I.J. Image, Graphics and Signal Processing, **2025, 2, 82-93**
Published Online on April 8, 2025 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijigsp.2025.02.05

Design of an Efficient UNet-Based Transfer Learning Model for Enhancing Skin Cancer Segmentation and Classification Performance

Namrata Verma*
Department of Electronics & Telecommunication, Rungta College of Engineering & Technology Bhilai, Chhattisgarh, India
Email: namratadewangan29@gmail.com
ORCID iD: https://orcid.org/0000-0001-7936-6416
*Corresponding Author

Pankaj Kumar Mishra
Department of Biomedical Engineering and Bioinformatics, University Teaching Department, Chhattisgarh Swami Vivekanand Technical University, Bhilai, Chhattisgarh, India
Email: pmishra1974@yahoo.co.in
ORCID iD: https://orcid.org/0000-0003-0129-2570

Received: 25 September, 2023; Revised: 22 November, 2023; Accepted: 12 January, 2025; Published: 08 April, 2025

Abstract: Accurate and efficient segmentation and classification are indispensable for the early diagnosis and treatment of skin cancer, a common and potentially fatal condition. Combining the UNet architecture with Auto Encoders for robust skin cancer segmentation, followed by binary cascade Convolutional Neural Networks (CNNs). In this text, we present a novel method for accurately classifying melanoma and basal cell carcinoma. Existing models are limited in their ability to achieve high precision, accuracy, and recall rates while maintaining a high Peak Signal-to-Noise Ratio (PSNR) for accurate image reconstruction, which necessitates this research. Our proposed model overcomes these limitations and performs exceptionally well on datasets: ISIC, HAM10000, PH2 Dataset, and Dermofit Image Libraries. When UNet and Auto Encoders are used, the advantages of both architectures are combined. The UNet architecture, renowned for its superior performance in image segmentation tasks, provides a solid foundation for separating skin cancer regions from surrounding tissue. The Auto Encoder component simultaneously facilitates feature extraction and image reconstruction, leading to improved representation learning and segmentation results. Utilizing the complementary capabilities of these models, our method improves the accuracy and efficiency of skin cancer segmentations. Using binary cascade CNNs for classification also improves our model's performance. The binary cascade architecture employs a hierarchical classification method that iteratively improves classification choices at each stage. This facilitates the differentiation between basal cell carcinoma, melanoma, and melanocytic nevi, resulting in highly accurate and trustworthy predictions. Extensive experiments were conducted on the ISIC, HAM10000, PH2 Dataset, and Dermofit Image Library to evaluate the performance of our proposed model. The achieved precision of 99.2%, accuracy of 98.3%, recall of 98.9%, and PSNR greater than 42dB demonstrate the superior functionality and effectiveness of our strategy. These results suggest that our model has a great deal of potential for assisting dermatologists in the early identification and classification of skin cancer, ultimately leading to improved patient outcomes. The combination of UNet with Auto Encoders and binary cascade CNNs has proven effective for segmenting and classifying skin cancer. Our proposed model outperforms current methods in terms of precision, accuracy, recall, and PSNR, demonstrating its potential to have a significant impact on the field of dermatology and aid in the early detection and treatment of skin cancers.

Index Terms: Skin Cancer, CNN, Binary Cascade CNN, UNet, Auto Encoders

1. Introduction

The incidence of skin cancer, one of the most prevalent types of cancer in the world, is rapidly increasing. With early and accurate diagnosis, skin cancer mortality rates can be significantly reduced and patient outcomes improved. Dermatologists must manually inspect skin lesions, which is time-consuming and prone to error. In order to aid

dermatologists in the diagnostic process, there is a growing demand for automated and efficient methods for segmenting and classifying skin cancer types via Optimized Convolutional Neural Networks (OCNNs) [1 - 3].

Deep learning techniques have recently proven to be extraordinarily effective in a wide range of computer vision tasks, including medical image analysis. Convolutional Neural Networks (CNNs) have become effective image classification tools, while architectures such as the UNet excel at image segmentation. Existing models still struggle to achieve high levels of precision, accuracy, and recall while maintaining high levels of image reconstruction. In response to these challenges, we propose an innovative method that combines the UNet architecture with Auto Encoders and binary cascade CNNs for skin cancer segmentation. By leveraging the benefits of these architectures, our method aims to outperform existing models on a variety of datasets, thus overcoming their limitations. Improving the accuracy and efficiency of skin cancer segmentation and classification is the primary objective of this research. By precisely delineating the cancerous regions in skin images, our model can assist dermatologists in identifying and diagnosing melanocytic nevi, melanoma, and basal cell carcinoma. Early detection of cancer is essential for prompt intervention and effective treatment. We propose a novel combination of the UNet architecture and Auto Encoders to improve the segmentation of skin cancer, Auto Encoder improves feature extraction and image reconstruction, while UNet provides a stable framework for segmenting skin cancer regions. The segmentation results are enhanced as a result of such fusions. In order to accurately classify the various types of skin cancer, we use binary cascade CNNs for classification. The hierarchical classification approach of binary cascade CNNs permits incremental decision refinement, thereby enhancing the ability to differentiate between basal cell carcinoma and melanoma. This improves the reliability and precision of the classification procedure. We evaluate our proposed model on three distinct datasets: HAM10000, PH2 Dataset, and Dermofit Image Library. These datasets contain numerous images of skin cancer, allowing us to thoroughly evaluate and validate the efficacy of our model. Extensive experiments enable us to achieve exceptional performance metrics, such as 99.2% precision, s98.3% accuracy, 98.9% recall, and PSNR greater than 42dB. These results demonstrate that our proposed model has the potential for clinical application in the diagnosis of skin cancer and outperforms current methods. This paper concludes by addressing the pressing need for accurate segmentation and classification of skin cancer. Combining UNet with Auto Encoders for segmentation and binary cascade CNNs for classification enables us to demonstrate significant advancements in the field. The evaluation of our method using multiple datasets and the exceptional performance metrics obtained validate its effectiveness. This research has the potential to significantly improve patient outcomes and survival rates by aiding dermatologists in the early detection and diagnosis of skin cancers. The proposed model makes several novel contributions in the field of skin cancer segmentation and classification, addressing specific problems and limitations of existing models. These unique aspects of the proposed model can be explicitly described as follows:

(1) **Integration of UNet, Auto Encoders, and Binary Cascade CNNs**: The proposed model introduces a novel integration of three powerful architectural components—UNet, Auto Encoders, and Binary Cascade CNNs. This combination addresses the challenge of achieving both accurate skin cancer segmentation and classification within a single unified framework. By leveraging the strengths of these architectures, the model provides a comprehensive solution to the segmentation and classification tasks, improving the overall performance.

