly classified into 4 types [22]:

- **Cascaded Architecture:** Here, the output of the previous task is used as the input to the subsequent task, with no shared layers between tasks.
- **Parallel Architecture:** This architecture has a common input, common layers, and separate task-specific layers. Typically, parallel architectures are trained end-to-end.
- **Interacted Architecture:** This is similar to parallel architecture, but the task-specific layers share information within themselves, often facilitated by skip connections or attention mechanisms.
- **Hybrid Architecture:** Combining elements of both parallel and cascaded designs, this architecture incorporates shared layers, with outputs from one task-specific layer feeding into another.

These architectures are illustrated in Figure 1. Previous works employing MTL in skin lesion analysis are detailed in Section 3.

One important aspect of training an MTL network is that, when trained end-to-end, different loss functions are optimized at the same time. These loss functions may be of the same type or different types. For instance, in the case of an MTL architecture that is trained for segmentation and classification, the segmentation network may use a dice loss and the classification network may use a cross-entropy loss. Since different tasks (whether of the same or different types) have different learning curves, it is essential to strategize the optimization of each loss function. However, this is not the case for a non-end-to-end architecture since each model is trained separately for optimal performance.

In a vanilla end-to-end MTL, the loss values of each task are added to get a scalar value of the overall loss. In this case, each task is considered equal throughout the training. However, this is not the best strategy. According to [23], two common loss weighting strategies are generally used in MTL frameworks:

- **Dynamic Weighting Strategy:** This involves adjusting the weights assigned to each task's loss value during training based on the gradients of each task. Tasks exhibiting steeper gradients are given greater importance for learning.
- **Uncertainty-Based Weighting:** Tasks with higher uncertainty in predictions are prioritized, allocating more weight to such tasks during training

In architectures that have tasks with different output types, a multi-phase training strategy is often adopted. This involves training tasks with similar output tasks together, followed by joint fine-tuning of all tasks.

2.2 Class Activation Maps

Class Activation Maps (CAMs) were first introduced in [24] by a team of researchers at MIT. This method is used to analyze the regions of interest within an image that a network focuses on while predicting the image's class. CAMs are visualizations that highlight the discriminative image regions, typically extracted from the final convolutional layer, just before the final output layer. Global average pooling is applied to compute the spatial average of each unit's feature map at this layer, resulting in a single value per feature map. The outputs of the global average pooling are used as features for a fully connected layer that produces the final output.

To generate CAMs, the weights of the output layer are projected back to the convolutional feature maps to highlight the importance of different regions for final classification. The resulting weighted combination is then upsampled to the image size for accurately visualizing the CAMs.

Several enhancements to the original CAM method have been developed to improve localization accuracy. One such development is Grad-CAM [25], which utilizes the gradients of the target class flowing into the final convolutional layer instead of using the weights as used in conventional CAM. Building on this, Grad-CAM++ [26] was introduced which extends Grad-CAM by incorporating both positive and negative gradients. Furthermore, Score-CAM [27] was introduced by Wang et al. (2020) which assigns importance scores to each pixel based on their contribution to the final output prediction score, utilizing the entire model rather than just the output layer. Finally, Ablation-CAM [28] explores a different angle by identifying significant regions through the iterative removal of image portions, subsequently observing the effects on prediction confidence.

CAMs are extremely useful for qualitative evaluation of a model. Visually inspecting the CAMs of a model gives a good idea of what region the model focuses on to make the predictions. However, to measure and compare the localization of CAMs among multiple models, it is important to have ground truth information about the location of objects within an image. This information is generally present in datasets with either bounding box annotations or with segmentation masks. Given the region information, here are a few commonly used metrics to evaluate the localization of CAMs:

- **Intersection over Union (IoU):** This metric was first used in the original paper of CAMs [24] and measures the overlap

between regions highlighted by CAM and the ground truth bounding box.

- **Pointing Game:** Introduced in [29], it is a metric to measure the percentage of images in a dataset where the highest activation point from the activation map lies inside the class region. If the highest activation point of the image lies in within the class region, it is considered a hit, else a miss. The point game is then calculated as follows:

$$\text{Pointing Game} = \frac{\text{Number of Hits}}{\text{Number of Hits} + \text{Number of Miss}}. \quad (1)$$

- **Weakly Supervised Localization**: This method of evaluation, also introduced in [24], involves treating the CAM outputs as an object detection problem. A bounding box is drawn around the largest connected region in the CAM heatmap, and compared with the ground truth bounding box of the image.

3 RELATED WORKS

This section provides a comprehensive overview of related studies in skin lesion analysis, categorized into subsections discussing available skin dataset, skin lesion segmentation, skin lesion classification, symmetry classification, and Multi-Task Learning (MTL) approaches.

3.1 Skin Datasets

The availability of larger annotated datasets has played a crucial role in improving the accuracy of skin lesion analysis over the years (as shown in Table 2). One of the pioneering datasets in this domain is the Interactive Atlas of Dermoscopy [12], popularized through the works of Celebi et al. [30]. Initially released on CD-ROM, this dataset comprises 1,039 dermoscopic images of 1,024x683 pixels, categorized into the classes melanomas, carcinomas, and nevus. Subsequently, the PH^2 dataset [31] emerged as the first publicly available dataset that contained segmentation masks comprising 200 dermoscopic images of 768x560 pixels from three classes - melanoma, atypical nevus, and common nevus.

Another notable release in the same year was the DermoFit Image Library [32], a collection of 1,300 clinical images captured in a controlled environment. Clinical images are captured using typical cameras and are different from dermoscopic images which are captured using a dermatoscope. This dataset has images from 10 categories of skin lesions of varying sizes. The 10 categories include actinic keratosis, basal cell carcinoma, melanocytic nevus, seborrhoeic keratosis, squamous cell carcinoma, intraepithelial carcinoma, pyogenic granuloma, haemangioma, dermatofibroma, malignant melanoma.

Currently, the most extensive collection of labeled dermoscopic images is the ISIC (International Skin Imaging Dataset) archive. Every year, the ISIC challenge [33–35] is conducted where a new dataset is publicly released for researchers to develop state-of-the-art results on segmentation and classification. The details of these datasets can be seen in Table 2. The ISIC 2016, 2017, and 2018 challenges comprised segmentation, feature extraction, and lesion classification tasks, whereas the ISIC 2019 and 2020 only comprised lesion classification. The type of lesion classification also varied throughout the years. The ISIC 2018 contained separate datasets for each task. While the challenge provided 3,694 images for segmentation and feature extraction, the HAM10000 [20] dataset was provided for classification. This is a large dataset consisting of 10,015 training images which was acquired over 20 years at the Medical University of Vienna. This dataset consisted of lesion images belonging to 7 classes namely melanoma, melanocytic nevus, basal cell carcinoma, actinic keratosis, benign keratosis, dermatofibroma, and vascular lesion. As part of a study [36] in 2020, the segmentation masks of the HAM10000 dataset were created and publicly released. This is the largest publicly available segmentation dataset. The ISIC 2019 and ISIC 2020 have much higher amounts of training images (25,331 and 33,126 respectively) but do not have segmentation masks like their predecessors. Both datasets have 8 classes but are highly imbalanced as shown in Table 2. Recently, an in-depth analysis was conducted on all the ISIC datasets (2016-2020) [37] where the authors found both overlaps among the datasets and duplicates within each dataset. The authors created a new dataset without the duplicates that consisted of 45,590 training images. The authors also released a balanced dataset (balance between melanoma and non-melanoma) with 7,848 images.

Table 1: Class distribution of SymDerm v2.0 along with data source distribution.

	Fully asymmetric	Symmetric to 1 axis	Symmetric to 2 axis
dermis	42	12	14
dermquest	89	18	27
EDRA	52	11	14
PH^2	52	31	117
HAM10000	936	505	559
ISIC2018	57	56	63
Total	1228	633	794

To classify the symmetry of a lesion, Talavera-Martinez et al. [16] created a dataset of labels called SymDerm following the taxonomy present in the PH^2 dataset [31]. As mentioned above, the PH^2 dataset consists of 200 images and also contains labels for the ABCD rule [10], hence also consists of labels for symmetry. This dataset was extended by first introducing simulated hair to the images which was done in [38]. This increased the dataset size from 200 to 438. The dataset was further extended by randomly selecting images from publicly available datasets and asking three expert dermatologists to label them. The images were classified into three categories: asymmetric, symmetric with respect to one axis, and symmetric with respect to two axes. Overall, the final dataset consists of 1,052 annotated images based on symmetry. In later works [17], the authors extended the SymDerm dataset and created the SymDerm v2.0 dataset which consists of around 2,000 new annotations. This dataset contains a total of 2,665 annotations which was curated using the same procedure used for the original SymDerm dataset. The class distribution of these datasets and the source of data can be seen in Table 1.

Table 2: An overview of datasets available for skin lesion segmentation and classification.

Year	Dataset	Number of Images	Modality	Tasks	Classes	Train/Val/Test	Distribution
2000	Interactive Atlas of Dermoscopy [39]	1,039	dermoscopy	Classification	3	-	Melanoma (26%) Carcinomas (4%) Nevi (4%)
2013	DermoFit [32]	1,300	clinical	Classification	10	-	Actinic keratosis (3.5%) Basal cell carcinoma (18.34%) Melanocytic nevus (35.46%) Seborrhoeic keratosis (19.76%) Squamous cell carcinoma (6.76%) Intraepithelial carcinoma(6%) Pyogenic granuloma (1.8%) Haemangioma (7.46%) Dermatofibroma (5%) Malignant melanoma (5.84%)
2013	Pedro Hispano Hospital (PH2) [31]	200	dermoscopy	Classification Segmentation	3	-	Melanomas (20%) Atypical Nevi (40%) Common Nevi (40%)
2016	ISIC 2016 [33]	1,279	dermoscopy	Classification Segmentation	2	900/-/379	Melanomas (42.6%) Benign (57.4%)
2017	ISIC 2017 [34]	2,750	dermoscopy	Classification Segmentation	3	2,000/150/600	Melanomas (18.7%) Seborrheic keratoses (12.7%) Nevi (68.6%)
2018	ISIC 2018 [35]	3,694	dermoscopy	Classification Segmentation	7	2, 594/100/1, 000	Melanoma (11.1%) Nevus (66.9%) Basal cell carcinoma (5.1%) Bowen's disease (3.3%) Benign keratosis (11%) Dermatofibroma (1.1%) Vascular lesion (1.4%)
2020	HAM10000 [20]	10,015	dermoscopy	Classification Segmentation	7	8,011/1,002/1,002	Melanoma (11.1%) Nevus (66.9%) Basal cell carcinoma (5.1%) Bowen's disease (3.3%) Benign keratosis (11%) Dermatofibroma (1.1%) Vascular lesion (1.4%)
2019	ISIC 2019	33,569	dermoscopy	Classification	8	25, 331/-/8,232	Melanoma (18%) Nevus (51%) Basal cell carcinoma (13%) Actinic keratosis (3.5%) Benign keratosis (10%)the Dermatofibroma (1%) Vascular lesion (1%) Squamous cell carcinomas (2.5%)
2020	ISIC 2020 [40]	44,108	dermoscopy	Classification	8	33,126/-/10,982	Melanomas (1.8%) Nevi (97.6%) Seborrheic keratoses (0.4%) Lentigines simplex (0.1%) Lichenoid keratoses (0.1%) Solar lentigines (0.02%) Café-au-lait macules (0.003%) Atypical melanocytic (0.003%)
2022	Combined ISIC with no duplicates [37]	56,987	dermoscopy	Classification	2	45,590/-/11,397	Melanoma (50%) Others (50%)
2022	Combined ISIC dataset balanced [37]	9,810	dermoscopy	Classification	2	7,848/-/1,962	Melanoma (50%) Others (50%)

3.2 Skin Lesion Segmentation

Semantic segmentation is a fundamental technique in computer vision that involves partitioning an image into multiple segments, each representing a distinct class. It is a pixel-level classification task that requires both translation-invariant global features for accurate categorization, and localized features for delineating specific regions. The use of CNNs for segmentation tasks began in the mid-2000s [41], but gained prominence after the paper on Fully Connected Networks (FCNs) [42] in 2015. FCNs, along with U-Net [43], became the basis of many state-of-the-art segmentation models.

Segmentation architectures for skin lesion analysis can be broadly categorized into four groups [8]: single-network models, multiple-network models, hybrid feature models, and transformer models. While the first two categories are self-explanatory, hybrid feature models integrate deep-learning-based features with manually selected features to enhance performance. Transformer models, on the other hand, utilize the transformer architecture [44].

Early works on skin lesion segmentation utilized single-network models that often adopted either FCN [42] or U-Net [43]. The U-Net architecture has been widely adopted for skin lesion segmentation, both in its original form [45, 46] and in modified forms [47, 48]. Both these modified U-Net works [47, 48] incorporate depth-wise separable convolutions to reduce the number of parameters. Similarly, recurrent CNN variations of U-Net were also introduced in RU-Net [49] and R2U-Net [50]. Another network family that has been frequently used to obtain state-of-the-art segmentation performance is DeepLab [51] and its variants DeepLabv2 [51], DeepLab v3 [52], and DeepLabv3+ [53]. These networks utilize atrous convolution, allowing flexible aggregation of contextual information effectively without compromising image resolution. For skin lesion segmentation, numerous works have adopted the DeepLab family architectures [54–56].

Multiple-network models could be further subdivided [8] into multi-task learning models (discussed in Section 3.5) and ensemble models. Ensemble networks, popular in machine learning, often improve stability and predictive performance by leveraging the complementary strengths of various models [57]. These models can be implemented via early fusion at the feature level or late fusion at the prediction level. In skin lesion segmentation, ensemble learning typically involves applying different configurations to a single deep-learning architecture. Variations in the ensemble models may include adjusting network hyperparameters like the number of filters per block and their size [45], optimization technique [58], training splits [59, 60], different color spaces [61], and different initialization techniques [55, 62].

3.3 Skin Lesion Classification

Numerous studies have led to advancements in the field of skin lesion classification. A dynamic graph cut algorithm combined with a Naive Bayes classifier was utilized for robust segmentation and classification [63] on the ISIC 2017 dataset [34]. A complete analysis [64] was conducted on the use of EfficientNets [65] for skin lesion classification, comparing performance across all versions (B0 to B7) and suggesting optimal hyperparameters; EfficientNet B4 was found to yield the best performance on the HAM10000 dataset [20]. A model that combines MobileNet V2 and LSTM, demonstrating superior performance over traditional single network architectures, was developed in [66].

Enhancements in multi-type skin disease classification were achieved using an Optimal Path Deep Neural Network (OP-DNN) that improved feature extraction, significantly enhancing classification accuracy [67]. A classification model that integrates DenseNet and ConvNeXt, benchmarking its performance across multiple datasets, was also developed [78]. Other significant works include [68], incorporating multi-scale attention blocks, and improving the specificity. An enhanced deep bottleneck transformer model was suggested by [69] to improve the performance of classification.

Further innovations include [70], which presented a hybrid model merging various deep learning models with traditional classifiers. An approach of optimized region-growing segmentation and an autoencoder-based classification model that accurately identified affected skin regions was developed by [71]. A model that used Wiener-filtering preprocessing and a whale optimization algorithm was introduced in [72]. This traditional technique showed promising results compared to more recent techniques. A study [73] trained 23 CNNs on a collection of skin lesion images for diagnosing Lyme disease using transfer learning and offered model selection advice, in which they concluded that ResNet50 gave the best performance. The authors also examined the explainability of the models using a qualitative evaluation of the Grad-CAMs. The performance comparison of the recent works in skin lesion classification on the HAM10000 [20] dataset is summarized in Table 3.

3.4 Symmetry Classification

As mentioned in Section 1, the symmetry of a lesion is one of the key features used in rule-based clinical criteria for skin disease identification like the ABCD rule [10], and the 7-point checklist [84]. Previous works on symmetry detection for skin lesions predominantly used classical image processing techniques. These techniques include analyzing the irregularity of color distributions [85], using Fourier descriptors to quantify the asymmetry of shape [86], and computing the non-overlapping areas of a lesion by folding it across an axis of symmetry [87]. However, these methods use symmetry classification as one of the features for classifying lesions without explicit symmetry evaluation.

