ТЕОРЕТИЧНІ ТА ПРИКЛАДНІ ПРОБЛЕМИ ІНТЕЛЕКТУАЛЬНИХ СИСТЕМ ПІДТРИМАННЯ ПРИЙНЯТТЯ РІШЕНЬ

UDC 004.855.5
DOI: 10.20535/SRIT.2308-8893.2025.2.03

NAVIGATING CHALLENGES IN DEEP LEARNING FOR SKIN CANCER DETECTION

V. NIKITIN, V. DANILOV

Abstract. Skin cancer is one of the most prevalent malignancies worldwide. A critical factor in reducing mortality rates is the early detection. It underscores the need for accessible Computer-Aided Diagnostic (CAD) systems. Recent advancements in Deep Learning (DL) have shown great promise in addressing this challenge. Despite this progress in the field of machine learning, researchers encounter numerous obstacles when it comes to skin cancer classification. This article examines the current state of DL-based skin cancer diagnostics. Critical aspects of system development, including data preprocessing, model training, and performance evaluation, are addressed. Moreover, the article highlights opportunities for innovation that could significantly advance the field. By providing a comprehensive overview, this article aims to guide researchers and practitioners in optimizing DL models, addressing existing limitations, and exploring emerging trends to enhance diagnostic accuracy and accessibility.

Keywords: skin cancer, deep learning, classification, transformers, CNN, GAN, data preprocessing, data augmentation.

INTRODUCTION

Skin cancer affects millions of individuals every year. Malignancy originates in the skin's primary layers — the epidermis, dermis, and hypodermis. Due to its exposure to ultraviolet (UV) radiation, the epidermis is the most common site for skin cancers. Keratinocytes and melanocytes in these layers can undergo mutations due to prolonged exposure to UV radiation, which can lead to malignancy. The three major types of skin cancer include Basal Cell Carcinoma (BCC), Squamous Cell Carcinoma (SCC), and Malignant Melanoma [1].

BCC is the most common skin cancer, typically presenting as a slow-growing, pearly lesion. While it rarely metastasizes, untreated BCC can cause significant local tissue damage. SCC, the other type, is more aggressive and capable of metastasis. It often appears as scaly, crusty lesions, especially on sun-exposed skin, SCC requires timely intervention to prevent systemic spread. Malignant melanoma, though rare, is the most aggressive form of skin cancer and is prone to rapid metastasis [1].

Regardless of the type of skin cancer, early detection is crucial for successful treatment. Still, there are many barriers preventing timely diagnosis. According to

the World Health Organization, there is a global shortage of dermatologists, particularly in low- and middle-income countries [2]. Studies show evidence that lower-income populations are more likely to have advanced stages of skin cancer [3]. Hence, obtaining professional help to determine whether a lesion is malignant can be challenging. Even in developed regions, busy lifestyle and/or financial constraints can prevent individuals from seeking medical attention.

Computer Aided Diagnostics (CAD) system is a solution for assisted or automated diagnostics. They can be especially helpful in reducing human errors and expanding access to advanced medical care. Deep Learning (DL) with its advancements in recent years is a promising solution for skin cancer CAD. Machine Learning (ML) models can help to detect malignant lesions early via specialized or common hardware, potentially covering gap in dermatological services. Such systems can be particularly useful as prescreening tools, indicating cases where professional help is necessary.

DL has revolutionized various areas over the recent years, including medical imaging. During the past decade, researchers have developed numerous DL models for skin cancer classification with varying levels of success. This diversity of models and methods creates a sophisticated landscape to navigate. The article aims to underline key challenges and obstacles that researchers face developing skin cancer DL model as of 2024, providing examples of solutions proposed in existing researcher. Finally, it highlights areas for improvement that could take current DL-based CAD systems to a new level.

To simplify navigation, the article is divided into sections, each focusing on a critical component in the development of DL-based CAD systems for skin cancer detection.

1. *Researchers and Communities*: An overview of labs, teams, and organizations that have made significant advancements in DL-based skin cancer tasks in recent years.

2. *Datasets and Data Challenges*: A list of widely used datasets with their descriptions. This section also includes an overview of data based obstacles.

3. *Development Pipeline*: Addressing data normalization, augmentation, and techniques to fix data challenges highlighted in the previous section.

4. *Model Training and Optimization*: This section explores the complexities of selecting model architectures, leveraging transfer learning, and optimizing hyperparameters.

5. *Evaluation and Validation*: Various metrics suitable for the task are explored in detail.

6. *Opportunities for Innovation*: Directions and areas that could significantly enhance the quality of DL-based CAD systems.

7. *Guidance for Emerging Researchers*: Practical advice for newcomers to the field, including best practices in research design, navigating the literature, and identifying emerging trends.

By exploring these themes, the article provides a concise resource highlighting the current challenges, solutions, and future directions in the application of DL to skin cancer detection. Ultimately, the aim is to contribute to the development of more accurate, accessible diagnostic tools that can improve patient outcomes and reduce healthcare disparities.