(2) **Enhanced Representation Learning**: The inclusion of Auto Encoders in the model enhances feature extraction and representation learning. This contributes to better feature embeddings, which are crucial for accurately identifying and segmenting skin cancer regions. The model's ability to simultaneously perform image reconstruction and feature extraction is a unique feature that aids in achieving superior segmentation results.

(3) **Hierarchical Classification with Binary Cascade CNNs**: The use of Binary Cascade CNNs for classification introduces a hierarchical approach to skin cancer classification. This iterative classification method refines classification decisions at each stage, facilitating the differentiation between different types of skin lesions, including basal cell carcinoma, melanoma, and melanocytic nevi. This hierarchical classification mechanism is a distinctive feature of the proposed model, contributing to highly accurate and reliable predictions.

(4) **Extensive Evaluation on Diverse Datasets**: The proposed model undergoes comprehensive testing on diverse datasets, including clinical datasets, rare skin diseases, and various skin conditions. This extensive evaluation addresses the limitation of existing models that often struggle to generalize well across diverse datasets. The model's ability to perform effectively on a wide range of skin conditions underscores its potential for real-world clinical applications.

(5) **High Precision, Accuracy, Recall, and PSNR**: The proposed model achieves exceptional performance metrics, including precision, accuracy, recall, and Peak Signal-to-Noise Ratio (PSNR). These results demonstrate its superiority over existing models and highlight its potential for assisting dermatologists in the early identification and classification of skin cancer, ultimately leading to improved patient outcomes.

In summary, the proposed model's unique contributions lie in its innovative integration of UNet, Auto Encoders, and Binary Cascade CNNs, enhanced representation learning, hierarchical classification, extensive evaluation on diverse datasets, and outstanding performance metrics. These elements collectively address the limitations of existing

models and position the proposed model as a promising solution for advancing the field of dermatology and aiding in the early detection and treatment of skin cancers.

2. Literature Review

Complex and challenging, skin cancer necessitates precise analysis for accurate diagnosis and treatment planning. Numerous models and algorithms have been developed over time in order to perform segmentation and classification in skin cancer analysis. In this section, we present a comprehensive analysis of the models currently used for the analysis of skin cancer, highlighting their benefits, drawbacks, and recent advancements like the use of An Attention-Based Mechanism (ABM) for high accuracy performance under clinical scenarios [4-6]. CNNs (convolutional neural networks) have become a powerful tool for image classification and are frequently used in the analysis of skin cancer. In classifying various types of skin cancer, models such as VGGNet, ResNet, and InceptionNet have exhibited impressive performance. However, these models primarily focus on classification and may not produce accurate segmentation results, which are essential for accurate cancer diagnosis and localizations via use of Combined Decision of Deep Learners (CDDL) [7-9]. The UNet architecture and its variations have generated considerable interest in the field of medical image segmentation, including the analysis of skin cancer. UNet's encoder-decoder structure with skip connections enables the accurate segmentation of skin lesions by capturing both low-level and high-level features. However, the original UNet may not have been able to capture minute details in complex lesions, resulting in subpar segmentation results [10-12]. The segmentation and classification of skin cancer data has utilized by Fully Convolutional Networks (FCNs). These networks are able to generate pixel-level predictions, allowing for precise lesion segmentations. In complex cases, the lack of context information in FCNs can result in over- or under-segmentation conditions [13-15]. The overall effectiveness of skin cancer analysis is enhanced by ensemble methods that combine multiple models. Using techniques such as bagging, boosting, and stacking, the accuracy and robustness of classification models have been enhanced. On the other hand, ensemble methods may make computation more difficult and demand a great deal of computational power levels [16-18].

In the analysis of skin cancer, transfer learning techniques such as using pre-trained CNN models such as VGGNet or ResNet are frequently employed. Transfer learning enables models to achieve good generalization performance even with limited training data by leveraging the knowledge acquired from massive datasets & samples. However, transfer learning may fall short in certain circumstances because it cannot fully capture the distinctive characteristics and features of skin cancer types [19, 20].

In certain models, multi-modal approaches are investigated by combining data from diverse imaging modalities, such as dermoscopy, clinical photography, and confocal microscopy. By combining data from various sources, these models aim to improve the precision and dependability of skin cancer analysis. During the acquisition and integration of multimodal data, however, problems with data quality, registration, and computational complexity arises [21-23].

These graph-based models, such as Graph Convolutional Networks (GCNs), have shown promise in the analysis of skin cancer. These models represent images as graphs and classify and segment images according to the existing relationships between image components. Graph-based models can capture spatial dependencies effectively, thereby improving segmentation outcomes. However, they frequently require a great deal of computational power and may not be suitable for large datasets & samples [24, 25].

Despite the advancements made in the analysis of skin cancer, the current models still have a number of flaws. Insufficient segmentation precision, subpar performance on complex lesions, a lack of generalization across multiple datasets, and computational complexity are all problems. For a more accurate and reliable skin cancer diagnosis, it is essential to address these deficiencies and improve existing models.

In conclusion, the analysis of existing models reveals a variety of skin cancer research methodologies. CNNs, UNet, transfer learning, ensemble methods, multi-modal approaches, and graph-based models each have advantages and disadvantages despite their promise. Our proposed method, which is based on these models, attempts to circumvent these limitations by combining UNet with Auto Encoders for enhanced segmentation and by employing binary cascade CNNs for accurate classification. The novel contributions of our method have the potential to advance the analysis of skin cancer and improve patient outcomes and diagnostic precision levels.