In 2020, [15] proposed novel methods to classify symmetry based on shape and texture. In this work, symmetry based on shape is assessed by parameterizing the candidates to be axes of symmetry by passing lines through the center of mass and using the Jaccard index to find the similarity between the two sets (in this case, lesion regions on both sides of the axis). A random forest classifier was used to classify the symmetry of the lesion into "1-axis symmetry", "2-axis symmetry" and "Asymmetry". Symmetry based on texture is based on determining the similarity of different lesion patches

Table 3: Summary of recent work on skin lesion classification on the HAM10000 dataset. The table provides a comparison of numerous works on the same dataset using six metrics: Accuracy, Sensitivity, Specificity, Recall, Precision, and F1 Score.

Method	Accuracy	Sensitivity	Specificity	Recall	Precision	F1Score
EfficientNet B4 [64]	87.7	88	88	88	88	87
DLNN + MobileNet V2 + LSTM [66]	90.21	90.24	95.1	92.24	-	-
ML + CNN [74]	95.18	94	-	85	88	86
DenseNet + ConvNext [75]	95.29	92.58	-	92.58	88.35	89.99
DenseNet 201 + ML [70]	99.94	91.48	98.82	91.48	97.01	-
S2C-DeLeNet [76]	91.03	90.58	90.58	90.58	90.38	90.48
EDBTM [69]	95.84	-	-	-	96.1	-
HSBSO-LSTM [77]	93.8	93.9	93.8	-	33.9	49.8
WT-DRNNet [78]	95.36	-	98.62	-	95.59	93.37
DBN-MEFOA [79]	97.99	92.99	97.00	-	96.99	91.99
ESVMKRF-HEAO [80]	97.4	95.9	96	-	96.3	97.4
DenseNet169 [81]	92.25	93.59	-	-	-	93.27
CNN-layered BLSTM [82]	89.47	88.33	97.17	-	-	-
MC-SVM [83]	98.57	93.89	96.37	-	-	94.98

across an axis. This method uses two random forest classifiers to assess the symmetry of two 32x32 patches and to aggregate information across different orientations.

Additionally, a few DCNN-based methods for classifying symmetry have emerged. For instance, [16] which is the first of such methods proposed a simple CNN architecture with three blocks to classify the symmetry of a lesion. This study also introduced the SymDerm dataset, which is a collection of publicly available dermoscopic images that have been labeled for symmetry by three expert dermatologists. The details of this dataset are described in Section 3.1. The proposed method outperforms the traditional symmetry classification techniques mentioned previously. SymDerm v2.0, an extension of the SymDerm dataset was introduced in [17], along with an MTL architecture performing symmetry and malignancy classification. Furthermore, the study analyses the effect of different augmentation techniques on symmetry classification. The study concludes that although symmetry classification does not show any improvement in performance measures by including the task of skin lesion classification, it does benefit from the auxiliary task of classifying a lesion as malignant or benign. The comparison of existing methods on the SymDerm and SymDerm v2.0 datasets is summarized in Table 4.

3.5 Multi-Task Learning

Based on the training strategy, multi-task learning can be broadly classified into two types: end-to-end and non-end-to-end. As discussed in Section 2, most cascade architectures are trained non-end-to-end, and most parallel architectures are trained end-to-end. For skin lesion analysis, MTL methods commonly focus on skin lesion segmentation and classification tasks.

Most non-end-to-end methods leverage segmentation masks as a background removal strategy to enhance classification performance [89–91]. These methods often employ a cascaded architecture, leveraging segmentation outcomes to improve classification. Yu et al. [89] proposed a non-end-to-end network that utilized a fully connected residual network (FCRN) for skin lesion segmentation. The output mask of this network was used to crop the image to obtain just the lesion patch. This patch was then used as an input to a deep residual network to classify the presence of melanoma. Another non-end-to-end approach was proposed in [92], which implemented a four-stage architecture for skin lesion segmentation. The first stage was feature extraction using a ResNet50 architecture. This was followed by a network for region proposal. The Proposed regions were used as input to an R-CNN network that was used for bounding box classification and adjustment. The bounding boxes that are classified as lesions were sent to the final stage which utilized the SkinNet [93] architecture for segmentation. Few methods propose sharing parameters between segmentation and classification. For instance, [18] proposes a mutual bootstrapping method called MB-DCNN for segmentation and classification. This method initially uses a coarse segmentation network to generate segmentation masks. These masks are then used with the original image as input to a classification network. Class activation maps [24] (as explained in Section 2) from the classification network are then utilized to train an enhanced segmentation network. This method shows that CAMs from the classification network help refine the segmentation results.

On the other hand, end-to-end methods generally consist of a shared encoder and distinct outputs for each task, either parallel or interacted architectures. For instance, [94] employs a shared encoder for feature extraction and three separate output branches for segmentation, melanoma classification, and seborrheic keratosis classification. Similarly, [56, 95] also propose end-to-end architectures for segmentation and classification. The study by [95] introduces an end-to-end architecture encompassing three primary tasks: skin lesion detection, segmentation, and classification. In contrast, [56] focuses solely on lesion segmentation and classification tasks, integrating a feature passing branch to facilitate feature transfer between the segmentation and classification branches. More

Table 4: Summary of recent work on skin lesion classification on the SymDerm and SymDerm v2.0 datasets. The table provides a comparison of numerous works on the two datasets using five metrics: Balanced Accuracy (B.Acc), Kappa Score, and Weighted Average of Precision, Recall, and F1 Score.

Method	Dataset	Num. Classes	B.Acc	Kappa Score	Weight. Avg. Precision	Weight. Avg. Recall	Weight. Avg. F1-Score
SIFT + Decision Tree [88]	SymDerm	3	0.479	0.256	0.563	0.523	0.539
SIFT + Decision Tree [88]	SymDerm	2	0.676	0.353	0.678	0.676	0.677
Shape + Colour based [15]	SymDerm	3	0.498 ± 0.018	0.291 ± 0.025	0.566 ± 0.015	0.562 ± 0.016	0.560 ± 0.015
CNN [16]	SymDerm	3	0.615 ± 0.019	0.429 ± 0.030	0.690 ± 0.026	0.627 ± 0.022	0.645 ± 0.021
CNN [16]	SymDerm	2	0.719 ± 0.029	0.441 ± 0.059	0.735 ± 0.035	0.722 ± 0.029	0.718 ± 0.029
CNN with data augmentation [17]	SymDerm v2.0	2	0.711 ± 0.010	0.422 ± 0.020	0.715 ± 0.012	0.714± 0.012	0.712 ± 0.011
CNN with MTL [17]	SymDerm v2.0	2	0.699 ± 0.035	0.398 ± 0.073	0.702 ± 0.035	0.699 ± 0.038	0.699 ± 0.038

recently, [96] proposed MTL-CNN that included the task of edge prediction to boost the performance of segmentation. This network consists of one encoder for feature extraction, two parallel decoders for the dense prediction tasks of edge detection and lesion segmentation, and one classification subnet for skin lesion classification. In this architecture, the output from the edge decoder is utilized to improve the segmentation results by using multiple Edge Information Enhancement (EIE) modules along the skip connections from the encoder to the segmentation decoder. Incorporating the EIE modules on each skip connection ensures enhancement at different scales of feature extraction. Similarly, the output of the segmentation decoder is utilized to improve the classification results by using multiple Lesion Area Extraction (LAE) modules along the skip connections from the encoder to the classification subnet.

4 METHODOLOGY

This section contains the details of the proposed model architectures and training strategies. The core idea of utilizing these networks is to leverage segmentation masks to improve symmetry classification and utilize CAMs from the symmetry classification network to improve the performance of skin lesion classification. For this purpose, a cascaded architecture, trained in a non-end-to-end manner, is proposed (as depicted in Figure 2).

4.1 Symmetry Enhanced Lesion Classification Network (SE-LCN)

The proposed architecture consists of three distinct networks, each trained for specific tasks:

4.1.1 ***The Segmentation Network*** is responsible for generating lesion masks of dermoscopic images to aid classification of symmetry and lesion type. For this task, we train and evaluate DeepLabv3+ [52], and U-Net [43], and the better-performing model is utilized as the network for segmentation. These two models were chosen for comparison because these two are the best-performing standalone models, as discussed in Section 3. The model is trained with a hybrid loss function that was introduced in [18], and is defined as follows:

$$L_{hybrid} = L_{dice} + \lambda L_{rank}, \tag{2}$$

where λ is a weighting factor and L_{dice} [97] and L_{rank} represent the dice loss and the rank loss respectively. The dice loss is a loss function derived from the dice similarity coefficient, which measures the similarity between two sets. The dice loss is defined as:

$$L_{dice} = 1 - \frac{|X \cap Y|}{|X| + |Y|}, \tag{3}$$

where $|X \cap Y|$ counts the number of true positive pixels (i.e., the intersection where both the prediction and the ground truth are positive), $|X|$ is the number of positive pixels in the ground truth, and $|Y|$ is the number of positive pixels in the prediction. The rank loss is employed to pose additional constraints on hard and easy pixels based on prediction error [18]. In this function, pixels from both foreground and background are ranked by their error after forward propagation of each batch. The top K pixels with the largest error in lesion or background are selected as hard pixels in this area. If H^0_{ni} and H^1_{nj} are prediction values of the i_{th} hard pixel of the background (denoted by 0) and the j_{th} hard pixel of the lesion (denoted by 1) for the n_{th} input image, and m be the margin value, the rank loss is defined as:

$$L_{\text{rank}}(X_n, Y_n) = \frac{1}{K^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \max(0, H^0_{ni}(X_n, Y_n) - H^1_{nj}(X_n, Y_n) + \text{m}), \tag{4}$$

which enforces $H^1_{ij} > H^0_{ij} + \text{m}$ in the training stage. Such a design enables a segmentation network to pay more attention to those hard pixels and thus learn more discriminative information.

The segmentation outcomes were utilized by both the classification networks by masking the input image. Throughout the experimentation, two distinct masking strategies were explored, each with the aim of optimizing classification accuracy. The strategy that demonstrated better results was integrated into the final architecture. The two masking strategies are:

- **Segmentation Masks:** These masks are obtained by simply using the binary masks (obtained from the segmentation model) to the input image. The mask keeps the segmented

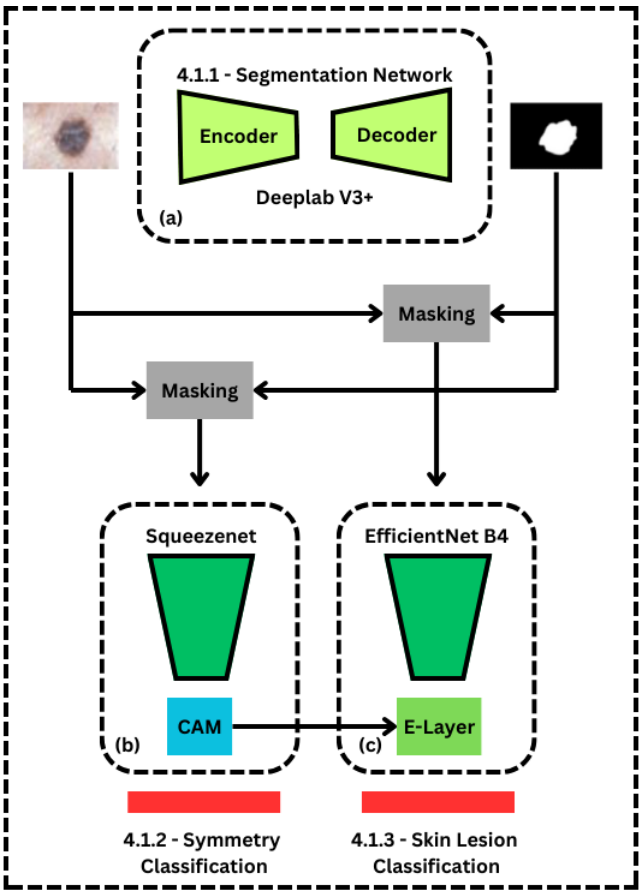


Figure 2: Symmetry Enhanced Lesion Classification Network (SE-LCN): The proposed non-end-to-end architecture includes (a) the segmentation network, the output of which is used to mask input images for both (b) the symmetry classification network and (c) the lesion classification network. Class Activation Maps (CAMs) from the symmetry classification networks are merged with the feature maps of the lesion classification network using an Enhanced Layer (E-Layer).

.

(a) Input Image **(b) Segmented Mask** **(c) LAE Mask**

Figure 3: Two types of masking strategies for the input image shown in (a) are the Segmentation mask (b) and the LAE mask (c).

region (foreground) of the input image as it is and converts the remaining pixels (background) to black pixels. Such a masking strategy ensures that the model only focuses on the exact region where the lesion is segmented. An example of this mask applied to an input image is shown in Figure 3b.

- **Lesion Area Extraction (LAE) Masks:** The LAE masks were implemented in [96] where a box region around the lesion is kept in the input image. In this mask, a bounding box is detected around the lesion, the box is dilated by a factor of 1.4, and this resultant box region is retained while the remaining pixels are converted to black pixels. This mask aims to ensure that the model focuses on the lesion area and the surrounding skin region, which may be helpful in classification. An example of this mask applied to an input image is shown in Figure 3c.

4.1.2 ***The Symmetry Classification Network*** utilizes the masked inputs (using masks generated by the segmentation network) to classify the symmetry of a lesion. Four light-weight architectures were considered for this network: ResNet34, EfficientNetB0, ShuffleNet, and SqueezeNet. On training and evaluating these networks on the SymDerm v2.0 dataset using the same hyperparameters, SqueezeNet gave the best performance. Hence, a SqueezeNet [98] architecture pre-trained on Imagenet [99] is trained to classify the symmetry of a skin lesion into two classes: symmetric and asymmetric. A weighted cross-entropy loss is employed to tackle the class imbalance between the two classes. The weight for a class 'i' (w_i) is calculated as follows:

$$w_i = \frac{\text{total\_samples}}{\text{num\_samples\_in\_class\_i} * \text{num\_classes}}, \tag{5}$$

where 'total_samples' is the total number of samples in the dataset 'num_samples_in_class_i' is the number of samples in class i, and 'num_classes' is the total number of classes in the dataset (in this case, 2). The weighted cross-entropy loss for binary classification is calculated as follows:

$$\text{Weighted CE}(p, y, w) = -\left(w \cdot y \log(p) + (1 - y) \log(1 - p)\right), \tag{6}$$

where p is the predicted probability of class 'symmetric', y is the actual label (0 for asymmetric and 1 for symmetric), and w is the weight of class 'symmetric'.

4.1.3 ***The Lesion Classification Network*** utilizes the masked inputs from the segmentation network and CAMs from the symmetry classification network to classify the type of skin lesion into 7 distinct classes. The EfficientNet B4 [65] architecture pre-trained on Imagenet [99] is fine-tuned using a weighted cross-entropy loss function to tackle class imbalance. The weights for each class are calculated using eq. (5), and the weighted cross-entropy loss for multi-class classification is calculated as:

$$\text{Weighted CE}(p, y, w) = -\sum_{i=1}^{N} w_i \cdot y_i \log(p_i), \tag{7}$$

where p_i is the predicted probability of class i, y_i is the label of class i, and w_i is the weight of class i.

The enhanced layer (E-Layer), introduced in [18], is applied to fuse the CAMs from the symmetry classification network into the features of the lesion classification network. The E-layer, present in

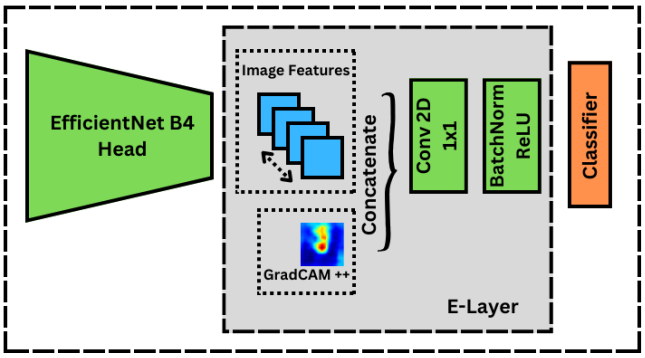


Figure 4: The E-Layer concatenates image features with the Grad-CAM++ output of the symmetry classification model. This is followed by a 1x1 convolution layer and a batch normalization layer with a ReLU activation function.

the lesion classification network, concatenates the incoming CAMs with feature maps extracted from the encoder, and employs a 1x1 convolution, which is followed by a batch normalization and a ReLU activation function to fuse the image features with the localization features (provided by the CAM). The resultant feature maps are flattened and fed to the classifier of the network. The architecture of the E-Layer can be seen in Figure 4. To extract CAMs from the symmetry classification network, we use Grad-CAM++ [26].