RESEARCHERS AND COMMUNITIES

Stanford University Team. In 2017, the study "Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks" was published in Nature by a team from Stanford University. They researched a dataset of approximately 130,000 clinical images representing over 2,000 different skin diseases. They took Google Inception v3 CNN architecture, pre-trained on the ImageNet dataset, and fine-tuned it for skin lesion classification. Performance was validated against a group of 21 dermatologists, ensuring clinical relevance. It turned out the model achieved performance on par with board-certified dermatologists [4].

International Skin Imaging Collaboration. The International Skin Imaging Collaboration (ISIC) is the largest consortium that made a significant contribution to the skin cancer classification problem. The collaboration created the largest publicly available skin cancer dataset and conducted multiple challenges to encourage the development of advanced algorithms for skin lesion analysis. In mid-2024, ISIC released its latest dataset [5]. ISIC promotes transparency and reproducibility in dermatological research while safeguarding patient privacy. Its extensive use in numerous studies has made it a cornerstone for developing and validating DL models for skin cancer diagnostics, significantly accelerating progress in the field.

Memorial Sloan Kettering Cancer Center. Memorial Sloan Kettering Cancer Center (MSKCC) is one of the world's leading institutions dedicated to cancer treatment, research, and education. In dermatology, MSKCC implemented AI to support the detection of melanoma. The institution is actively working on implementing CAD systems to improve the effectiveness of diagnostics. They consider AI a complementary tool for medical experts [6].

Google Health and DeepMind. Google's research teams actively worked to apply DL to medical images. For instance, Yuan Liu and Peggy Bui published a paper titled "A Deep Learning System for Differential Diagnosis of Skin Diseases" in 2019 [7]. They developed a system that was able to accurately differentiate 26 common skin conditions.

Individual Researchers. There are numerous ML enthusiasts and experts working on the problem all over the world. For example, Yilmaz et al. achieved an accuracy of 82% with on the ISIC 2017 dataset with the NASNetMobile model [8]. Baygin et al. integrated textural and deep features to develop a pyramidal hybrid model receiving a classification accuracy of 91.54% with 10-fold cross-validation [9]. Agarwal and Singh were able to get 86.65% accuracy on ISIC Archive utilising convolutional neural networks with transfer learning [10].

Many more studies report remarkable results with the models and the data used. Each year, hundreds of new articles on the topic are being published, some of them making breakthroughs through usage of new technology or metric. The article mentions key pioneering research as examples of successful problem-solving for medical imaging. The main idea of this article is to guide through the current landscape of skin cancer images with DL leading to even more advances in the field.

DATASETS AND DATA CHALLENGES

Data is the base of any ML algorithm. The success of any DL project heavily depends on access to robust and comprehensive datasets as well as the correct usage

of those. By processing the data, the model extracts the most important features from it so that it can make correct predictions on unseen data of a similar nature. D. Wen et al. (2022) made an exhaustive list of datasets that were publicly available at the time [11]. The section below mentions key datasets from that list, adding the data that was released since 2022.

Datasets. *ISIC Archive.* The ISIC Archive [5] is the largest data repository with cancerous skin images so far. As of today, the number of open-for-usage samples is nearing half a million. More than 400,000 of those are images released as part of the latest ISIC Challenge — ISIC 2024 [12]. The archive includes a diverse range of skin lesions: melanomas, nevi, BCCs, SCC, etc. It includes both annotated and unannotated data sourced from various international centers. Some of the images contain precise borders of the lesion, allowing not only classification but segmentation as well. It might be the most impactful contribution to DL-based skin cancer CAD systems in terms of data.

HAM10000. The dataset [13], also known as "Human Against Machine with 10,000 training images" (HAM10000), published by Philipp Tschandl et al. (2018), is the most popular dataset for DL skin cancer classification. It contains a total of 10,015 dermoscopic images, categorized into seven common types of skin lesions: actinic keratoses, basal cell carcinoma, benign keratosis-like lesions, dermatofibroma, melanoma, melanocytic nevi, and vascular lesions. Each image is labeled with a confirmed diagnosis, verified through histopathology, follow-up examinations, or expert consensus. It also includes metadata detailing patient demographics such as age, gender, and lesion location, which adds context for model training.

Due to its high-quality data, HAM10000 was used in numerous articles, normally combined with transfer learning to compensate for the lack of size. For instance, A.T. Priyeshkumar et al. (2024) developed an ensemble DL model for skin lesion classification using HAM10000, achieving high accuracy in differentiating between lesion types [14]. T.M. Alam et al. (2022) proposed a novel CNN architecture trained on HAM10000, focusing on improving classification performance through data augmentation techniques [15]. H.A. Owida et al. (2024) investigated the use of transfer learning with pre-trained models on HAM10000 to enhance melanoma detection [16]. A. Ameri et al. (2020) explored feature extraction methods using HAM10000 to improve the interpretability of DL models [17]. S.S. Chaturvedi et al. (2020) conducted a comparative study of DL algorithms on the HAM10000 dataset to identify the most effective approaches for skin lesion classification [18].

Non-ISIC Data. While the ISIC Archive and HAM10000 datasets are widely used in recent skin lesion research, several other datasets have historically contributed to the field, particularly before the availability of ISIC data.