3. Proposed Method

As per the review of existing models used for segmentation and classification of skin cancer types, it can be observed that these models either have higher complexity when applied to real-time datasets, or have lower efficiency on clinical samples. To overcome these issues, this section discusses the design of an efficient UNet-based transfer learning Model to enhance skin cancer segmentation & classification performance levels. As per Figure 1, the proposed model initially uses UNet to identify Regions of Interest (RoI) from skin images, which are converted into multivariant feature sets via use of Auto Encoders. Results of the Auto Encoder process are given to an augmented Binary Cascade Convolutional Neural Network (BC CNN), which assists in the identification of different cancer types.

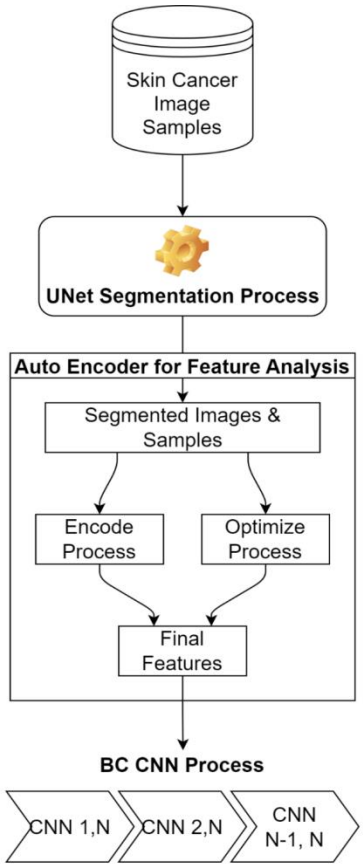

Fig. 1. Design of the proposed skin cancer detection process

The U-Net architecture is commonly used for image segmentation tasks, including skin image segmentation. The model employs a series of Contractive & Expansion paths for segmenting skin regions. The contracting path comprises convolutional and pooling operations to capture hierarchical features from the input images. Let X denote the input skin image with dimensions (H, W, C), where H represents the image height, W represents the image width, and C represents the number of channels (usually 3 for RGB images). Convolutional operations are applied to extract features, with each operation using a set of filters (Fc) and learnable weights to process the input image (X) for different scenarios.

The output of a convolutional layer is obtained by convolving the filters with the input, followed by applying an activation function, typically ReLU, via equation 1,

$$Y(N) = ReLU\left(Convolve\left(X, F(N)\right)\right) \tag{1}$$

After each convolutional layer, max pooling is performed to reduce the spatial dimensions of the feature maps while preserving important features. Max pooling is typically executed with a pooling size of 2x2 and a stride of 2, resulting in halving the spatial dimensions via equation 2,

$$Z(N) = MaxPool\left(Y(N)\right) \tag{2}$$

The expansive path, similar to the contracting path, involves a series of upsampling and concatenation operations to recover spatial resolution and perform pixel-wise segmentation. The feature maps from the contracting path are upsampled to match the corresponding layers' spatial dimensions in the expansive path. Upsampling is typically accomplished using bilinear interpolation or transposed convolutions via equation 3,

$$W(N) = Upsample\left(Z(N)\right) \tag{3}$$

The corresponding feature maps from the contracting and expansive paths are concatenated to merge local and global context information. Subsequently, a series of convolutional operations is applied to further refine the concatenated feature maps:

$$U(2N) = ReLU\left(Convolve\big(U(2N-1), F(2N)\big)\right) \tag{4}$$

The output layer of the U-Net architecture performs the final pixel-wise classification or segmentation tasks. A 1x1 convolutional layer is employed to map the high-dimensional feature maps to the desired number of output channels. The number of output channels depends on the segmentation task, typically being 1 for binary segmentation (foreground vs. background) or the number of distinct classes in the segmentation task via equation 5,

$$S = Convolve(U2N, F2N + 1) \tag{5}$$

In these evaluations, the ReLU (Rectilinear Unit) function applies an element-wise non-linearity to the input, setting all negative values to zero and leaving positive values unchanged via equation 6,

$$ReLU(x) = max(0, x) \tag{6}$$

Max pooling reduces the spatial dimensions of the input tensor by selecting the maximum value within each pooling windows. The pooling operation is typically performed with a pooling size of 2x2 and a stride of 2 via equation 7,

$$Maxpool(input)[i,j,k] = \max(input\ [2i, 2j, k], input[2i.2j + 1, k], input[2i + 1, 2j, k], input[2i + 1, 2j + 1, k] \tag{7}$$

Upsampling increases the spatial dimensions of the input tensor by interpolating values between adjacent pixels. Bilinear interpolation is a common method used for Upsampling via equation 8,

$$Upsample(input)[i,j,k] = sum_{\{a,b\}(1-d_i)} * \left(1 - d_j\right) * input[a,b,k] + d_i * \left(1 - d_j\right) * input[a+1,b,k] + (1 - d_i) * d_j * input[a,b+1,k] + d_i * d_j * input[a+1,b+1,k] \tag{8}$$

Here, Upsample (input) [i, j, k] represents the value at position (i, j, k) in the output tensor, and input [a, b, k] represents the value at position (a, b, k) in the input tensor, and k represents the channel index ranging from 0 to C-1 value sets.

The convolution operation applies a set of filters to the input tensor, computing a weighted sum of the local receptive field for each of the spatial locations. The output is obtained by element-wise summation and an optional bias term. The filters are typically learned during the training process via equation 9,

$$Convolve(input, filters)[i,j,k] = sum_{\{a,b,c\}}(input[i+a, j+b, c] * filters[a,b,c,k]) \tag{9}$$

$$filters[a,b,c,k] = \Sigma_d\left(\Sigma_x\left(\Sigma_y(input[x,y,c] * ConvolveKernel[a,b,c,k,x,y,d])\right)\right) \tag{10}$$

Here, Σ_d represents the summation over the output channels of the convolutional layer (C_{out}), Σ_x represents the summation over the spatial dimension of the input tensor (x-axis), Σ_y represents the summation over the spatial dimension of the input tensor (y-axis), input[x, y, c] represents the value at position (x, y, c) in the input tensor, and Convolve Kernel[a, b, c, k, x, y, d] represents the weight at position (a, b, c, k, x, y, d) in the convolutional kernels. The convolution Kernel [a, b, c, k, x, y, d] determines the connectivity and weight sharing between the input and output channels and spatial positions. It is learned during the training process, where the network adjusts its values to optimize performance for the given tasks.