5 EXPERIMENTAL SETUP AND RESULTS

This section contains a detailed description of the data preprocessing steps used for each network, the experimental framework for the experiments, and the results of all the experiments conducted in this work.

5.1 Data and Preprocessing

To train the lesion segmentation and the lesion classification networks, the HAM10000 dataset [20] was used. This dataset, however, consists of numerous images containing skin hair, which could cause distractions to the model and negatively impact its performance. To remove these distractions, we implemented a hair removal technique that was used in [100]. This technique uses a thresholded image from the output of blackhat transform for segmenting the hair pixels, and uses the fast marching method [101] for inpainting using the surrounding pixels. Representative examples of skin lesion images, before and after the application of hair removal, are illustrated in Figure 5.

HAM10000 dataset consists of lesion images belonging to 7 classes which include: melanoma, melanocytic nevi, basal cell carcinoma, actinic keratosis, benign keratosis-like lesions, dermatofibroma, and vascular lesions. However, the distribution of images in these classes is highly imbalanced. To prevent class bias in the model, we balanced the dataset, aiming to have 2,000 images for each class. The classes that had a higher number of images were downsampled by randomly selecting 2,000 images, and the classes with a lower number of images were upsampled using data augmentation. The

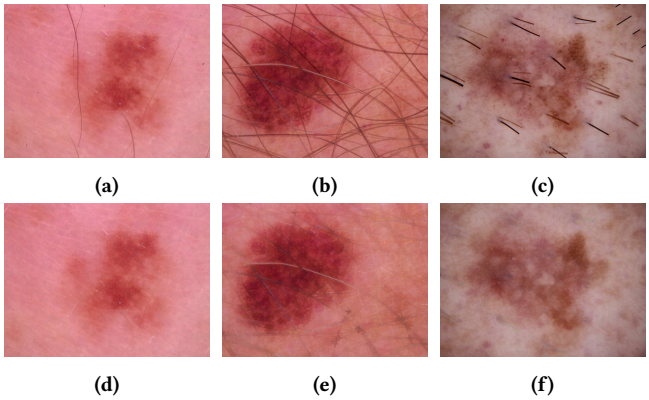


Figure 5: Examples of the hair removal step on three images of the HAM10000 dataset. Subfigures (a), (b), and (c) represent images with hair, while subfigures (d), (e), and (f) represent the corresponding images after hair removal.

.

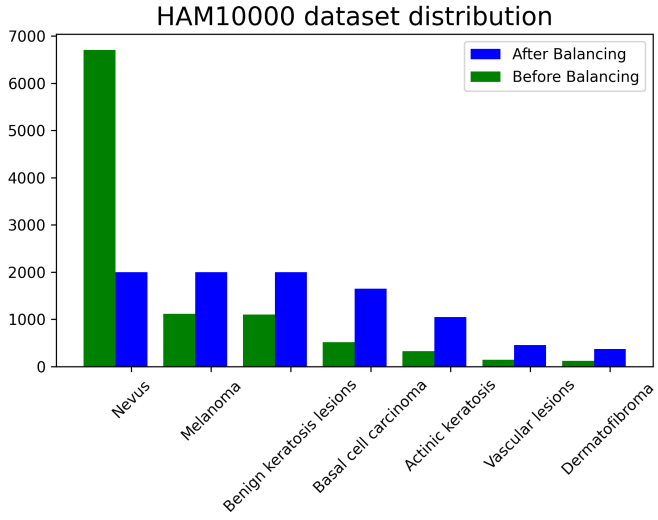


Figure 6: Class distribution of HAM10000 dataset before and after dataset balance.

augmentation techniques used were vertical flip, horizontal flip, and 90-degree rotation. The distribution of classes before and after balancing can be seen in Figure 6. Notably, certain classes such as basal cell carcinoma, actinic keratosis, vascular lesions, and dermatofibroma could not be augmented to the 2,000-image target due to their limited initial sample quantities.

To train the symmetry classification model, the SymDerm v2.0 dataset [17] is used. Although this dataset contains images belonging to three classes, we consolidate the classes 'Symmetry w.r.t. 1 axis' and 'Symmetry w.r.t.2 axes' into a single 'Symmetric' class for simplicity, as done in [17].

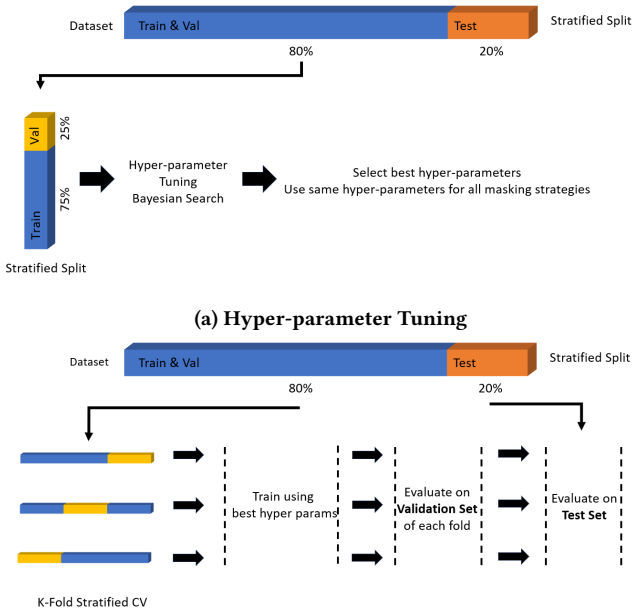
5.2 Implementation and Learning Details

The experiments in this research are carried out using the PyTorch framework and were run on the EEMCS HPC Cluster provided by the University of Twente. The cluster consists of several GPU nodes that use SLURM to manage submitted jobs.

Each network of the proposed architecture is trained separately in a sequential order. Each dataset is split where 80% of the data is used for training and validation, and 20% is used for testing. A stratified split is used to ensure that the class ratio remains the same on both splits. The training and validation data is initially further split where 75% is used for training and 25% is used for validation. Using this split data, a Bayesian hyperparameter search [102] is conducted to find the optimal hyperparameters. The hyperparameters yielding the highest validation accuracy are selected. The hyperparameters used in this search are batch size, learning rate, momentum, weight decay, and weighted loss function. The selected hyperparameters are then used to conduct experiments on a network (comparing masking strategies). The experiments are conducted by employing a K-fold stratified cross-validation on the initial training and validation set (80% of the entire dataset). The average performance is evaluated across K folds on the validation set of each fold, and the separate test set. The experimental setup is visualized in Figure 7. Each fold is trained for a maximum of 100 epochs, where the weights are saved from the epoch with the lowest validation loss. An early stopping strategy is applied with a patience of 10 epochs to prevent overfitting. All three networks are trained with the ADAM optimizer [103]. For the symmetry classification network, a 3-fold cross-validation is implemented, whereas for the final SE-LCN, a 4-fold cross-validation is implemented. The final SE-LCN model that is trained on the HAM10000 dataset is also evaluated on the ISIC 2019 dataset to test the generalization capabilities of the model. The ISIC 2019 dataset consists of 8 classes, the seven classes from the HAM10000 dataset and squamous cell carcinoma (SCC). The images belonging to SCC category were omitted to restrict the evaluation to images belonging to 7 classes.

The following experiments are designed in this work:

- **Experiment 1:** Training and evaluating DeepLabv3+ (segmentation network) on HAM10000 dataset.
- **Experiment 2:** Training and evaluating SqueezeNet (symmetry classification network) on SymDerm v2.0 dataset
  - Without masked input.
  - Using Segmentation masks as input.
  - Using LAE masks as input.
- **Experiment 3:** Training and evaluating EfficientNet B4 (lesion classification network) on HAM10000 dataset:
  - Without masked input.
  - Using Segmentation masks as input.
  - Using LAE masks as input.
- **Experiment 4:** Training and evaluating EfficientNet B4 (lesion classification network) on HAM10000 dataset with CAMs from symmetry classification network using all combinations of masking the symmetry classification network and lesion classification network.



(a) Hyper-parameter Tuning

(b) Cross-validation training and evaluation

Figure 7: The experimental framework includes (a) a Bayesian search for hyper-parameter tuning and (b) a k-fold cross-validation training and evaluation strategy.

- **Experiment 5:** Evaluating the models trained in Experiment 4 (using the HAM10000 dataset) on the ISIC2019 dataset.
- **Additional Experiment:** Training and evaluating the final SE-LCN models on the ISIC2016 dataset (binary classification) using all combinations of masking the symmetry classification network and lesion classification network. Results are presented in Appendix C.

5.3 Evaluation Metrics

5.3.1 ***Evaluation of predictive performance*** To evaluate the performance of the segmentation network, the dice index, pixel-wise sensitivity, pixel-wise specificity, and pixel-wise accuracy are used. To evaluate the performance of the symmetry classification network, the balanced accuracy, kappa score, and weighted average of precision, recall, and F1-score are calculated. To evaluate the lesion classification network, the accuracy, kappa score, and macro average of precision, recall, and F1-score are calculated.

5.3.2 ***Evaluation of localization of CAMs*** To evaluate the best masking strategy for the symmetry classification network and the impact of including class activation maps to SE-LCN, apart from a quantitative measure of performance using the metrics mentioned previously, we measure the localization accuracy of the generated CAMs. This is done by comparing the CAMs generated by these models and the ground truth segmentation masks. For this comparison, we use a threshold of 0.5 on the activation map (which consists of values between 0 and 1) and consider any value above

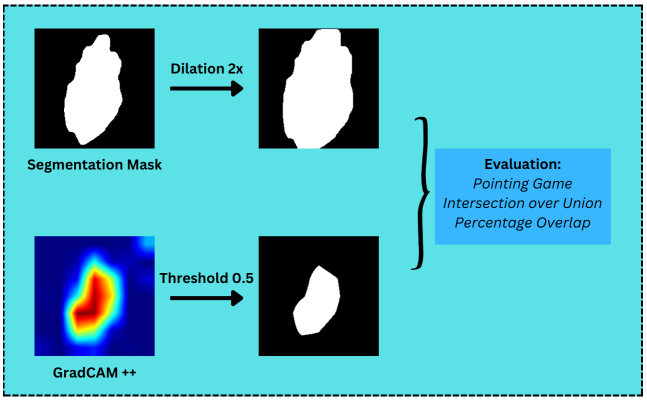


Figure 8: To evaluate the localization of CAMs, the ground truth segmentation mask is used as a reference. The segmentation mask is dilated twice, and a threshold of 0.5 is used to find the activation points in the CAM.

the threshold as an activation point. Next, the ground truth segmentation mask is dilated twice to increase the area of the ground truth region. This is done to consider activation points on the edges of the lesion and surrounding skin tissues as true positives. Finally, the localization is evaluated using three metrics: Intersection over Union (IoU), Percentage Overlap, and Pointing Game (refer eq. (1)). The IoU is calculated as:

$$\text{IoU} = \frac{|X \cap Y|}{|X \cup Y|}, \tag{8}$$

and the percentage overlap is calculated as

$$\text{Percentage Overlap} = \frac{|X \cap Y|}{|X|}, \tag{9}$$

where |X| represents the number of activation points in the CAM, and |Y| represents the number of points in the lesion region calculated after dilation. The process of evaluating the localization of CAMs is illustrated in Figure 8.

5.4 Results

5.4.1 ***Skin Lesion Segmentation*** Table 5 presents the performance measures of the DeepLabv3+ and U-Net segmentation models, assessed on the HAM10000 test set. The DeepLabv3+ model achieves a dice index of 0.8882, a pixel-wise sensitivity of 0.9083, a pixel-wise specificity of 0.9178, and a pixel-wise accuracy of 0.9184. The U-Net model achieves a dice index of 0.8807, a pixel-wise sensitivity of 0.8873, a pixel-wise specificity of 0.9328, and a pixel-wise accuracy of 0.9150. Both models have similar performances on the HAM10000 dataset. However, since DeepLabv3 + gave better performance in 3 metrics, we decided to utilize the same in the SE-LCN pipeline.

5.4.2 ***Symmetry Classification*** The comparative analysis of SqueezeNet models trained on the SymDerm v2.0 dataset, employing different masking strategies is summarized in Table 6a and Table 6b. These models were evaluated using five metrics, where the values represent the mean and standard deviation across 3 cross-validation folds. Four models are compared in these tables,

Table 5: The performance of DeepLabv3+ and U-Net models evaluated on the HAM10000 test set. Performance is evaluated using four metrics: dice index, pixel-wise sensitivity, pixel-wise specificity, and pixel-wise accuracy.

Metric	DeepLabv3+ [52]	U-Net [43]
Dice Index	**0.8882**	0.8807
Pixel-wise Sensitivity	**0.9083**	0.8873
Pixel-wise Specificity	0.9178	**0.9328**
Pixel-wise Accuracy	**0.9184**	0.9150

one model trained from scratch without any masking (model 1), one pre-trained model without any masking (model 2), and two pre-trained models with segmentation masking and LAE masking (models 3 and 4). The pre-trained models (models 2-4) were fine-tuned to the SymDerm v2.0 dataset. Table 6a shows the model performances on the corresponding validation sets across 3-folds. Across three folds, the model that was trained from scratch had a mean balanced accuracy of 0.664, a mean kappa score of 0.392, and a mean weighted average precision, recall, and F1-score of 0.669, 0.668, and 0.665 respectively on the validation set. The values indicate a consistent enhancement in model performance with a pre-training and fine-tuning approach for symmetry classification. Among the three pre-trained models, the model employing segmentation masking (model 3) outperformed the others in validation metrics, having a mean balanced accuracy of 0.740, a mean kappa score of 0.481, and a mean weighted average precision, recall, and F1-score of 0.744, 0.742, and 0.742 respectively. This was followed by the model employing LAE masking (model 4) while the model without input masking (model 2) gave the worst performance.

Table 6b shows the model performances on a separate test set across 3-folds. The model that was trained from scratch had a mean balanced accuracy of 0.673, a mean kappa score of 0.347, and a mean weighted average precision, recall, and F1-score of 0.680, 0.679, and 0.678 respectively on the test set. A similar trend can be observed in which the pre-trained networks that are fine-tuned to classify symmetry perform better than the model trained from scratch. However, the masking strategies do not show improvement in results over the model trained without input masking. The pre-trained model without input masking (model 2) exhibits superior performance across all metrics, having a mean balanced accuracy of 0.758, a mean kappa score of 0.511, and a mean weighted average precision, recall, and F1-score of 0.761, 0.757, and 0.758 respectively. Among the two masking strategies, using the LAE masks (model 4) demonstrated better performance on all metrics except for the kappa score.

Furthermore, the localization of CAMs generated by the three pre-trained models (models 2-4) was evaluated using the segmentation masks of the HAM10000 dataset, as shown in Table 7. The best-performing fold (using test accuracy) of each model was used to generate CAMs. Out of the three models, the model with no input masking (model 2) gave the best mean IoU of 0.321, whereas

the segmentation masking strategy (model 3) gave the best performance on all mean percentage overlap and pointing game scores with values of 0.940 and 0.812 respectively. Notably, the pointing game score is significantly higher for this model. A few examples of CAMs generated by different masking strategies are illustrated in Figure 9. A general difference among the CAMs generated by the models using different masking strategies can be observed from Figure 9. The models that do not use input masking have activation points spread across the image, including the background region. On the other hand, the models that use segmentation masking have activation points along the borders of the lesion area. The models that use LAE masking have activation points along the edges of the LAE box.

5.4.3 ***Symmetry Enhanced Lesion Classification Network (SE-LCN)*** The performance comparison of the SE-LCN models trained on the HAM10000 dataset, employing various input masking strategies for the lesion classification network and the symmetry classification is summarized in Table 8a and Table 8b. It comprises 12 models, of which three are baseline models (models 1, 5, and 9), with no transfer of CAMs' information, and the remaining nine are complete symmetry-enhanced models with different masking combinations. The tables are divided into three groups, based on the masking strategy of the lesion classification model. Models 1-4 do not utilize input masking for the classification model, models 5-8 utilize the segmentation masking strategy, and experiments 9-12 utilize the LAE masking strategy.

Table 8a shows the SE-LCN performances on the corresponding validation sets across 4-folds. First, we analyze the inclusion of CAMs in each group of models. For the first group (models 1-4), where the classification model does not have an input mask, all three models that utilize transfer of CAMs (models 2-4) show an improvement in performance as compared to the baseline model (model 1). Model 1 gives a mean accuracy of 0.859, whereas models 2,3, and 4 give a mean accuracy of 0.871, 0.869, and 0.883 on the validation set. A similar trend is observed in the 3rd group (models 9-12) where the transfer of CAMs improves the performance when compared to the baseline (model 9). However, the trend is different for the second group (models 5-8), which utilizes segmentation masks for the lesion classification model. In this group, the baseline model with no transfer of CAMs (model 5) has a mean accuracy of 0.800 and slightly outperforms the other three models of the group, which have mean accuracies of 0.797, 0.798, and 0.787 respectively.

It can be noted that irrespective of the masking strategy used in the symmetry classification model, using no mask in the lesion classification model (model 1-4) yields the best performance, where the mean accuracies range from 0.859 to 0.883. Among the two masking strategies, using the LAE masks for the lesion classification models (model 9-12) gives better performance with mean accuracies ranging from 0.787 to 0.800. Using the segmentation masks (model 5-8). Among the four best models (model 1-4), using LAE masks on the symmetry classification network gave the best performance on all metrics. Therefore the best-performing model on the validation set is a symmetry network trained with LAE input masks transferring CAMs to a lesion classification network trained with no input masks.

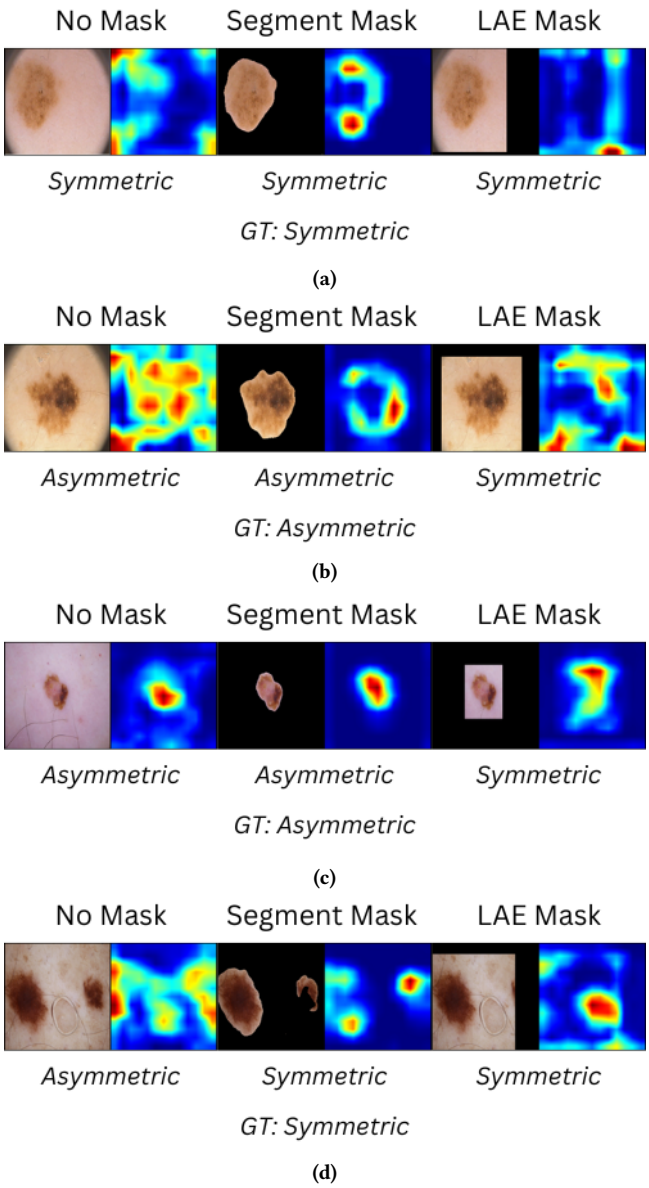


Figure 9: Four examples of Grad-CAM ++ of different masking strategies for symmetry classification on SymDerm v2.0 test set. The type of input masking (no mask, segmentation mask, and LAE mask) is mentioned above each input-CAM pair, and the predicted class is mentioned below. GT denotes the ground truth prediction of that image.

Table 8b shows the SE-LCN performances on the separate test set across 4-folds. A trend similar to Table 8a is observed where the lesion classification networks trained with no input masks (model 1-4) perform the best, followed by classification networks trained with LAE input masks (model 9-12). Among the four models with

Table 6: Performance comparison of SqueezeNet models trained on the SymDerm v2.0 dataset, employing various masking strategies to classify lesion symmetry. Each model's performance is quantified using five metrics: Balanced Accuracy, Kappa Score, Weighted Average Precision, Recall, and F1-Score, with values representing the mean and standard deviation calculated on the corresponding validation and test sets across 3 folds of cross-validation. The highest value of each metric is highlighted in bold.

(a) Evaluation on corresponding validation sets

Model			Validation Set *Mean ± Std. Dev*				
Model	**Masking**	**Pretrained**	**Balanced Acc.**	**Kappa Score**	**Weight Avg Precision**	**Weight Avg Recall**	**Weight Avg F1**
1	None	FALSE	0.664 ± 0.017	0.329 ± 0.033	0.669 ± 0.015	0.668 ± 0.015	0.665 ± 0.017
2	None	TRUE	0.733 ± 0.021	0.465 ± 0.041	0.735 ± 0.022	0.733 ± 0.020	0.733 ± 0.020
3	Segmentation	TRUE	**0.740 ± 0.013**	**0.481 ± 0.029**	**0.744 ± 0.016**	**0.742 ± 0.016**	**0.742 ± 0.015**
4	LAE (bbox)	TRUE	0.736 ± 0.009	0.475 ± 0.019	0.741 ± 0.010	0.741 ± 0.009	0.739 ± 0.009

(b) Evaluation on separate test set

Model			Test Set *Mean ± Std. Dev*				
Model	**Masking**	**Pretrained**	**Balanced Acc.**	**Kappa Score**	**Weight Avg Precision**	**Weight Avg Recall**	**Weight Avg F1**
1	None	FALSE	0.673 ± 0.025	0.347 ± 0.045	0.680 ± 0.023	0.679 ± 0.019	0.678 ± 0.021
2	None	TRUE	**0.758 ± 0.010**	**0.511 ± 0.019**	**0.761 ± 0.010**	**0.757 ± 0.009**	**0.758 ± 0.009**
3	Segmentation	TRUE	0.730 ± 0.010	0.458 ± 0.018	0.733 ± 0.009	0.729 ± 0.009	0.729 ± 0.009
4	LAE (bbox)	TRUE	0.732 ± 0.006	0.466 ± 0.013	0.736 ± 0.006	0.736 ± 0.006	0.735 ± 0.006

Table 7: Evaluation of localization of CAMs of SqueezeNet models trained on the SymDerm v2.0 dataset, employing various masking strategies to classify lesion symmetry. The best fold (from cross-validation) of each model is used for evaluation on the HAM10000 dataset. Metrics shown include Intersection over Union (IoU), Percentage Overlap, and Pointing Game scores, utilizing segmentation masks for reference comparison.

Model		HAM10000 dataset		
Model	**Masking**	**IoU**	**Overlap**	**Point Game**
2	None	**0.321 ± 0.182**	0.758 ± 0.263	0.673
3	Segmentation	0.249 ± 0.148	**0.940 ± 0.130**	**0.812**
4	LAE	0.161 ± 0.159	0.564 ± 0.367	0.407

no mask for the lesion classification network (model 1-4), the model with segmentation masks (model 3) used in the symmetry classification network, and the baseline model (model 1) achieved the highest accuracy of 0.849. However, the highest kappa score and the highest macro average of precision and recall are achieved by the model that uses segmentation masks in the symmetry classification network (model 3) with mean values of 0.721, 0.770, and 0.792 respectively. The highest macro average of recall is 0.824, achieved by the model that uses LAE masks (model 4) for symmetry classification. Overall, on the test set, the transfer of CAMs does not improve the classification accuracy but shows considerable improvements in all other metrics.

Table 9 summarizes the evaluation of localization of CAMs generated by all 12 models. The three baseline models (model 1, 5, and 9) have low mean IoU scores of 0.273, 0.270, and 0.260 respectively. However, these models have high mean overlap scores of 0.964, 0.967, and 0.967 respectively, and high pointing game scores of 0.952, 0.956, and 0.956 respectively. The transfer of CAMs to these baseline models shows a significant increase in IoU scores. The mean IoU scores for models with the transfer of CAMs range from 0.415 to 0.487. However, a slight drop in percentage overlap and pointing game score is observed with the inclusion of CAMs. The pointing game score for the remaining models ranges from 0.828 to 0.941 and the mean percentage overlap ranges from 0.888 to 0.959. The Grad-CAM ++ from all masking combinations of the SE-LCN framework on one example image is illustrated in Figure 10.

Table 10 summarizes the same SE-LCN model performances when evaluated on the ISIC2019 dataset across 4-folds. The dataset used for evaluation consisted of 24,703 images (the entire dataset except for images belonging to the class SCC). It can be observed from the table that the baseline model (model 1) with no masking and no CAM transfer gives the worst performance with a mean accuracy of 0.522. The inclusion of CAM in the classification models without masks (models 2-4) improves the performance on this dataset. Model 4, which transfers CAMs from symmetry models with LAE

Table 8: Performance comparison of SE-LCN models trained on the HAM10000 dataset, employing various masking strategies on inputs of both the symmetry classification network (that generates CAMs) and the lesion classification network. Each model's performance is quantified using five metrics: Accuracy, Kappa Score, Macro Average Precision, Recall, and F1-Score, with values representing the mean and standard deviation calculated on the corresponding validation and test sets across 4 folds of cross-validation. The highest value of each metric among a lesion classifier masking strategy is highlighted in bold, and the highest value of each metric among all experiments is highlighted in blue text. The baseline model is model 1, which uses no input masking and no transfer of CAMs

(a) Evaluation on corresponding validation sets

Model			Validation Set *Mean ± Std. Dev*				
Model	**Classification Model**	**Symmetry Model**	**Accuracy**	**Kappa Score**	**Macro Avg Precision**	**Macro Avg Recall**	**Macro Avg F1**
1	Without Masks	No CAM transfer	0.859 ± 0.014	0.828 ± 0.017	0.876 ± 0.013	0.884 ± 0.015	0.879 ± 0.014
2	Without Masks	Without Masks	0.871 ± 0.009	0.843 ± 0.010	0.889 ± 0.011	0.897 ± 0.008	0.892 ± 0.009
3	Without Masks	Segmentation Masks	0.869 ± 0.026	0.841 ± 0.032	0.883 ± 0.026	0.893 ± 0.022	0.887 ± 0.025
4	Without Masks	LAE Masks	**0.883 ± 0.006**	**0.858 ± 0.007**	**0.902 ± 0.003**	**0.909 ± 0.005**	**0.905 ± 0.003**
5	Segmentation Masks	No CAM transfer	**0.800 ± 0.017**	**0.757 ± 0.021**	0.813 ± 0.013	**0.825 ± 0.016**	0.818 ± 0.014
6	Segmentation Masks	Without Masks	0.797 ± 0.007	0.753 ± 0.008	0.817 ± 0.004	0.820 ± 0.009	0.817 ± 0.003
7	Segmentation Masks	Segmentation Masks	0.798 ± 0.005	0.754 ± 0.006	**0.817 ± 0.007**	0.824 ± 0.006	**0.819 ± 0.006**
8	Segmentation Masks	LAE Masks	0.787 ± 0.023	0.741 ± 0.028	0.809 ± 0.025	0.812 ± 0.019	0.808 ± 0.023
9	LAE Masks	No CAM transfer	0.847 ± 0.011	0.814 ± 0.013	0.860 ± 0.013	0.874 ± 0.009	0.865 ± 0.011
10	LAE Masks	Without Masks	0.859 ± 0.014	0.829 ± 0.017	0.875 ± 0.008	0.884 ± 0.016	0.878 ± 0.012
11	LAE Masks	Segmentation Masks	0.852 ± 0.016	0.820 ± 0.019	0.864 ± 0.013	0.879 ± 0.015	0.870 ± 0.014
12	LAE Masks	LAE Masks	**0.869 ± 0.014**	**0.841 ± 0.017**	**0.882 ± 0.008**	**0.894 ± 0.013**	**0.888 ± 0.011**

(b) Evaluation on separate test set

Model			Test Set *Mean ± Std. Dev*				
Model	**Classification Model**	**Symmetry Model**	**Accuracy**	**Kappa Score**	**Macro Avg Precision**	**Macro Avg Recall**	**Macro Avg F1**
1	Without Masks	No CAM transfer	0.849 ± 0.012	0.713 ± 0.005	0.750 ± 0.020	0.814 ± 0.013	0.777 ± 0.009
2	Without Masks	Without Masks	0.845 ± 0.004	0.716 ± 0.005	0.756 ± 0.013	0.822 ± 0.010	0.785 ± 0.008
3	Without Masks	Segmentation Masks	**0.849 ± 0.008**	**0.721 ± 0.011**	**0.770 ± 0.027**	0.819 ± 0.005	**0.792 ± 0.014**
4	Without Masks	LAE Masks	0.844 ± 0.009	0.716 ± 0.012	0.766 ± 0.014	**0.824 ± 0.005**	0.792 ± 0.008
5	Segmentation Masks	No CAM transfer	0.782 ± 0.009	**0.618 ± 0.013**	0.638 ± 0.020	0.738 ± 0.012	0.680 ± 0.015
6	Segmentation Masks	Without Masks	0.781 ± 0.009	0.617 ± 0.010	**0.647 ± 0.029**	**0.742 ± 0.007**	**0.684 ± 0.016**
7	Segmentation Masks	Segmentation Masks	0.776 ± 0.010	0.610 ± 0.011	0.642 ± 0.025	0.738 ± 0.017	0.679 ± 0.010
8	Segmentation Masks	LAE Masks	**0.782 ± 0.007**	0.616 ± 0.006	0.644 ± 0.021	0.723 ± 0.008	0.674 ± 0.014
9	LAE Masks	No CAM transfer	0.804 ± 0.002	0.653 ± 0.001	0.670 ± 0.014	0.778 ± 0.008	0.713 ± 0.008
10	LAE Masks	Without Masks	0.817 ± 0.005	0.672 ± 0.007	0.689 ± 0.002	0.784 ± 0.012	0.728 ± 0.005
11	LAE Masks	Segmentation Masks	0.812 ± 0.008	0.666 ± 0.011	0.671 ± 0.018	0.776 ± 0.017	0.714 ± 0.016
12	LAE Masks	LAE Masks	**0.819 ± 0.004**	**0.677 ± 0.006**	**0.702 ± 0.017**	**0.801 ± 0.005**	**0.742 ± 0.011**

input gives the best performance among all models with a mean accuracy of 0.676. Interestingly, the models that use some type of masking for lesion classification, but do not transfer CAMs (models 5 and 9) have a good performance with mean accuracies 0.622 and 0.643 respectively. However, all the models show a significant drop in performance when compared to performance on the HAM10000 test set. The localization accuracy of CAMs could not be evaluated on the ISIC2019 dataset since the dataset does not consist of segmentation masks.