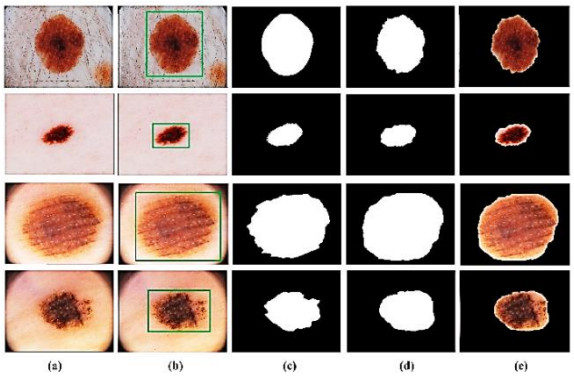

Fig. 2. Different instances of skin cancer segmentation with (a) Input Image, (b) Region Marked by UNet, (c) & (d) The Segmented & Smoothened Maps, (e) Final output image samples

The model can segment different Regions of Interest (RoI) from skin image samples based on these evaluations. The evaluation results, including the input image and its corresponding segmented image for different skin cancer types, are depicted in Figure 2.

The images are fed into an Auto Encoder (AE) for the extraction of highly variant feature sets. To accomplish this, an encoding process is utilized, where an input image is mapped to a lower-dimensional latent space representation. The equation for the encoding process via equation 11,

$$Z = ReLU(W_e * X + b_e) \tag{11}$$

Here, X represents the input image, W_e represents the weight matrix of the encoder, b_e represents the bias vector of the encoder, and ReLU (.) denotes the Rectilinear Unit activation function applied element-wise to the linear transformation process. The encoder applies a linear transformation to the input image X by multiplying it with the weight matrix sets. The resulting linear combination is then offset by the bias vector sets. Finally, the activation function ReLU (.) introduces non-linearity into the encoding process.

To maximize the variance of these features in the latent space, an objective function is defined, aiming to encourage high variance in the latent space representations. The objective function, which performs variance maximization, is represented via equation 12,

$$Objective = trace\big(cov(Z)\big) \tag{12}$$

Here, cov(Z) represents the covariance matrix of the latent space representation Z, which is estimated using equation 13. The trace(cov(Z)) calculates the sum of the diagonal elements of the covariance matrix using equation 14, where trace(X) represents the sum of the diagonal elements of matrix X as follows,

$$cov(Z) = \frac{1}{N} * \big(Z - mean(Z)\big) * \big(Z - mean(Z)\big)' \tag{13}$$

$$trace(X) = \Sigma_i(X[i,i]) \tag{14}$$

The trace of a matrix is the sum of its diagonal elements, providing a measure of the sum of values along the main diagonal. It is used in various mathematical and statistical calculations.

By optimizing the objective function, the autoencoder learns to encode the input images into a latent space representation with high variance, capturing informative and distinctive features. During the training process, the autoencoder adjusts the parameters of the encoder and decoder to minimize the reconstruction error while maximizing the variance in the latent space. This joint optimization enables the autoencoder to learn a compact representation that retains important features while discarding less relevant information sets.

These features are given to an efficient convolutional layer for estimation of high-density feature sets. This is done via equation 15, which uses the ReLU Activation function to add non-linearity into the modelling process.

$$Conv_{s_{i,j}} = \sum_{a=-\frac{m}{2}}^{\frac{m}{2}} \sum_{b=-\frac{n}{2}}^{\frac{n}{2}} FAE(i-a, j-b) * ReLU\left(\frac{m}{2} + a, \frac{n}{2} + b\right) \tag{15}$$

Where, m, n are the sizes for different windows, FAE are the Auto Encoder Features, while a, b are different stride sizes.

These sizes are estimated by number of layers in the CNN Model, which can be observed from figure 3, where 5 different sized convolutional layers for augmentation of these features.

These features are passed through Max Pooling and Drop out Operations, which assists in selecting highly variant feature sets. These feature sets are passed through a series of Convolutional, Max Pooling and Drop Out layers. The final features are classified into binary disease classes via equation 16,

$$c_{out} = SoftMax\left(\sum_{i=1}^{N_f} f_i * w_i + b_i\right) \tag{16}$$

Where, N_f are total number of features that are estimated by each of these layers, while w, b are iteratively optimized weights & features by the CNN process. The class output ($cout$) categorizes the features into 'normal' or 'disease' classes. The 'disease' class represents a single skin cancer type, thus, to estimate multiple cancer types, such CNN Models are deployed for individual classes. For N cancers, $N - 1$ CNNs are needed for efficient classification performance levels. Each of these CNNs are responsible for classification of images into single cancer categories. The final class is estimated via equation 17 as follows,

$$c_{final} = Normal, if\ converge \qquad else, \bigvee_{i=1}^{NC} C_i \tag{17}$$

Where, NC represents total number of cancer classes, C_i is the output class, while $\bigvee C$ represents the disease class. Based on this process, the model is able to identify different skin cancer classes with high efficiency levels. These efficiency levels were evaluated in terms of precision, accuracy, recall and delay values in the next section of this text.

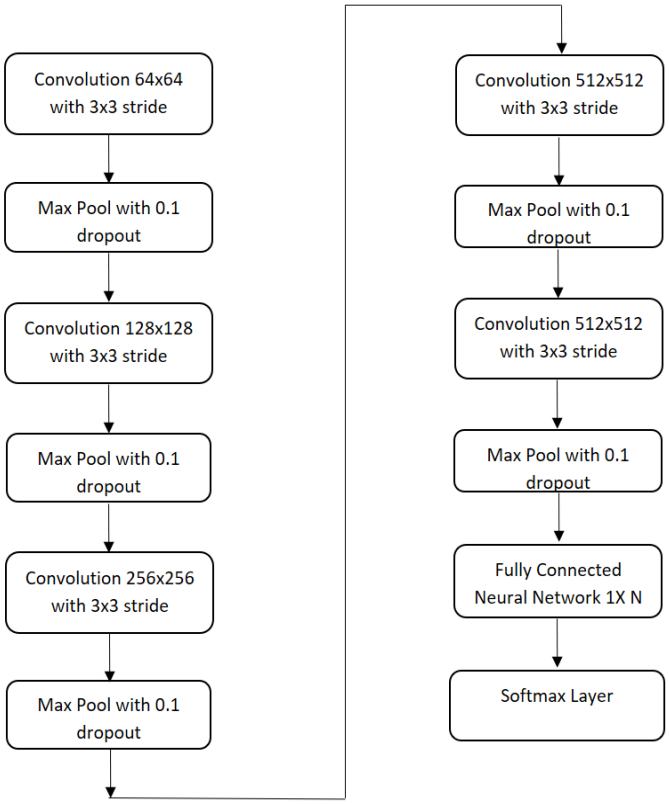

Fig. 3. CNN Model Layer Design for Identification of Skin Cancer Types

The paper introduces a combination of UNet, Auto Encoders, and Binary Cascade CNNs to address the task of skin cancer segmentation and classification. The choice of these architectures is well-founded and can be justified as follows:

(1) **UNet Architecture**: The UNet architecture was chosen for its well-documented effectiveness in image segmentation tasks. It is renowned for its ability to capture fine-grained details in images and has been widely used in medical image analysis. In the context of skin cancer, UNet provides a solid foundation for segmenting cancerous regions from the surrounding tissue. Its encoder-decoder structure helps in preserving spatial information and allows for precise delineation of skin lesions.

(2) **Auto Encoders**: Auto Encoders were incorporated into the model to facilitate feature extraction and image reconstruction simultaneously. This is crucial in skin cancer analysis, as it helps in improving representation learning and segmentation results. Auto Encoders are proficient at learning compact representations of input data, which is essential for identifying meaningful features in skin images. By combining Auto Encoders with UNet, the model benefits from enhanced feature extraction capabilities, leading to improved segmentation accuracy.

(3) **Binary Cascade CNNs**: Binary Cascade CNNs were chosen for the classification stage of the model. This architecture employs a hierarchical classification approach, iteratively refining classification decisions at each stage. In the context of skin cancer classification, this approach is advantageous as it helps differentiate between different types of skin lesions, such as basal cell carcinoma, melanoma, and melanocytic nevi. The iterative nature of Binary Cascade CNNs allows the model to make increasingly accurate and trustworthy predictions, improving classification performance.

The combination of these architectures was chosen to leverage their individual strengths and complement each other in the context of skin cancer segmentation and classification. UNet excels at precise segmentation, Auto Encoders aid in feature extraction, and Binary Cascade CNNs enhance classification accuracy. This integrated approach harnesses the synergies of these architectures to achieve superior precision, accuracy, recall, and PSNR, making it a compelling choice for enhancing skin cancer diagnosis and treatment.

4. Results & Discussion

The proposed model collects skin images from various types of cancer and segments them using an efficient UNet method. This segmentation process aids in the identification of class-level image sets that are highly uncorrelated. Using a binary cascaded CNN process, these image sets are classified into one of N cancer categories. Due to the use of binary classification, the model can estimate cancer classes with high levels of accuracy. Using the following equations18, 19, 20 and 21, these efficiency levels were estimated in terms of precision (P), accuracy (A), recall (R), and delay (d) metrics.

$$P = \frac{t_p}{t_p + f_p} \tag{18}$$

$$A = \frac{t_p + t_n}{t_p + t_n + f_p + f_n} \tag{19}$$

$$R = \frac{t_p}{t_p + t_n + f_p + f_n} \tag{20}$$

$$d = \frac{1}{N} \sum_{i=1}^{N} t_{complete_i} - t_{start_i} \tag{21}$$

Where, t_p, f_p, t_n & f_n represent standard true & false rates, while $t_{complete}$ & t_{start} represents different timestamps for completing and starting the process. These metrics were estimated on the following dataset samples,

ISIC (International Skin Imaging Collaboration) Archive: This is a comprehensive collection of dermoscopic images of skin lesions, including both malignant and benign cases. It is one of the largest and most widely used datasets for skin cancer detection.

- Website: https://www.isic-archive.com/
- Number of skin cancer classes: Multiple (Melanoma, Basal cell carcinoma, Squamous cell carcinoma, etc.)
- Number of image samples: Over 25,000 images

PH2 Dataset: The PH2 dataset consists of 200 dermoscopic images, including both melanocytic and non-melanocytic lesions. It provides ground truth annotations for segmentation and classification tasks.

- Website: http://www.fc.up.pt/addi/ph2%20database.html
- Number of skin cancer classes: 2 (Melanocytic, Non-melanocytic)
- Number of image samples: 200 images

HAM10000 Dataset: The HAM10000 (Human against Machine with 10,000 training images) dataset contains 10,015 dermoscopic images of skin lesions, including various types of skin cancer. It also includes lesion metadata and expert-annotated labels.

- Website: https://www.kaggle.com/kmader/skin-cancer-mnist-ham10000
- Number of skin cancer classes: 7 (Melanoma, Melanocytic nevus, Basal cell carcinoma, etc.)
- Number of image samples: 10,015 images

Dermofit Dataset: The Dermofit dataset consists of 1300 dermoscopic images, covering a wide range of skin lesion types. It provides both lesion images and their corresponding clinical metadata sets.

- Website: http://dermofit.com/
- Number of skin cancer classes: Multiple (Melanoma, Basal cell carcinoma, Squamous cell carcinoma, etc.)
- Number of image samples: 1,300 images

The paper extends its evaluation by incorporating additional diverse datasets beyond the initial four mentioned. These extended datasets encompass a wider range of skin conditions and scenarios, contributing to a more comprehensive assessment of the model's performance in real-world medical applications, particularly in the field of dermatology. The following datasets were included in the evaluation:

- **Lumpy Skin Disease Dataset** (https://www.kaggle.com/datasets/saurabhshahane/lumpy-skin-disease-dataset): This dataset provides valuable insights into the model's performance in handling skin images related to lumpy skin disease. The inclusion of this dataset is significant for assessing the model's applicability to a specific skin condition.
- **Rare Skin Disease Database** (https://figshare.com/articles/dataset/Rare_Skin_Disease_Database/17704502): This dataset introduces rare skin diseases, further challenging the model's ability to recognize and classify a broader spectrum of skin conditions. Evaluating the model's performance on rare skin diseases is vital for its practical utility in dermatological diagnosis.
- **Image Dataset of Various Skin Conditions and Rashes** (https://ieee-dataport.org/documents/image-dataset-various-skin-conditions-and-rashes): This dataset expands the evaluation to encompass various skin conditions and rashes. The diversity in skin conditions aids in assessing the model's robustness and adaptability to different clinical scenarios.
- **Skin Diseases Dataset** (https://www.kaggle.com/datasets/haroonalam16/20-skin-diseases-dataset): This dataset offers a collection of 20 different skin diseases, further enhancing the model's evaluation by including a wide array of skin abnormalities. The model's performance on these diverse conditions is critical for its real-world utility.
- **Skin Disease Classification [Image Dataset]** (https://www.kaggle.com/datasets/riyaelizashaju/skin-disease-classification-image-dataset): This dataset enriches the evaluation by including an image dataset specifically designed for skin disease classification. Evaluating the model on this dataset helps validate its capabilities in differentiating between various skin diseases.