Table 9: Evaluation of localization of CAMs of SE-LCN trained on the HAM10000 dataset, employing various masking strategies. The best fold (from cross-validation) of each model is used for evaluation on the HAM10000 dataset. Metrics shown include Intersection over Union (IoU), Percentage Overlap, and Pointing Game scores, utilizing segmentation masks for reference comparison. The highest value of each metric among a lesion classifier masking strategy is highlighted in bold. The baseline model is model 1, which uses no input masking and no transfer of CAMs

Model			CAM Localization *Entire Dataset*		
Model	**Classification Model**	**Symmetry Model**	**IoU**	**Overlap**	**Point Game**
1	Without Masks	No CAM transfer	0.273 ± 0.118	**0.964 ± 0.156**	**0.952**
2	Without Masks	Without Masks	**0.461 ± 0.187**	0.895 ± 0.175	0.844
3	Without Masks	Segmentation Masks	0.458 ± 0.190	0.888 ± 0.180	0.828
4	Without Masks	LAE Masks	0.427 ± 0.171	0.894 ± 0.175	0.941
5	Segmentation Masks	No CAM transfer	0.270 ± 0.110	**0.967 ± 0.156**	**0.956**
6	Segmentation Masks	Without Masks	0.427 ± 0.171	0.958 ± 0.137	0.941
7	Segmentation Masks	Segmentation Masks	0.415 ± 0.165	0.959 ± 0.139	0.94
8	Segmentation Masks	LAE Masks	**0.468 ± 0.193**	0.955 ± 0.140	0.892
9	LAE Masks	No CAM transfer	0.260 ± 0.107	**0.967 ± 0.159**	**0.956**
10	LAE Masks	Without Masks	0.468 ± 0.193	0.950 ± 0.134	0.892
11	LAE Masks	Segmentation Masks	0.475 ± 0.189	0.951 ± 0.133	0.899
12	LAE Masks	LAE Masks	**0.487 ± 0.195**	0.949 ± 0.133	0.896

Table 10: Performance comparison of SE-LCN models trained on the HAM10000 dataset and evaluated on the ISIC2019 dataset, employing various masking strategies on inputs of both the symmetry classification network (that generates CAMs) and the lesion classification network. Each model's performance is quantified using five metrics: Accuracy, Kappa Score, Macro Average Precision, Recall, and F1-Score, with values representing the mean and standard deviation calculated on the entire ISIC2019 dataset across 4 folds of cross-validation. The highest value of each metric among a lesion classifier masking strategy is highlighted in bold, and the highest value of each metric among all experiments is highlighted in blue text. The baseline model is model 1, which uses no input masking and no transfer of CAMs

Model			ISIC 2019 Dataset *Mean ± Std. Dev*				
Model	**Classification Model**	**Symmetry Model CAM**	**Accuracy**	**Kappa Score**	**Macro Avg Precision**	**Macro Avg Recall**	**Macro Avg F1**
1	Without Masks	No CAM Transfer	0.522 ± 0.014	0.339 ± 0.012	0.376 ± 0.011	0.516 ± 0.008	0.403 ± 0.012
2	Without Masks	Without Masks	0.669 ± 0.007	0.499 ± 0.008	0.505 ± 0.007	**0.541 ± 0.009**	0.511 ± 0.008
3	Without Masks	Exact Segment Masks	0.675 ± 0.004	0.501 ± 0.007	0.512 ± 0.007	0.535 ± 0.006	0.514 ± 0.009
4	Without Masks	LAE Masks	**0.676 ± 0.009**	**0.505 ± 0.005**	**0.518 ± 0.007**	0.539 ± 0.007	**0.522 ± 0.004**
5	Exact Segment Masks	No CAM Transfer	0.622 ± 0.007	0.430 ± 0.012	0.465 ± 0.016	**0.509 ± 0.008**	0.465 ± 0.006
6	Exact Segment Masks	Without Masks	0.622 ± 0.008	**0.434 ± 0.009**	0.453 ± 0.016	0.503 ± 0.006	0.457 ± 0.014
7	Exact Segment Masks	Exact Segment Masks	0.621 ± 0.008	0.433 ± 0.009	0.461 ± 0.022	0.501 ± 0.008	0.461 ± 0.015
8	Exact Segment Masks	LAE Masks	**0.624 ± 0.007**	0.430 ± 0.012	**0.469 ± 0.009**	0.495 ± 0.013	**0.465 ± 0.018**
9	LAE Masks	No CAM Transfer	**0.643 ± 0.009**	**0.454 ± 0.010**	0.460 ± 0.017	**0.525 ± 0.011**	**0.457 ± 0.018**
10	LAE Masks	Without Masks	0.639 ± 0.005	0.433 ± 0.017	**0.472 ± 0.012**	0.506 ± 0.018	0.457 ± 0.012
11	LAE Masks	Exact Segment Masks	0.635 ± 0.005	0.434 ± 0.013	0.464 ± 0.009	0.507 ± 0.010	0.451 ± 0.012
12	LAE Masks	LAE Masks	0.636 ± 0.007	0.425 ± 0.013	0.470 ± 0.013	0.502 ± 0.016	0.452 ± 0.019

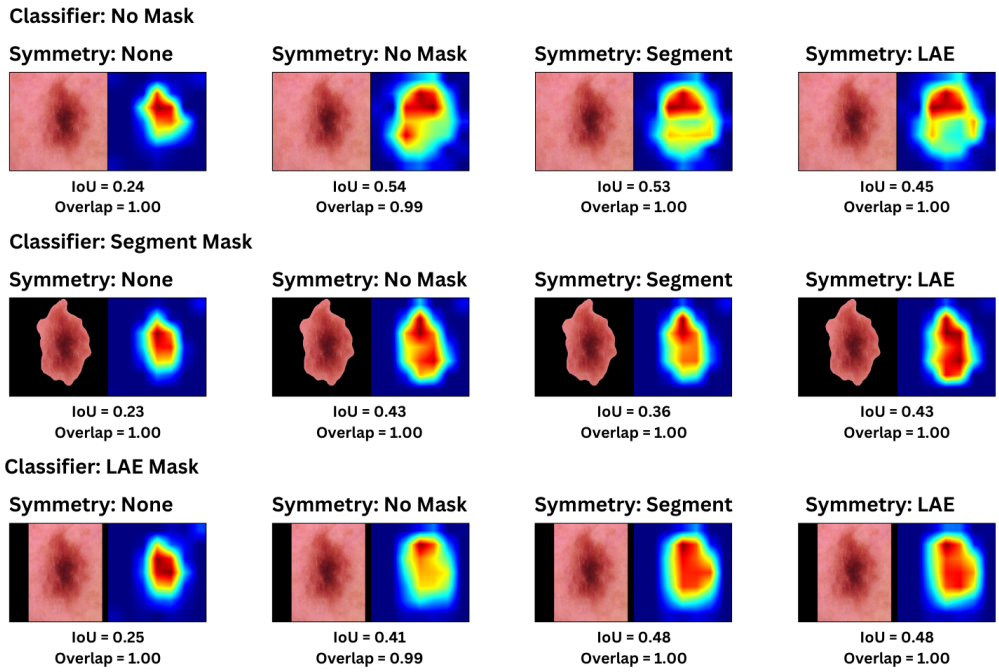


Figure 10: The figure illustrates Grad-CAM++ from all the masking combinations of the SE-LCN framework. Each row of images represents four images having the same masking strategy for the lesion classification network. The first image of each row represents a baseline approach with no CAM transfer. The following three outputs come from models with CAM transfer from the symmetry classification network using different masking strategies.

6 DISCUSSION

This section provides a detailed analysis of the results described in section 5.4.

6.1 Symmetry Classification

The results (refer Table 6a and Table 6b) show a clear performance improvement when networks pre-trained on the ImageNet dataset are fine-tuned for symmetry classification. This is because, pre-training facilitates model generalization [104], allowing the network to develop a robust initial understanding of generic features before adapting to more specific data characteristics found in dermoscopic images. Notably, all three pre-trained networks also outperform the existing works in the literature (mentioned in Table 4, which include classical approaches [15, 88] and D-CNN approaches trained from scratch [16, 17].

The impact of masking strategies was compared among the three pre-trained models on the validation sets and the test set. While the performance on the validation sets is indicative of the robustness of the model to different data distributions (since each fold has a different training and validation split), the performance on the test set indicates the model performance on unseen data. Notably, the trends differ in the validation sets and the test sets. While input masking methods showed superior performance on the validation set, they caused a decline in performance on the test set. The performance comparison among the two masking strategies also differs in both datasets. The segmentation masks perform better when evaluated on the validation set, the LAE masks perform better on the test set.

On evaluating the localization of CAMs on the HAM10000 dataset (refer to Table 7), the mean IoU was the highest for the model with no input masking, whereas the percentage overlap and pointing game score were the highest for the model with segmentation masking. One possible reason for this could be because the CAMs for segmentation masking focused more on the edges of the lesion, as seen in Figure 9, which could be useful while classifying symmetry. Since the segmentation masking forces the model to focus on the edges, the overall intersection between the activation points and the segmentation mask is less, giving it a lower IoU score. However, since the edges lie inside the dilated region of the segmentation mask, the overlap and point game score are high for this masking strategy, which indicates that most activation points lie within the dilated segmentation region. All three CAM localization metrics give the lowest score for the model using LAE masking. As observed in Figure 9, in many cases, this masking strategy forces the model to focus on the edges of the LAE box. Therefore, most activation

points do not lie within the dilated segmentation region. Hence, the segmentation masking strategy yields the best CAM localization.

6.2 Symmetry Enhanced Lesion Classification Network

Similar to the symmetry classification network, the SE-LCN was evaluated on validation sets and the test set (refer Table 8a and Table 8b), using a 4-fold cross-validation. We observed that on the validation set, the inclusion of CAMs had a definitive improvement in the performance over the baseline models when the lesion classifier did not utilize segmentation masks (model 1-4), and when the lesion classifier utilized LAE input masking (model 9-12). The same was, however, only true for the LAE masking (model 9-12) models when evaluated on the test set.

On evaluating on both the validation set and test set, the lesion classification models trained without input masking (models 1-4) showed the best performance, followed by the lesion classification models trained with LAE masks applied to the input (model 9-12). The segmentation input masks (model 5-8) yielded the worst performance. This indicates that restricting the input image to the lesion region negatively impacts the classification performance.

The trends again differ on both evaluation sets while varying the masking strategy on the symmetry classification model. Among the four best models (models 1-4), using the LAE mask on the symmetry classification model (model 4) gives a superior performance on the validation set, while using segmentation masks (model 3) gives the best performance on 3 metrics of the test set. Interestingly, the accuracy on the test set remains the same for the baseline network (model 1) and the network with segmentation masks on the symmetry model (model 3), but the latter shows better performances on macro averages of precision, recall, and F1-score. The class-wise evaluation metrics of these two models are summarized in Appendix B.

The SE-LCN models trained on the HAM10000 dataset were also evaluated on the ISIC2019 dataset, to analyze the impact of masking and CAM transfer on the generalization capabilities of the model. It can be noticed that input masking and CAM transfer improve the model performance on ISIC2019 over the baseline model (model 1). Interestingly, the models that did not utilize CAM transfer and utilized input masking (models 5 and 9) had good performance when compared to the baseline. Similarly, the models with no masking that utilized CAM transfer also performed well (models 2-4). This shows that using either input masking or CAM transfer improves the model's generalizability. Overall the performance of every model drops when tested on the ISIC2019 dataset. This shows that the models do not maintain high performance when tested on a larger variety of data.

Additionally, the SE-LCN models were trained and evaluated on the ISIC2016 [33] dataset, a binary classification dataset to classify lesions as malignant or benign. The ISIC2016 dataset consists of 900 images. The same approach of training and evaluation, as mentioned in Figure 7, was implemented in this experiment where 20% of the data was used for testing and a 4-fold cross validation was used on the remaining data. The results of these experiments are summarized in Appendix C. Table 13 shows the results on the corresponding validation set and Table 14 shows the results on a separate test set. On the test set, the transfer of CAMs shows an improvement in performance except when using segmentation masks (models 5-8). Overall, using segmentation masking in the classifier (models 5-8) gives the worst performance, similar to the models when trained on the HAM10000 dataset. Among the masking strategies used in the symmetry model, using no input masking gives the best performance.

Upon evaluation of the localization of CAMs generated by the lesion classification model, the baseline models have a high percentage overlap score and a pointing game score. The inclusion of CAMs slightly decreased the performance on these metrics. However, the inclusion of CAMs increased the mean IoU scores over the baseline models. Figure 10 illustrates the impact of including CAMs with different masking strategies on an input image. A general observation is that the three baseline approaches (with no transfer of CAMs from the symmetry model) generates smaller regions of activation. When symmetry CAMs are transferred to these models, the output GradCAM covers more of the lesion region, which explains the increase in IoU. The smaller region of CAMs for the baseline approaches can explain the higher values of the mean percentage overlap and the pointing game score since the majority of the activation points would lie inside the dilated segmented region.

6.3 Answers to Research Questions

Based on the results obtained from the experiments, the following are the answers to the research questions mentioned in the Section 1:

(1) **Impact of inclusion of CAMs from symmetry classification network on the predictive performance of the lesion classification network:** The impact of inclusion of CAMs on the predictive performance can be observed in Table 8 where Table 8a denotes the model performances on the validation sets and Table 8b denotes the model performances on the test set. First, we look at the inclusion of CAMs (models 2-4) in the baseline model (model 1) which uses no input masking on the classifier. In both datasets, we observe that models 2-4 outperform the baseline, indicating a positive impact on performance. There is no significant improvement in accuracy in the test set, but all other metrics show an improvement over the baseline. A similar trend can be seen in the models that use LAE masking for the lesion classifier (models 9-12), where the models with CAM transfer (models 10-12) show an improvement in performance across all metrics over model 9 in both datasets. Hence, while using LAE masking (models 9-12) or no masking (models 1-4) on the lesion classifier, the inclusion of CAMs has a positive impact on the predictive performance. However, this is not the case when using segmentation masks (models 5-8). Here, the inclusion of symmetry CAM does not show any improvement, and in certain cases decreases

the performance in few of the metrics. Given that the models using segmentation masking in the lesion classifier have lower performance than the other masking strategies, it could have a negative impact on CAM transfer. Therefore, the results suggest that the inclusion of CAMs from the symmetry classification network has a positive impact on the predictive performance of the SE-LCN model except when segmentation masking is used on the lesion classifier.

(2) **Impact of inclusion of CAMs from symmetry classification network on localization accuracy of CAMs generated by lesion classification network:** The impact of inclusion of CAMs on the localization accuracy can be observed in Table 9. The baseline models (models 1,5 and 9) already have a high Percentage Overlap and Pointing Game score which decreases in all cases with the inclusion of CAMs. However, the IoU score significantly increases in all cases with the inclusion of CAMs. Figure 10 illustrates the CAM for an input image generated by all the models (1-12). It can be observed that the three baseline models have a smaller region of activation. This could explain why these models have a higher Percentage Overlap and Pointing game score since most of the activation points lie inside the dilated lesion area. The models that utilize symmetry CAMs generate larger CAMs, adding focus to the edges and immediate surrounding regions of the lesion. This is why they give a larger intersection with the segmentation masks, yielding a higher IoU. Therefore, while the inclusion of CAMs from the symmetry classification model does not cause an improvement in two of the CAM localization metrics, it helps the model focus more on the edges (and thereby symmetry) of a lesion while classifying the lesion.

The following are the answers to the subquestions mentioned in Section 1:

(1) **Impact of input masks on the predictive performance of symmetry classification and lesion classification networks:** Table 6 describes the predictive performance of the symmetry classification models using different masking strategies. On the validation set (Table 6a), using input masking (models 3 and 4) showed superior performance over the baseline pre-trained model (model 2). However, on the test set, the best performance was achieved by model 2 which does not use input masking. A higher mean performance on the validation sets could indicate robustness to varying data distributions since a cross-validation approach was implemented, but the performance on the test set is more indicative of the real-world performance. Hence input masking does not benefit the predictive performance of the model in real-world data.

Next, we analyze the impact of segmentation masking on the predictive performance of the lesion classification model (indicated in Table 8). The performances of models evaluated on the validation set and the test set show a similar trend in which the best performance is obtained by the models that do not use any masking (models 1-4), followed by the models that use LAE masking (models 9-12). The worst performance is obtained by the models that utilize segmentation masks (models 5-8). Hence, the use of input masking does not benefit the predictive performance of the lesion classification network.

(2) **Impact of input masks on the localization accuracy of CAMs generated by the symmetry classification and lesion classification networks:** Table 7 describes the localization accuracy of CAMs generated by the symmetry classification model using different masking strategies. While the baseline model with no masking (model 2) gives the best IoU score, the model that uses segmentation masking (model 3) gives the highest Percentage Overlap and Pointing Game score. The model that utilizes LAE masking gives the worst performance in all metrics. On visually analyzing the generated CAMs illustrated in Figure 8, it can be observed that the activation points of segmentation masking focus on the edges of the lesion, which is useful in determining the symmetry. This could explain the low IoU score since the intersection region of the activation points and segmentation masks is limited to the edges of the lesion. Since most of these points lie within the dilated lesion region, segmentation masking performs the best on the other two metrics. Hence, using segmentation masks generated by the lesion segmentation network gives the best CAM localization, helping with the explainability of the model.

The evaluation of localization accuracy of CAMs of the SE-LCN model is shown in Table 9. The three baseline models that do not include CAMs from the symmetry model (models 1, 5, and 9) show similar performances irrespective of the masking strategy. The use of input masking on the lesion classification network does not have any significant impact on the already highly accurate baseline model.

7 CONCLUSIONS AND FUTURE WORK

This section provides an overview of the conclusions drawn from this research, followed by its limitations and future scope for improvement.

7.1 Conclusions

This work introduces the Symmetry Enhanced Lesion Classification Network (SE-LCN), an MTL approach for skin lesion segmentation, lesion symmetry classification, and skin lesion disease classification. The main focus of this research was two-fold: analyzing the impact of input masking on symmetry classification, and analyzing the impact of transfer of CAMs on skin lesion classification.

This study utilized the SqueezeNet architecture for symmetry classification. Utilizing a network pre-trained on a larger dataset (Imagenet) and fine-tuning for symmetry classification outperformed

training a network from scratch. This method also outperformed existing works in the literature [16, 17] that trained CNN from scratch. While input masking did not improve the predictive performance of the model on the test set, it significantly improved the localization of the CAM generated by the model. Hence, using segmentation masks ensures that the activation points in an image lie within the lesion region. Having accurate CAMs indicates the reliability of model predictions, ensuring that the model focuses on the right parts of the image while making predictions. This could be a useful tool for diagnosticians who want to verify the credibility of certain predictions made by a model.

The SE-LCN utilizes the E-Layer [18] to integrate CAMs with features in the EfficienNet B4 model, which is trained to classify skin lesions. An ablation study was conducted to analyze the impact of CAM transfer and analyze the impact of different input masking strategies. The best results were achieved when a lesion classification model without input masking utilized CAMs from symmetry classification models that employed input masking. Evaluation on the HAM10000 dataset suggests that masking the inputs to the lesion classifier does not enhance predictive performance, whereas CAM integration improves it. Assessing the generalizability on the ISIC2019 dataset indicated that both input masking and CAM transfer improve generalizability. Additionally, including CAMs from symmetry classification models increased the Intersection over Union (IoU) score of the CAMs extracted from the classifier but decreased the percentage overlap and pointing game scores. This is because the inclusion of symmetry CAMs expanded the region of activation points in the classifier network, focusing on the edges of the lesion and surrounding pixels, as observed through visual inspection.

7.2 Limitations and Future Work

This research has several limitations that can be addressed in future work:

(1) Hyperparameter tuning for each experiment was conducted for the baseline model, and the same hyperparameters were used for every trained model within the experiment for a fair comparison. However, each model may have different optimal hyperparameter settings. It is evident from Figure 11, Figure 12, and Figure 13 that each model, when trained with the same hyperparameters, converges after a different number of epochs. Tuning the hyperparameters for each model before training could maximize the performance of each CAM transfer and masking strategy.
(2) The proposed SE-LCN model consists of three individual networks trained separately, making training and inference computationally expensive. Future work could investigate a symmetry-enhanced end-to-end classification model, which would be more computationally efficient.
(3) The segmentation model implemented in SE-LCN is a standalone model with high performance. However, it does not match the performance of state-of-the-art segmentation approaches. It would be interesting to analyze the impact of using a better segmentation model for masking both classifiers, especially when using segmentation masking.

Overall, this work presents a comprehensive analysis of how the combination of the inclusion of CAMs and the usage of segmentation masks impacts the performance of a cascaded MTL architecture. This analysis could be used in modifying the architecture and training models that could generate state-of-the-art results that provide explainable predictions.

References

[1] Mercedes Filho, Zhen Ma, and Joao Manuel RS Tavares. A review of the quantification and classification of pigmented skin lesions: from dedicated to hand-held devices. *Journal of medical systems*, 39:1–12, 2015.
[2] Dan Popescu, Mohamed El-Khatib, Hassan El-Khatib, and Loretta Ichim. New trends in melanoma detection using neural networks: a systematic review. *Sensors*, 22(2):496, 2022.
[3] Xiaoying Sun, Na Zhang, Chengqian Yin, Bo Zhu, and Xin Li. Ultraviolet radiation and melanomagenesis: from mechanism to immunotherapy. *Frontiers in Oncology*, 10:951, 2020.
[4] Hassan El-Khatib, Dan Popescu, and Loretta Ichim. Deep learning–based methods for automatic diagnosis of skin lesions. *Sensors*, 20(6):1753, 2020.
[5] Abder-Rahman A Ali and Thomas M Deserno. A systematic review of automated melanoma detection in dermatoscopic images and its ground truth data. *Medical Imaging 2012: Image Perception, Observer Performance, and Technology Assessment*, 8318:421–431, 2012.
[6] Adekanmi Adegun and Serestina Viriri. Deep learning techniques for skin lesion analysis and melanoma cancer detection: a survey of state-of-the-art. *Artificial Intelligence Review*, 54:811–841, 2021.
[7] Keiron O'Shea and Ryan Nash. An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*, 2015.
[8] Zahra Mirikharaji, Kumar Abhishek, Alceu Bissoto, Catarina Barata, Sandra Avila, Eduardo Valle, M Emre Celebi, and Ghassan Hamarneh. A survey on deep learning for skin lesion segmentation. *Medical Image Analysis*, page 102863, 2023.
[9] Robert J Friedman, Darrell S Rigel, and Alfred W Kopf. Early detection of malignant melanoma: the role of physician examination and self-examination of the skin. *CA: a cancer journal for clinicians*, 35(3):130–151, 1985.
[10] Naheed R Abbasi, Helen M Shaw, Darrell S Rigel, Robert J Friedman, William H McCarthy, Iman Osman, Alfred W Kopf, and David Polsky. Early diagnosis of cutaneous melanoma: revisiting the abcd criteria. *Jama*, 292(22):2771–2776, 2004.
[11] J Daniel Jensen and Boni E Elewski. The abcdef rule: combining the "abcde rule" and the "ugly duckling sign" in an effort to improve patient self-screening examinations. *The Journal of clinical and aesthetic dermatology*, 8(2):15, 2015.
[12] Giuseppe Argenziano, Gabriella Fabbrocini, Paolo Carli, Vincenzo De Giorgi, Elena Sammarco, and Mario Delfino. Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions: comparison of the abcd rule of dermatoscopy and a new 7-point checklist based on pattern analysis. *Archives of dermatology*, 134(12):1563–1570, 1998.
[13] Scott W Menzies, Christian Ingvar, Kerry A Crotty, and William H McCarthy. Frequency and morphologic characteristics of invasive melanomas lacking specific surface microscopic features. *Archives of dermatology*, 132(10):1178–1182, 1996.
[14] Giuseppe Argenziano, H Peter Soyer, Sergio Chimenti, Renato Talamini, Rosamaria Corona, Francesco Sera, Michael Binder, Lorenzo Cerroni, Gaetano De Rosa, Gerardo Ferrara, et al. Dermoscopy of pigmented skin lesions: results of a consensus meeting via the internet. *Journal of the American Academy of Dermatology*, 48(5):679–693, 2003.
[15] Vincent Toureau, Pedro Bibiloni, Lidia Talavera-Martínez, and Manuel González-Hidalgo. Automatic detection of symmetry in dermoscopic images based on shape and texture. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 625–636. Springer, 2020.
[16] Lidia Talavera-Martínez, Pedro Bibiloni, Aniza Giacaman, Rosa Taberner, Luis Javier Del Pozo Hernando, and Manuel González-Hidalgo. A novel approach for skin lesion symmetry classification with a deep learning model. *Computers in biology and medicine*, 145:105450, 2022.
[17] Lidia Talavera-Martínez, Pedro Bibiloni, Aniza Giacaman, Rosa Taberner, Luis Javier Del Pozo Hernando, and Manuel González-Hidalgo. Model regularisation for skin lesion symmetry classification: Symderm v2. 0. In *International Conference on Computer Analysis of Images and Patterns*, pages 99–109. Springer, 2023.
[18] Yutong Xie, Jianpeng Zhang, Yong Xia, and Chunhua Shen. A mutual bootstrapping model for automated skin lesion segmentation and classification. *IEEE transactions on medical imaging*, 39(7):2482–2493, 2020.

[19] Rich Caruana. Multitask learning. *Machine learning*, 28:41–75, 1997.

[20] Philipp Tschandl, Cliff Rosendahl, and Harald Kittler. The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific data*, 5(1):1–9, 2018.

[21] Partoo Vafaeikia, Khashayar Namdar, and Farzad Khalvati. A brief review of deep multi-task learning and auxiliary task learning. *arXiv preprint arXiv:2007.01126*, 2020.

[22] Yan Zhao, Xiuying Wang, Tongtong Che, Guoqing Bao, and Shuyu Li. Multi-task deep learning for medical image computing and analysis: A review. *Computers in Biology and Medicine*, 153:106496, 2023.

[23] Ting Gong, Tyler Lee, Cory Stephenson, Venkata Renduchintala, Suchismita Padhy, Anthony Ndirango, Gokce Keskin, and Oguz H Elibol. A comparison of loss weighting strategies for multi task learning in deep neural networks. *IEEE Access*, 7:141627–141632, 2019.

[24] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2921–2929, 2016.

[25] RR Selvaraju, M Cogswell, A Das, R Vedantam, D Parikh, and D Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. arxiv 2016. *arXiv preprint arXiv:1610.02391*, 2022.

[26] Adiya Chattopadhyay, Anirban Sarkar, Prantik Howlader, and Vineeth N Balasubramanian. Grad-cam: Improved visual explanations for deep convolutional networks. *arXiv: 1710.11063*, 2017.

[27] Haofan Wang, Zifan Wang, Mengnan Du, Fan Yang, Zijian Zhang, Sirui Ding, Piotr Mardziel, and Xia Hu. Score-cam: Score-weighted visual explanations for convolutional neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 24–25, 2020.

[28] Harish Guruprasad Ramaswamy et al. Ablation-cam: Visual explanations for deep convolutional network via gradient-free localization. In *proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 983–991, 2020.

[29] Jianming Zhang, Sarah Adel Bargal, Zhe Lin, Jonathan Brandt, Xiaohui Shen, and Stan Sclaroff. Top-down neural attention by excitation backprop. *International Journal of Computer Vision*, 126(10):1084–1102, 2018.

[30] M Emre Celebi, Hassan A Kingravi, Bakhtiyar Uddin, Hitoshi Iyatomi, Y Alp Aslandogan, William V Stoecker, and Randy H Moss. A methodological approach to the classification of dermoscopy images. *Computerized Medical imaging and graphics*, 31(6):362–373, 2007.

[31] Teresa Mendonça, Pedro M Ferreira, Jorge S Marques, André RS Marcal, and Jorge Rozeira. Ph 2-a dermoscopic image database for research and benchmarking. In *2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, pages 5437–5440. IEEE, 2013.

[32] Lucia Ballerini, Robert B Fisher, Ben Aldridge, and Jonathan Rees. A color and texture based hierarchical k-nn approach to the classification of non-melanoma skin lesions. *Color medical image analysis*, pages 63–86, 2013.

[33] David Gutman, Noel CF Codella, Emre Celebi, Brian Helba, Michael Marchetti, Nabin Mishra, and Allan Halpern. Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (isbi) 2016, hosted by the international skin imaging collaboration (isic). *arXiv preprint arXiv:1605.01397*, 2016.

[34] NCF Codella, D Gutman, ME Celebi, B Helba, MA Marchetti, SW Dusza, A Kalloo, K Liopyris, N Mishra, and H Kittler. Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging. *Int Skin Imag Collab (ISIC). arXiv Prepr arXiv171005006*, 2017.

[35] Noel Codella, Veronica Rotemberg, Philipp Tschandl, M Emre Celebi, Stephen Dusza, David Gutman, Brian Helba, Aadi Kalloo, Konstantinos Liopyris, Michael Marchetti, et al. Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic). *arXiv preprint arXiv:1902.03368*, 2019.

[36] Philipp Tschandl, Christoph Rinner, Zoe Apalla, Giuseppe Argenziano, Noel Codella, Allan Halpern, Monika Janda, Aimilios Lallas, Caterina Longo, Josep Malvehy, et al. Human–computer collaboration for skin cancer recognition. *Nature Medicine*, 26(8):1229–1234, 2020.

[37] Bill Cassidy, Connah Kendrick, Andrzej Brodzicki, Joanna Jaworek-Korjakowska, and Moi Hoon Yap. Analysis of the isic image datasets: Usage, benchmarks and recommendations. *Medical image analysis*, 75:102305, 2022.

[38] Lidia Talavera-Martinez, Pedro Bibiloni, and Manuel Gonzalez-Hidalgo. Hair segmentation and removal in dermoscopic images using deep learning. *IEEE Access*, 9:2694–2704, 2020.

[39] Giuseppe Argenziano, HP Soyer, Vincenzo De Giorgi, Domenico Piccolo, Paolo Carli, Mario Delfino, et al. Interactive atlas of dermoscopy (book and cd-rom). 2000.

[40] Veronica Rotemberg, Nicholas Kurtansky, Brigid Betz-Stablein, Liam Caffery, Emmanouil Chousakos, Noel Codella, Marc Combalia, Stephen Dusza, Pascale Guitera, David Gutman, et al. A patient-centric dataset of images and metadata for identifying melanomas using clinical context. *Scientific data*, 8(1):34, 2021.

[41] Feng Ning, Damien Delhomme, Yann LeCun, Fabio Piano, Léon Bottou, and Paolo Emilio Barbano. Toward automatic phenotyping of developing embryos from videos. *IEEE Transactions on Image Processing*, 14(9):1360–1371, 2005.

[42] Jonathan Long, Evan Shelhamer, Trevor Darrell, and UC Berkeley. Fully convolutional networks for semantic segmentation. arxiv 2015. *arXiv preprint arXiv:1411.4038*, 2014.

[43] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015.

[44] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[45] Noel CF Codella, Q Nguyen, Sharath Pankanti, David Gutman, Brian Helba, Allan Halpern, and John R Smith. Deep learning ensembles for melanoma recognition in dermoscopy images. corr abs/1610.04662 (2016), 2016.

[46] D Roja Ramani and S Siva Ranjani. U-net based segmentation and multiple feature extraction of dermascopic images for efficient diagnosis of melanoma. In *Computer Aided Intervention and Diagnostics in Clinical and Medical Images*, pages 81–101. Springer, 2019.

[47] Peng Tang, Qiaokang Liang, Xintong Yan, Shao Xiang, Wei Sun, Dan Zhang, and Gianmarc Coppola. Efficient skin lesion segmentation using separable-unet with stochastic weight averaging. *Computer methods and programs in biomedicine*, 178:289–301, 2019.

[48] Md Kamrul Hasan, Lavsen Dahal, Prasad N Samarakoon, Fakrul Islam Tushar, and Robert Martí. Dsnet: Automatic dermoscopic skin lesion segmentation. *Computers in biology and medicine*, 120:103738, 2020.

[49] Md Zahangir Alom, Chris Yakopcic, Mahmudul Hasan, Tarek M Taha, and Vijayan K Asari. Recurrent residual u-net for medical image segmentation. *Journal of medical imaging*, 6(1):014006–014006, 2019.

[50] Md Zahangir Alom, Theus Aspiras, Tarek M Taha, and Vijayan K Asari. Skin cancer segmentation and classification with improved deep convolutional neural network. In *Medical Imaging 2020: Imaging informatics for healthcare, research, and applications*, volume 11318, pages 291–301. SPIE, 2020.

[51] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. arxiv 2016. *arXiv preprint arXiv:1606.00915*, 1, 2016.

[52] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017.

[53] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018.

[54] Manu Goyal, Amanda Oakley, Priyanka Bansal, Darren Dancey, and Moi Hoon Yap. Automatic lesion boundary segmentation in dermoscopic images with ensemble deep learning methods. *arXiv preprint arXiv:1902.00809*, 2019.

[55] Zhiying Cui, Longshi Wu, Ruixuan Wang, and Wei-Shi Zheng. Ensemble transductive learning for skin lesion segmentation. In *Pattern Recognition and Computer Vision: Second Chinese Conference, PRCV 2019, Xi'an, China, November 8–11, 2019, Proceedings, Part II 2*, pages 572–581. Springer, 2019.