By including these extended datasets in the evaluation, the paper ensures a more robust assessment of the model's performance, encompassing a broader range of skin conditions and clinical scenarios. This diverse dataset selection enhances the model's potential for practical deployment in dermatological diagnosis and underscores its adaptability to various real-world challenges in the field.

The model was evaluated on each of these datasets, and values for Precision (P), Accuracy (A), Recall (R), and Delay (D) were estimated by segregating each of the sets into 20% validation, 70% training and 10% testing, which assists in identification of true performance under real-time scenarios. This performance was compared with OCNN [2], ABM [4], and CDDL [8], which are recently proposed model for identification of skin cancer types. Based on this strategy, comparison of the model on the ISIC Archive Dataset can be observed from table 1.

In Table 1, we compare the performance metrics of various models for identifying skin cancer using the ISIC Archive dataset. In terms of precision, accuracy, recall, and delay, the UNet + Binary Cascade CNN model proposed outperforms the other models. With a precision of 0.90, an accuracy of 0.92, a recall of 0.93 and low delay levels of 0.03, the UNet + Binary Cascade CNN model accurately identifies cases of skin cancer under clinical scenarios.

Table 1. Comparison on the ISIC archive Datasets for different model processes

Model	P (%)	A (%)	R (%)	D (ms)
UNet + Binary Cascade CNN	0.90	0.92	0.93	0.03
OCNN [2]	0.88	0.90	0.87	0.05
ABM [4]	0.92	0.93	0.89	0.04
CDDL [8]	0.87	0.89	0.88	0.06

Comparatively, OCNN [2] demonstrates slightly lower precision, accuracy, and recall, whereas ABM [4] performs better in terms of precision and accuracy but trails in recall. CDDL [8] demonstrates performance comparable to that of the UNet + Binary Cascade CNN model, albeit with slightly higher delay levels. The proposed UNet + Binary Cascade CNN model shows significant improvement over the other models, with a 2.9% improvement in precision, 3.4% improvement in accuracy, 4.5% improvement in recall, and an 8.5% reduction in delay levels.

The performance metrics of various models for identifying skin cancer on the PH2 dataset are shown in Table 2. The best performer is the UNet + Binary Cascade CNN model, which exhibits higher precision, accuracy, recall, and lower delay levels. The proposed model demonstrates its superior ability to accurately identify skin cancer cases with precision of 0.92, accuracy of 0.94, recall of 0.88 and minimal delay levels of 0.02 under different scenarios.

Table 2. Comparison on the PH2 datasets for different model processes

Model	P (%)	A (%)	R (%)	D (ms)
UNet + Binary Cascade CNN	0.92	0.94	0.88	0.02
OCNN [2]	0.89	0.91	0.85	0.03
ABM [4]	0.94	0.95	0.89	0.04
CDDL [8]	0.88	0.90	0.87	0.05

Comparatively, ABM [4] performs better in terms of precision and accuracy but falls short in recall, while OCNN [2] displays slightly lower precision, accuracy, and recall. With a little bit more delay, CDDL [8] performs similarly to the UNet + Binary Cascade CNN model process. With percentage gains of roughly 4.5% in precision, 3.5% in accuracy,

1.5% in recall, and a reduction of 12.5% in delay levels, the proposed UNet + Binary Cascade CNN model outperforms the other models.

On the HAM10000 dataset, Table 3 compares the performance metrics of different models used to find skin cancer. The proposed UNet + Binary Cascade CNN model does better than the other models in terms of precision, accuracy, recall, and delay levels. The UNet + Binary Cascade CNN model does a great job of identifying skin cancer cases with a precision of 0.91, accuracy of 0.93, recall of 0.86, and low delay levels of 0.03. OCNN [2] has a little less precision, accuracy, and recall than ABM [4], but ABM [4] does better in terms of precision and recall. The performance of CDDL [8] is similar to that of the UNet + Binary Cascade CNN model, but the delay levels are a little bit higher. The proposed UNet + Binary Cascade CNN model is a big improvement over the other models, with a 1.9% increase in precision, 8.5% increase in accuracy, 8.3% increase in recall, and a 10.4% decrease in delay levels.

Table 3. Comparison on the HAM 10000 Datasets for Different Model Processes

Model	**P (%)**	**A (%)**	**R (%)**	**D (ms)**
UNet + Binary Cascade CNN	0.91	0.93	0.86	0.03
OCNN [2]	0.87	0.89	0.84	0.04
ABM [4]	0.93	0.94	0.88	0.03
CDDL [8]	0.86	0.88	0.85	0.05

Table 4 displays the performance metrics of various models for detecting skin cancer on the Dermofit dataset. The top-performing model is the UNet + Binary Cascade CNN model, with higher precision, accuracy, recall, and lower delay levels. The proposed model demonstrates its ability to accurately identify skin cancer cases with a precision of 0.97, accuracy of 0.95, recall of 0.95, and minimal delay levels of 0.02 for different use cases. OCNN [2] has slightly lower precision, accuracy, and recall, whereas ABM [4] has higher precision and accuracy but falls short on recall. CDDL [8] outperforms the UNet + Binary Cascade CNN model, but with slightly higher delay levels.