[56] Sheng Chen, Zhe Wang, Jianping Shi, Bin Liu, and Nenghai Yu. A multi-task framework with feature passing module for skin lesion classification and segmentation. In *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, pages 1126–1129. IEEE, 2018.

[57] Omer Sagi and Lior Rokach. Ensemble learning: A survey. *Wiley interdisciplinary reviews: data mining and knowledge discovery*, 8(4):e1249, 2018.

[58] Teck Yan Tan, Li Zhang, Chee Peng Lim, Ben Fielding, Yonghong Yu, and Emma Anderson. Evolving ensemble models for image segmentation using enhanced particle swarm optimization. *IEEE access*, 7:34004–34019, 2019.

[59] Yading Yuan, Ming Chao, and Yeh-Chi Lo. Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance. *IEEE transactions on medical imaging*, 36(9):1876–1886, 2017.

[60] Yading Yuan and Yeh-Chi Lo. Improving dermoscopic image segmentation with enhanced convolutional-deconvolutional networks. *IEEE journal of biomedical and health informatics*, 23(2):519–526, 2017.

[61] Federico Pollastri, Federico Bolelli, Roberto Paredes, and Costantino Grana. Augmenting data with gans to segment melanoma skin lesions. *Multimedia Tools and Applications*, 79(21):15575–15592, 2020.

[62] Laura Canalini, Federico Pollastri, Federico Bolelli, Michele Cancilla, Stefano Allegretti, and Costantino Grana. Skin lesion segmentation ensemble with diverse training strategies. In *Computer Analysis of Images and Patterns: 18th International Conference, CAIP 2019, Salerno, Italy, September 3–5, 2019, Proceedings, Part I 18*, pages 89–101. Springer, 2019.

[63] VR Balaji, ST Suganthi, R Rajadevi, V Krishna Kumar, B Saravana Balaji, and Sanjeevi Pandiyan. Skin disease detection and segmentation using dynamic graph

cut algorithm and classification through naive bayes classifier. *Measurement*, 163:107922, 2020.

[64] Karar Ali, Zaffar Ahmed Shaikh, Abdullah Ayub Khan, and Asif Ali Laghari. Multiclass skin cancer classification using efficientnets–a first step towards preventing skin cancer. *Neuroscience Informatics*, 2(4):100034, 2022.

[65] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019.

[66] Parvathaneni Naga Srinivasu, Jalluri Gnana SivaSai, Muhammad Fazal Ijaz, Akash Kumar Bhoi, Wonjoon Kim, and James Jin Kang. Classification of skin disease using deep learning neural networks with mobilenet v2 and lstm. *Sensors*, 21(8):2852, 2021.

[67] Arushi Jain, Annavarapu Chandra Sekhara Rao, Praphula Kumar Jain, and Ajith Abraham. Multi-type skin diseases classification using op-dnn based feature extraction approach. *Multimedia Tools and Applications*, pages 1–26, 2022.

[68] Shenyi Qian, Kunpeng Ren, Weiwei Zhang, and Haohan Ning. Skin lesion classification using cnns with grouping of multi-scale attention and class-specific loss weighting. *Computer Methods and Programs in Biomedicine*, 226:107166, 2022.

[69] Katsuhiro Nakai, Yen-Wei Chen, and Xian-Hua Han. Enhanced deep bottleneck transformer model for skin lesion classification. *Biomedical Signal Processing and Control*, 78:103997, 2022.

[70] Saleh Naif Almuayqil, Sameh Abd El-Ghany, and Mohammed Elmogy. Computer-aided diagnosis for early signs of skin diseases using multi types feature fusion based on a hybrid deep learning model. *Electronics*, 11(23):4009, 2022.

[71] Dasari Anantha Reddy, Swarup Roy, Sanjay Kumar, and Rakesh Tripathi. A scheme for effective skin disease detection using optimized region growing segmentation and autoencoder based classification. *Procedia Computer Science*, 218:274–282, 2023.

[72] Areej A Malibari, Jaber S Alzahrani, Majdy M Eltahir, Vinita Malik, Marwa Obayya, Mesfer Al Duhayyim, Aloísio V Lira Neto, and Victor Hugo C de Albuquerque. Optimal deep neural network-driven computer aided diagnosis model for skin cancer. *Computers and Electrical Engineering*, 103:108318, 2022.

[73] Sk Imran Hossain, Jocelyn de Goër de Herve, Md Shahriar Hassan, Delphine Martineau, Evelina Petrosyan, Violaine Corbin, Jean Beytout, Isabelle Lebert, Jonas Durand, Irene Carravieri, et al. Exploring convolutional neural networks with transfer learning for diagnosing lyme disease from skin lesion images. *Computer Methods and Programs in Biomedicine*, 215:106624, 2022.

[74] Bhuvaneshwari Shetty, Roshan Fernandes, Anisha P Rodrigues, Rajeswari Chengoden, Sweta Bhattacharya, and Kuruva Lakshmanna. Skin lesion classification of dermoscopic images using machine learning and convolutional neural network. *Scientific Reports*, 12(1):18134, 2022.

[75] Mingjun Wei, Qiwei Wu, Hongyu Ji, Jingkun Wang, Tao Lyu, Jinyun Liu, and Li Zhao. A skin disease classification model based on densenet and convnext fusion. *Electronics*, 12(2):438, 2023.

[76] Md Jahin Alam, Mir Sayeed Mohammad, Md Adnan Faisal Hossain, Ishtiaque Ahmed Showmik, Munshi Sanowar Raihan, Shahed Ahmed, and Talha Ibn Mahmud. S2c-delenet: A parameter transfer based segmentation-classification integration for detecting skin cancer lesions from dermoscopic images. *Computers in Biology and Medicine*, 150:106148, 2022.

[77] Mohamed A Elashiri, Arunachalam Rajesh, Surya Nath Pandey, Surendra Kumar Shukla, Shabana Urooj, et al. Ensemble of weighted deep concatenated features for the skin disease classification model using modified long short term memory. *Biomedical Signal Processing and Control*, 76:103729, 2022.

[78] Fayadh Alenezi, Ammar Armghan, and Kemal Polat. Wavelet transform based deep residual neural network and relu based extreme learning machine for skin lesion classification. *Expert Systems with Applications*, 213:119064, 2023.

[79] Qirui Huang, Huan Ding, and Fatima Rashid Sheykhahmad. A skin cancer diagnosis system for dermoscopy images according to deep training and metaheuristics. *Biomedical Signal Processing and Control*, 83:104705, 2023.

[80] B Kalpana, AK Reshmy, S Senthil Pandi, and S Dhanasekaran. Oesv-krf: Optimal ensemble support vector kernel random forest based early detection and classification of skin diseases. *Biomedical Signal Processing and Control*, 85:104779, 2023.

[81] Ioannis Kousis, Isidoros Perikos, Ioannis Hatzilygeroudis, and Maria Virvou. Deep learning methods for accurate skin cancer recognition and mobile application. *Electronics*, 11(9):1294, 2022.

[82] Belal Ahmad, Mohd Usama, Tanvir Ahmad, Shabnam Khatoon, and Chaudhary Maqbool Alam. An ensemble model of convolution and recurrent neural network for skin disease classification. *International Journal of Imaging Systems and Technology*, 32(1):218–229, 2022.

[83] Sarmad Maqsood and Robertas Damaševičius. Multiclass skin lesion localization and classification using deep learning based features fusion and selection framework for smart healthcare. *Neural networks*, 160:238–258, 2023.

[84] MF Healsmith, JF Bourke, JE Osborne, and RAC Graham-Brown. An evaluation of the revised seven-point checklist for the early diagnosis of cutaneous malignant melanoma. *British Journal of Dermatology*, 130(1):48–50, 1994.

[85] Stefania Seidenari, Giovanni Pellacani, and Costantino Grana. Asymmetry in dermoscopic melanocytic lesion images: a computer description based on colour distribution. *Acta dermato-venereologica*, 86:123–128, 2006.

[86] Kathy M Clawson, Philip J Morrow, Bryan W Scotney, D John McKenna, and Olivia M Dolan. Determination of optimal axes for skin lesion asymmetry quantification. In *2007 IEEE international conference on image processing*, volume 2, pages II–453. IEEE, 2007.

[87] Arve Kjoelen, Marc J Thompson, Scott E Umbaugh, Randy Hays Moss, and William V Stoecker. Performance of ai methods in detecting melanoma. *IEEE Engineering in Medicine and Biology Magazine*, 14(4):411–416, 1995.

[88] Abder-Rahman Ali, Jingpeng Li, and Sally Jane O'Shea. Towards the automatic detection of skin lesion shape asymmetry, color variegation and diameter in dermoscopic images. *Plos one*, 15(6):e0234352, 2020.

[89] Lequan Yu, Hao Chen, Qi Dou, Jing Qin, and Pheng-Ann Heng. Automated melanoma recognition in dermoscopy images via very deep residual networks. *IEEE transactions on medical imaging*, 36(4):994–1004, 2016.

[90] Ivan Gonzalez-Diaz. Dermaknet: Incorporating the knowledge of dermatologists to convolutional neural networks for skin lesion diagnosis. *IEEE journal of biomedical and health informatics*, 23(2):547–559, 2018.

[91] Mohammed A Al-Masni, Dong-Hyun Kim, and Tae-Seong Kim. Multiple skin lesions diagnostics via integrated deep convolutional networks for segmentation and classification. *Computer methods and programs in biomedicine*, 190:105351, 2020.

[92] Sulaiman Vesal, Shreyas Malakarjun Patil, Nishant Ravikumar, and Andreas K Maier. A multi-task framework for skin lesion detection and segmentation. In *International Workshop on Computer-Assisted and Robotic Endoscopy*, pages 285–293. Springer, 2018.

[93] Sulaiman Vesal, Nishant Ravikumar, and Andreas Maier. Skinnet: A deep learning framework for skin lesion segmentation. In *2018 IEEE nuclear science symposium and medical imaging conference proceedings (NSS/MIC)*, pages 1–3. IEEE, 2018.

[94] Xulei Yang, Zeng Zeng, Si Yong Yeo, Colin Tan, Hong Liang Tey, and Yi Su. A novel multi-task deep learning model for skin lesion segmentation and classification. *arXiv preprint arXiv:1703.01025*, 2017.

[95] Lei Song, Jianzhe Lin, Z Jane Wang, and Haoqian Wang. An end-to-end multi-task deep learning framework for skin lesion analysis. *IEEE journal of biomedical and health informatics*, 24(10):2912–2921, 2020.

[96] Xiaoyu He, Yong Wang, Shuang Zhao, and Xiang Chen. Joint segmentation and classification of skin lesions via a multi-task learning convolutional neural network. *Expert Systems with Applications*, page 120174, 2023.

[97] Carole H Sudre, Wenqi Li, Tom Vercauteren, Sebastien Ourselin, and M Jorge Cardoso. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings 3*, pages 240–248. Springer, 2017.

[98] Forrest N Iandola, Song Han, Matthew W Moskewicz, Khalid Ashraf, William J Dally, and Kurt Keutzer. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and< 0.5 mb model size. *arXiv preprint arXiv:1602.07360*, 2016.

[99] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[100] Hritam Basak, Rohit Kundu, and Ram Sarkar. Mfsnet: A multi focus segmentation network for skin lesion segmentation. *Pattern Recognition*, 128:108673, 2022.

[101] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[102] Jasper Snoek, Hugo Larochelle, and Ryan P Adams. Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems*, 25, 2012.

[103] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[104] Dan Hendrycks, Kimin Lee, and Mantas Mazeika. Using pre-training can improve model robustness and uncertainty. In *International conference on machine learning*, pages 2712–2721. PMLR, 2019.

A Training Plots

This section presents the training plots of the first fold of 12 masking combinations of the SE-LCN network.

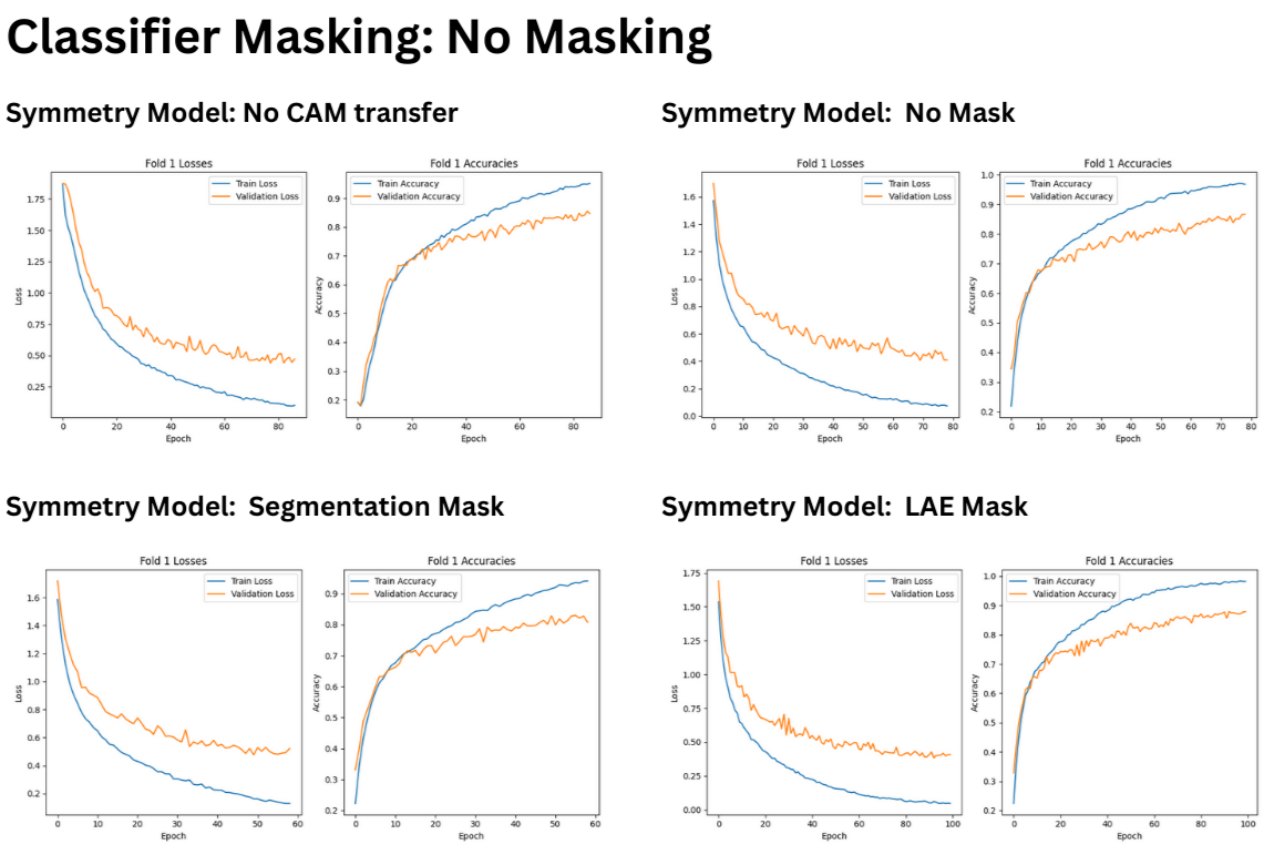


Figure 11: Training plots of the first fold of four models that use no input masking for the lesion classifier model. The training and validation loss and accuracies have been plotted across number of epochs.

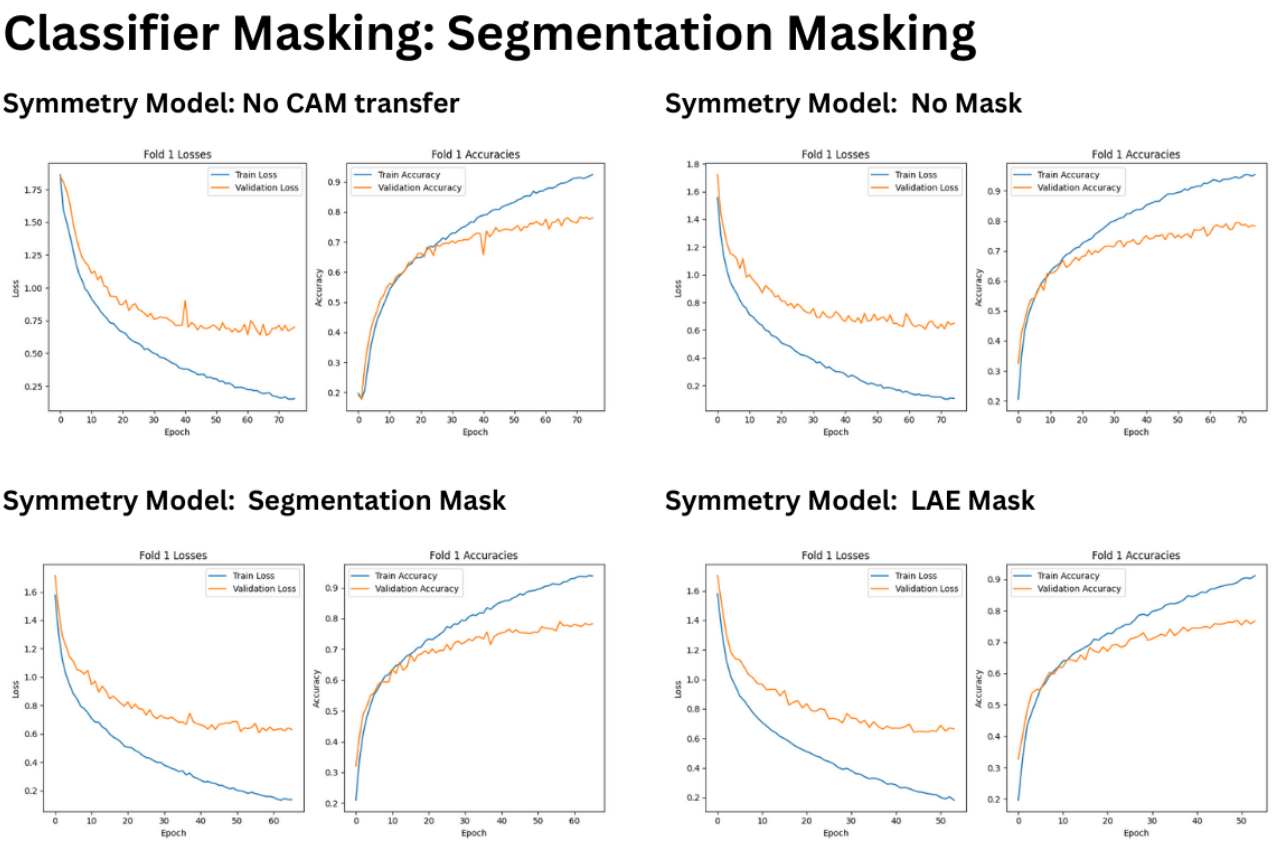


Figure 12: Training plots of the first fold of four models that use segmentation masking for the lesion classifier model. The training and validation loss and accuracies have been plotted across number of epochs.

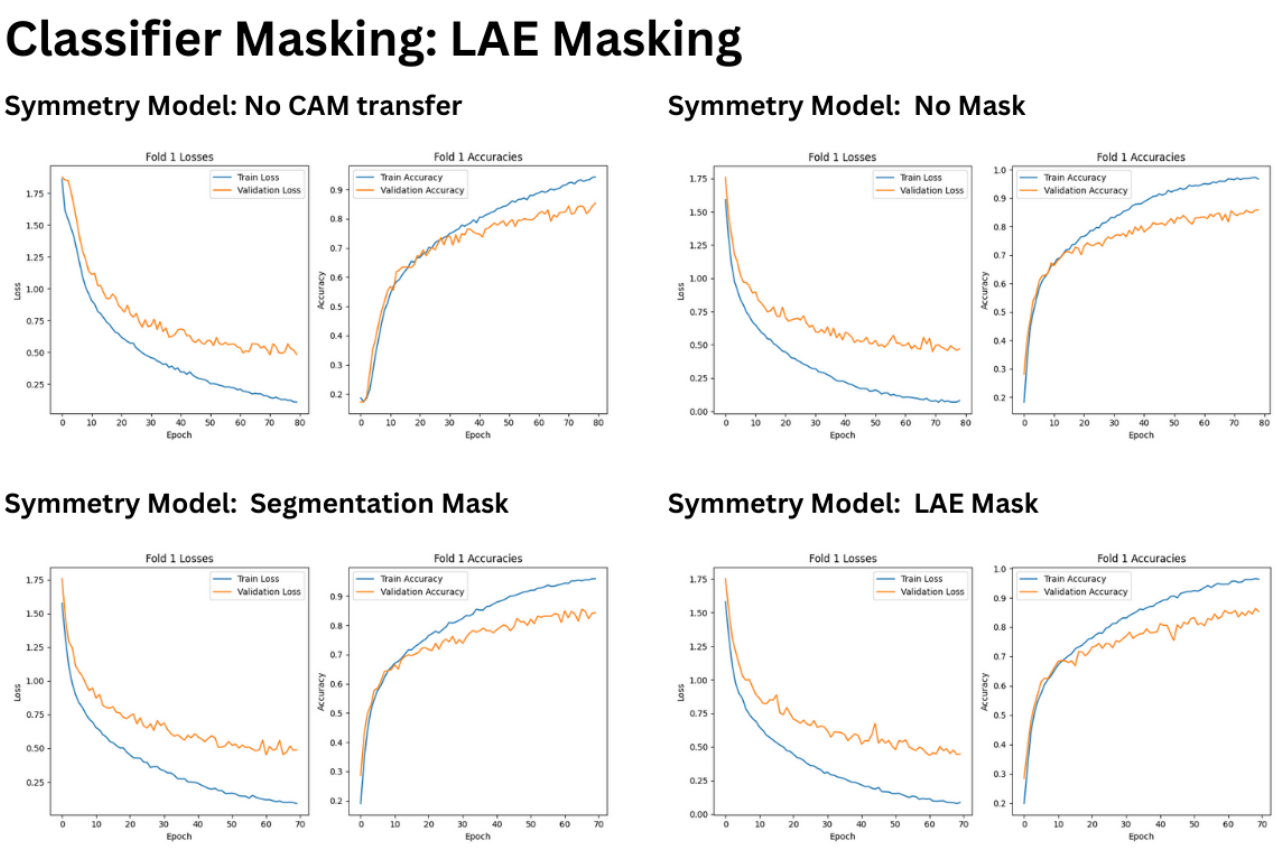


Figure 13: Training plots of the first fold of four models that use LAE input masking for the lesion classifier model. The training and validation loss and accuracies have been plotted across number of epochs.

B Class-wise Evaluation

Table 11: Class-wise classification metrics mean and standard deviation for model with lesion classifier having no mask and no transfer of CAMs (model 1) evaluated on the test set.

Class	Precision	Recall	F1-Score	Support
Nevus	0.94 ± 0.0071	0.90 ± 0.0158	0.92 ± 0.0071	1337
Melanoma	0.58 ± 0.0187	0.6875 ± 0.0228	0.6275 ± 0.0148	223
Benign keratosis-like lesions	0.725 ± 0.0229	0.6975 ± 0.0259	0.71 ± 0.0000	220
Basal cell carcinoma	0.75 ± 0.0596	0.865 ± 0.0296	0.8025 ± 0.0286	103
Actinic keratoses	0.765 ± 0.0403	0.815 ± 0.0572	0.7775 ± 0.0179	65
Vascular lesions	0.8025 ± 0.0622	0.875 ± 0.0150	0.835 ± 0.0364	28
Dermatofibroma	0.81 ± 0.0212	0.905 ± 0.0087	0.865 ± 0.0180	23

Table 12: Class-wise classification metrics mean and standard deviation for model with lesion classifier having no mask and symmetry classifier having segmentation mask (model 3) evaluated on the test set.

Class	Precision	Recall	F1-Score	Support
Nevus	0.9425 ± 0.0083	0.8925 ± 0.0109	0.9175 ± 0.0043	1341
Melanoma	0.5725 ± 0.0192	0.6850 ± 0.0287	0.6225 ± 0.0109	223
Benign keratosis-like lesions	0.7100 ± 0.0361	0.7100 ± 0.0354	0.7075 ± 0.0083	220
Basal cell carcinoma	0.7675 ± 0.0130	0.8400 ± 0.0224	0.8000 ± 0.0122	103
Actinic keratoses	0.7700 ± 0.0458	0.7775 ± 0.0642	0.7700 ± 0.0300	65
Vascular lesions	0.7550 ± 0.0335	0.9025 ± 0.0295	0.8225 ± 0.0192	28
Dermatofibroma	0.7350 ± 0.0702	0.8900 ± 0.0200	0.8000 ± 0.0464	23

C SE-LCN on ISIC2016 dataset

Table 13: Performance comparison of SE-LCN models trained and evaluated on the ISIC2016 dataset, employing various masking strategies on inputs of both the symmetry classification network and the lesion classification network. Each model's performance is quantified using five metrics, with values representing the mean and standard deviation calculated on the corresponding validation sets across 4 folds of cross-validation. The highest value of each metric among a lesion classifier masking strategy is highlighted in bold, and the highest value of each metric among all experiments is highlighted in blue text. The baseline is model 1.

Model			ISIC 2016 Validation Dataset (2 classes) *Mean ± Std. Dev*				
Model	**Masking**	**Symmetry Model CAM**	**Accuracy**	**Kappa Score**	**Macro Avg Precision**	**Macro Avg Recall**	**Macro Avg F1**
1	Without Masks	No CAM Transfer	0.788 ± 0.024	**0.392 ± 0.045**	0.682 ± 0.025	**0.719 ± 0.017**	0.694 ± 0.023
2	Without Masks	Without Masks	0.791 ± 0.024	0.365 ± 0.083	0.674 ± 0.038	0.694 ± 0.048	0.682 ± 0.041
3	Without Masks	Exact Segment Masks	0.776 ± 0.018	0.345 ± 0.081	0.660 ± 0.034	0.693 ± 0.051	0.671 ± 0.040
4	Without Masks	LAE Masks	**0.797 ± 0.034**	0.383 ± 0.095	**0.684 ± 0.048**	0.702 ± 0.049	**0.691 ± 0.047**
5	Exact Segment Masks	No CAM Transfer	0.773 ± 0.039	**0.366 ± 0.069**	0.669 ± 0.034	**0.708 ± 0.030**	**0.680 ± 0.037**
6	Exact Segment Masks	Without Masks	0.787 ± 0.017	0.351 ± 0.041	0.668 ± 0.022	0.685 ± 0.019	0.675 ± 0.021
7	Exact Segment Masks	Exact Segment Masks	0.781 ± 0.017	0.340 ± 0.052	0.661 ± 0.024	0.682 ± 0.029	0.669 ± 0.026
8	Exact Segment Masks	LAE Masks	**0.787 ± 0.015**	0.353 ± 0.031	**0.669 ± 0.016**	0.687 ± 0.017	0.676 ± 0.016
9	LAE Masks	No CAM Transfer	0.810 ± 0.017	0.448 ± 0.037	0.709 ± 0.019	0.748 ± 0.022	0.723 ± 0.019
10	LAE Masks	Without Masks	0.820 ± 0.024	0.450 ± 0.058	0.716 ± 0.030	0.737 ± 0.029	0.725 ± 0.029
11	LAE Masks	Exact Segment Masks	**0.827 ± 0.014**	0.456 ± 0.031	**0.725 ± 0.018**	0.734 ± 0.021	0.728 ± 0.015
12	LAE Masks	LAE Masks	0.824 ± 0.017	**0.458 ± 0.035**	0.723 ± 0.018	**0.739 ± 0.024**	**0.729 ± 0.018**

Table 14: Performance comparison of SE-LCN models trained and evaluated on the ISIC2016 dataset, employing various masking strategies on inputs of both the symmetry classification network and the lesion classification network. Each model's performance is quantified using five metrics, with values representing the mean and standard deviation calculated on a separate test set across 4 folds of cross-validation. The highest value of each metric among a lesion classifier masking strategy is highlighted in bold, and the highest value of each metric among all experiments is highlighted in blue text. The baseline is model 1.

Model			ISIC 2016 Test Dataset (2 classes) *Mean ± Std. Dev*				
Model	**Masking**	**Symmetry Model CAM**	**Accuracy**	**Kappa Score**	**Macro Avg Precision**	**Macro Avg Recall**	**Macro Avg F1**
1	Without Masks	No CAM Transfer	0.785 ± 0.023	0.392 ± 0.037	0.683 ± 0.021	0.717 ± 0.011	0.694 ± 0.020
2	Without Masks	Without Masks	**0.807 ± 0.013**	**0.418 ± 0.032**	**0.702 ± 0.017**	**0.719 ± 0.016**	**0.709 ± 0.016**
3	Without Masks	Exact Segment Masks	0.799 ± 0.020	0.411 ± 0.041	0.695 ± 0.024	0.719 ± 0.016	0.705 ± 0.021
4	Without Masks	LAE Masks	0.802 ± 0.011	0.410 ± 0.021	0.696 ± 0.012	0.716 ± 0.011	0.704 ± 0.011
5	Exact Segment Masks	No CAM Transfer	0.768 ± 0.028	**0.362 ± 0.042**	**0.668 ± 0.021**	**0.706 ± 0.023**	**0.677 ± 0.023**
6	Exact Segment Masks	Without Masks	**0.774 ± 0.016**	0.356 ± 0.023	0.667 ± 0.012	0.696 ± 0.014	0.676 ± 0.012
7	Exact Segment Masks	Exact Segment Masks	0.772 ± 0.004	0.345 ± 0.021	0.661 ± 0.008	0.689 ± 0.015	0.671 ± 0.010
8	Exact Segment Masks	LAE Masks	0.770 ± 0.006	0.338 ± 0.020	0.658 ± 0.009	0.685 ± 0.014	0.668 ± 0.010
9	LAE Masks	No CAM Transfer	0.782 ± 0.019	0.388 ± 0.026	0.680 ± 0.016	0.717 ± 0.012	0.692 ± 0.014
10	LAE Masks	Without Masks	0.815 ± 0.012	**0.438 ± 0.017**	**0.714 ± 0.013**	**0.727 ± 0.009**	**0.719 ± 0.009**
11	LAE Masks	Exact Segment Masks	**0.818 ± 0.010**	0.430 ± 0.018	0.714 ± 0.013	0.716 ± 0.006	0.715 ± 0.009
12	LAE Masks	LAE Masks	0.807 ± 0.017	0.416 ± 0.021	0.703 ± 0.016	0.716 ± 0.011	0.707 ± 0.011

Table 15: Evaluation of localization of CAMs of SE-LCN trained on the ISIC2016 dataset, employing various masking strategies. The best fold (from cross-validation) of each model is used for evaluation on the HAM10000 dataset. Metrics shown include Intersection over Union (IoU), Percentage Overlap, and Pointing Game scores, utilizing segmentation masks for reference comparison. The highest value of each metric among a lesion classifier masking strategy is highlighted in bold. The baseline model is model 1, which uses no input masking and no transfer of CAMs

Model			CAM Localization *Entire Dataset*		
Model	**Classification Model**	**Symmetry Model**	**IoU**	**Overlap**	**Point Game**
1	Without Masks	No CAM transfer	0.015 ± 0.034	0.134 ± 0.249	0.033
2	Without Masks	Without Masks	0.025 ± 0.056	0.171 ± 0.287	0.072
3	Without Masks	Segmentation Masks	0.032 ± 0.063	0.212 ± 0.316	0.11
4	Without Masks	LAE Masks	**0.041 ± 0.078**	**0.223 ± 0.311**	**0.12**
5	Segmentation Masks	No CAM transfer	0.048 ± 0.093	0.232 ± 0.314	0.063
6	Segmentation Masks	Without Masks	0.114 ± 0.118	0.505 ± 0.355	0.293
7	Segmentation Masks	Segmentation Masks	0.122 ± 0.133	0.500 ± 0.345	0.251
8	Segmentation Masks	LAE Masks	**0.139 ± 0.131**	**0.562 ± 0.352**	**0.408**
9	LAE Masks	No CAM transfer	0.068 ± 0.094	0.341 ± 0.340	0.177
10	LAE Masks	Without Masks	**0.123 ± 0.124**	**0.527 ± 0.343**	**0.342**
11	LAE Masks	Segmentation Masks	0.116 ± 0.123	0.492 ± 0.337	0.281
12	LAE Masks	LAE Masks	0.061 ± 0.091	0.324 ± 0.344	0.153