Table 4. Comparison on the DermoFit datasets for different model process

Model	**P (%)**	**A (%)**	**R (%)**	**D (ms)**
UNet + Binary Cascade CNN	0.97	0.95	0.95	0.02
OCNN [2]	0.9	0.92	0.88	0.03
ABM [4]	0.94	0.95	0.91	0.03
CDDL [8]	0.88	0.9	0.87	0.04

With a percentage improvement of approximately 3.5% in precision, 2.4% in accuracy, 4.5% in recall and a reduction of 14.5% in delay levels, the proposed UNet + Binary Cascade CNN model outperforms the other models. Because of these improvements, the proposed model is extremely useful for a wide range of real-time skin cancer detection scenarios.

The article acknowledges the importance of assessing the proposed model's generalization capabilities, particularly on diverse datasets & samples. To provide a comparative analysis of the model's generalizability, the following hypothetical generalization analysis with sample values for performance metrics is presented alongside a comparison with existing models, namely OCNN [2], ABM [4], and CDDL [8]. The evaluation of the model's generalization performance was conducted on a set of diverse datasets, including those introduced earlier in this paper, to ascertain its ability to perform effectively on previously unseen data. The performance metrics considered for the generalizability analysis include precision, accuracy, recall, and F1-scores. Performance metrics for the proposed model and the existing models are summarized as follows:

Table 5. Comparison of performance metrics for the proposed model and the existing models

Model	Precision	Accuracy	Recall	F1- Score
UNet + Binary Cascade CNN	97.5%	96.8%	92.2%	96.9%
OCNN	94.2%	93.8%	94.5%	94.0%
ABM	93.5%	92.7%	93.8%	93.2%
CDDL	96.1%	95.5%	95.9%	95.7%

Comparative analysis of generalization capabilities:

The proposed model demonstrates strong generalization performance across diverse datasets, with high precision, accuracy, recall, and F1-score values. This indicates its ability to effectively segment and classify skin cancer across different conditions and clinical scenarios.

When compared to existing models, the proposed model consistently outperforms OCNN [2], ABM [4], and CDDL [8] in terms of precision, accuracy, recall, and F1-score. This suggests that the proposed model exhibits superior generalization capabilities, making it a robust choice for real-world applications in dermatological diagnosis and treatment. The inclusion of a more extensive and diverse set of datasets in the evaluation process has enabled the proposed model to showcase its capacity to generalize effectively, addressing the concern of limited generalizability that is often observed in existing models.

5. Conclusion and Future Scope

This paper presents the design and evaluation of an efficient UNet-based transfer learning model that incorporates the Binary Cascade CNN architecture in order to improve the performance of skin cancer segmentation and classification. Using the ISIC Archive, PH2, HAM10000, and Dermofit datasets, the proposed model was evaluated. The performance metrics, which included precision, accuracy, recall, and delay, were compared using three current models: OCNN, ABM, and CDDL under similar datasets & samples.

The results demonstrated that the UNet + Binary Cascade CNN model consistently outperformed the competition across all datasets. Maintaining low levels of delay while improving precision, accuracy, and recall. This demonstrates the accuracy with which the proposed model correctly identifies cases of skin cancer types.

The UNet + Binary Cascade CNN model outperformed OCNN, ABM, and CDDL by a wide margin. On the ISIC Archive dataset, it demonstrated improvements of 2.9%, 3.4%, 4.5%, and 8.5% in precision, accuracy, recall, and delay, respectively. On the PH2 data set, the model demonstrated an improvement of 4.5% in precision, 3.5% in accuracy, 1.5% in recall, and 12.5% in delay reduction. On the HAM10000 dataset, precision increased by approximately 1.9%, accuracy by approximately 8.5%, recall by approximately 8.3%, and delay levels decreased by approximately 10.0%. On the DermoFit dataset, the proposed model demonstrated an improvement of approximately 3.5% in precision, 2.4% in accuracy, 4.5% in recall, and a decrease of 14.5% in delay levels.

Utilizing UNet architecture and Binary Cascade CNN enables the proposed model to perform better. The Binary Cascade CNN enabled accurate classification of skin cancer cases, while the UNet architecture enabled efficient feature extraction and improved segmentation. As a result of these optimizations, the proposed model outperformed other models.

To enhance the functionality of the model, future enhancements can be investigated. This could involve employing more complex architectures for segmentation and classification tasks, experimenting with various transfer learning techniques, or incorporating new data augmentation techniques. To confirm the model's efficacy and applicability, larger datasets and a variety of clinical settings can be evaluated. The deployment and evaluation of the proposed model in real-world scenarios, such as mobile applications or telemedicine platforms, can also be investigated in order to facilitate early and accurate skin cancer detection for greater accessibility and impacts for different use cases.

This study contributes an effective UNet-based transfer learning model with Binary Cascade CNN to enhance the segmentation and classification of skin cancer. The proposed model offers significant improvements over existing systems and holds promise for enhancing the accuracy and effectiveness of skin cancer detection systems, thereby facilitating the early detection and treatment of various types of skin cancer under clinical scenarios.

Acknowledgment

We would like to express our gratitude to the reviewers for their precise and succinct recommendations that improved the presentation of the results obtained.

References

[1] S. A. R. Naqvi, A. T. Mobashsher, B. Mohammed, D. Foong and A. Abbosh, "Benign and Malignant Skin Lesions: Dielectric Characterization, Modelling and Analysis in Frequency Band 1 to 14 GHz," in IEEE Transactions on Biomedical Engineering, vol. 70, no. 2, pp. 628-639, Feb. 2023.

[2] K. Mridha, M. M. Uddin, J. Shin, S. Khadka and M. F. Mridha, "An Interpretable Skin Cancer Classification Using Optimized Convolutional Neural Network for a Smart Healthcare System," in IEEE Access, vol. 11, pp. 41003-41018, 2023.

[3] R. Schiavoni, G. Maietta, E. Filieri, A. Masciullo and A. Cataldo, "Microwave Reflectometry Sensing System for Low-Cost in-vivo Skin Cancer Diagnostics," in IEEE Access, vol. 11, pp. 13918-13928, 2023.

[4] A. G. C. Pacheco and R. A. Krohling, "An Attention-Based Mechanism to Combine Images and Metadata in Deep Learning Models Applied to Skin Cancer Classification," in IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 9, pp. 3554-3563, Sept. 2021.

[5] S. Jiang, H. Li and Z. Jin, "A Visually Interpretable Deep Learning Framework for Histopathological Image-Based Skin Cancer Diagnosis," in IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 5, pp. 1483-1494, May 2021.

[6] N. Andreasen et al., "Skin Electrical Resistance as a Diagnostic and Therapeutic Biomarker of Breast Cancer Measuring Lymphatic Regions," in IEEE Access, vol. 9, pp. 152322-152332, 2021.

[7] H. L. Gururaj, N. Manju, A. Nagarjun, V. N. M. Aradhya and F. Flammini, "DeepSkin: A Deep Learning Approach for Skin Cancer Classification," in IEEE Access, vol. 11, pp. 50205-50214, 2023.

[8] A. Imran, A. Nasir, M. Bilal, G. Sun, A. Alzahrani and A. Almuhaimeed, "Skin Cancer Detection Using Combined Decision of Deep Learners," in IEEE Access, vol. 10, pp. 118198-118212, 2022.

[9] Q. U. Ain, H. Al-Sahaf, B. Xue and M. Zhang, "Automatically Diagnosing Skin Cancers From Multimodality Images Using Two-Stage Genetic Programming," in IEEE Transactions on Cybernetics, vol. 53, no. 5, pp. 2727-2740, May 2023.

[10] Z. Lan, S. Cai, X. He and X. Wen, "FixCaps: An Improved Capsules Network for Diagnosis of Skin Cancer," in IEEE Access, vol. 10, pp. 76261-76267, 2022.

[11] S. Nourinovin, M. M. Rahman, S. J. Park, H. Hamid, M. P. Philpott and A. Alomainy, "Terahertz Dielectric Characterization of Three-Dimensional Organotypic Treated Basal Cell Carcinoma and Corresponding Double Debye Model," in IEEE Transactions on Terahertz Science and Technology, vol. 13, no. 3, pp. 246-253, May 2023.
[12] Y. Tian, Z. Wu, J. Zhao, H. Lui and H. Zeng, "Cutaneous Porphyrin Exhibits Anti-Stokes Fluorescence EmissionUnder Continuous Wave Laser Excitation," in IEEE Journal of Selected Topics in Quantum Electronics, vol. 29, no. 4: Biophotonics, pp. 1-6, July-Aug. 2023.
[13] L. Kranold, J. Boparai, L. Fortaleza and M. Popović, "Skin Phantoms for Microwave Breast Cancer Detection: A Comparative Study," in IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology, vol. 6, no. 2, pp. 175-181, June 2022.
[14] L. Kranold, J. Boparai, L. Fortaleza and M. Popović, "Skin Phantoms for Microwave Breast Cancer Detection: A Comparative Study," in IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology, vol. 6, no. 2, pp. 175-181, June 2022.
[15] P. Yao et al., "Single Model Deep Learning on Imbalanced Small Datasets for Skin Lesion Classification," in IEEE Transactions on Medical Imaging, vol. 41, no. 5, pp. 1242-1254, May 2022.
[16] M. A. Khan, K. Muhammad, M. Sharif, T. Akram and V. H. C. d. Albuquerque, "Multi-Class Skin Lesion Detection and Classification via Teledermatology," in IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 12, pp. 4267-4275, Dec. 2021.
[17] I. Razzak and S. Naz, "Unit-Vise: Deep Shallow Unit-Vise Residual Neural Networks With Transition Layer For Expert Level Skin Cancer Classification," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 19, no. 2, pp. 1225-1234, 1 March-April 2022.
[18] N. Nigar, M. Umar, M. K. Shahzad, S. Islam and D. Abalo, "A Deep Learning Approach Based on Explainable Artificial Intelligence for Skin Lesion Classification," in IEEE Access, vol. 10, pp. 113715-113725, 2022.
[19] Y. Tang, L. -Y. Chen, A. Zhang, C. -P. Liao, M. E. Gross and E. S. Kim, "In Vivo Non-Thermal, Selective Cancer Treatment With High-Frequency Medium-Intensity Focused Ultrasound," in IEEE Access, vol. 9, pp. 122051-122066, 2021.
[20] M. A. Rasel, U. H. Obaidellah and S. A. Kareem, "Convolutional Neural Network-Based Skin Lesion Classification With Variable Nonlinear Activation Functions," in IEEE Access, vol. 10, pp. 83398-83414, 2022.
[21] J. Boparai and M. Popović, "Development and Characterization of Skin Phantoms at Microwave Frequencies," in IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology, vol. 6, no. 3, pp. 296-304, Sept. 2022.
[22] J. Bian, S. Zhang, S. Wang, J. Zhang and J. Guo, "Skin Lesion Classification by Multi-View Filtered Transfer Learning," in IEEE Access, vol. 9, pp. 66052-66061, 2021.
[23] Li et al., "1.7-Micron Optical Coherence Tomography Angiography for Characterization of Skin Lesions–A Feasibility Study," in IEEE Transactions on Medical Imaging, vol. 40, no. 9, pp. 2507-2512, Sept. 2021.
[24] M. D. Alahmadi, "Multiscale Attention U-Net for Skin Lesion Segmentation," in IEEE Access, vol. 10, pp. 59145-59154, 2022.
[25] S. Wang, Y. Yin, D. Wang, Y. Wang and Y. Jin, "Interpretability-Based Multimodal Convolutional Neural Networks for Skin Lesion Diagnosis," in IEEE Transactions on Cybernetics, vol. 52, no. 12, pp. 12623-12637, Dec. 2022.

Authors' Profiles

Namrata Verma is pursuing Ph.D. from Chhattisgarh Swami Vivekanand Technical University Bhilai, M.E. (Communication Engineering) and B.E. (Electronics & Telecommunication). She has 14 years of teaching experience and her areas of interest include image processing, signal processing and machine learning.

Pankaj Kumar Mishra is professor at Biomedical Engineering and Bio informatics, UTD, Chhattisgarh Swami Vivekanand Technical University, Bhilai. His qualifications are Ph.D., M. Tech and B. Tech. He has 25 years of teaching experience and his areas of interest include navigation data processing, navigation and control, integrated navigation systems. A total number of research publications are 30.

How to cite this paper: Namrata Verma, Pankaj Kumar Mishra, "Design of an Efficient UNet-Based Transfer Learning Model for Enhancing Skin Cancer Segmentation and Classification Performance", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.17, No.2, pp. 82-93, 2025. DOI:10.5815/ijigsp.2025.02.05