The **PH²** dataset [19], developed by Pedro Hispano Hospital in Portugal, contains 200 dermoscopic images focusing on melanocytic nevi and melanomas. Despite its small size, the dataset is highly valued for its detailed manual segmentation masks and clinical annotations that include colors and dermoscopic structures. It is often used in studies emphasizing precise lesion segmentation and the analysis of dermoscopic features.

The **MED-NODE** dataset [20], created by the University of Twente in the Netherlands, includes 1,700 clinical (non-dermoscopic) images of melanomas and benign melanocytic nevi. Diagnoses in this dataset are confirmed through histopa-

thology. It is particularly useful for research focusing on classification based on clinical photographs rather than dermoscopic images. This dataset presents challenges like variations in lighting, skin tone, and image quality, making it a valuable resource for developing models that can handle real-world conditions.

The **Dermofit** Image Library [21], developed by the University of Edinburgh, includes approximately 1,300 high-resolution images spanning 10 skin lesion categories. However, it is accessible only under a commercial license, which can limit its availability for some researchers. The library is often used in studies requiring high-resolution images and advanced feature extraction techniques.

The **SD-198** dataset [22], compiled from dermatology atlases, contains 6,584 images across 198 different skin disease classes. It is particularly valuable for multi-class classification research and for developing models that can identify rare skin conditions.

These datasets, while smaller in scale compared to ISIC, offer unique advantages by addressing diverse imaging conditions and focusing on specific research challenges, continuing to play a crucial role in dermatology research.

Data Challenges. *Class Imbalance.* Datasets often have a disproportionately large number of benign lesions compared to malignant ones. Such skewed data causes model bias and poor sensitivity in detecting rare but critical cases like melanomas. This problem is typical for skin cancer data. For instance, in the ISIC 2020 dataset, the benign nevi-to-melanoma ratio is approximately 9 to 1 [5]. If not handled properly, this imbalance can lead to high overall accuracy but poor performance when detecting minority classes, which are crucial for early diagnosis.

Underrepresentation of Diverse Skin Types. In recent years, it was pointed out that available skin cancer data is insufficient in terms of darker skin tone representation. It can affect a model's accuracy across different ethnicities and exacerbate health disparities [11], [23]. The research community considers this a serious problem as, even though skin cancer is less prevalent among individuals with darker skin, it is often detected at later stages, leading to drastic outcomes. For instance, Black individuals are three times more likely to be diagnosed with late-stage melanoma [24]. One of the reasons is that melanoma for that part of the population sometimes appears on less visible areas, such as the palms, soles, or under the nails, which are less frequently examined [25].

Privacy Concerns and Data Sharing Limitations. Skin cancer imaging data is part of a person's personal information. Consequently, it falls under many laws and acts aimed at protecting users' data. As this data is of medical nature, it complicates the situation even more. Among the commonly known regulations are:

- Health Insurance Portability and Accountability Act [26] (HIPAA) in the United States.
- The General Data Protection Regulation [26] (GDPR) in the European Union.

These regulations impose strict guidelines on the handling, sharing, and processing of medical data. Understanding these restrictions is crucial for researchers aiming to access and utilize dermatological datasets while ensuring compliance with legal and ethical standards.

Reluctance to Share Proprietary Datasets. Self-collected datasets offer researchers distinct competitive advantages. This method of data acquisition allows influencing the data flow from the very start, forming a more precise understanding of the data nature. At the same time, exclusive access to this data lets institu-

tions uniquely fine-tune algorithms for enhanced performance, potentially leading to breakthroughs in skin cancer detection and diagnosis. Sulaiman Khan et al. (2022) [28] found that researchers often use private datasets exclusively or combine them with open data to achieve superior results in skin cancer classification tasks. Other reasons to withhold data might range from ethical to licensing concerns. Another issue is data privacy. This practice highlights the advantages of proprietary data but also underscores a significant barrier to progress in the field. When models are trained on private datasets, it becomes challenging to replicate studies, validate findings, or compare the effectiveness of different methodologies.

Data Labeling. Many of the publicly available skin cancer image samples are currently unlabeled, which makes them impossible to use with supervised DL algorithms. For instance, the ISIC 2020 dataset [5] includes 10,982 images without assigned classes. Meanwhile, accurately annotating medical images requires a high level of domain expertise, making the process both time-consuming and costly.

Artefacts in Image Acquisition. The most common way in which skin cancer images are obtained is dermatoscopy. It is a non-invasive imaging technique that involves examining and capturing skin lesions via a dermatoscope. Nevertheless, images of a dataset are often collected in different centers and institutions that utilize instruments with varying characteristics. However, when all the data is collected by a single facility, it may cause variations in resulting images, leading to models overfitting to device-specific artifacts instead of focusing on generic lesion features. Studies have shown that models trained on homogeneous datasets perform poorly on images from diverse sources [29], [30].

Ethical Considerations. Besides norms and regulations, one should realize that medical data is deeply personal. Processing it unavoidably raises some ethical concerns. Syed F.H. Shah in "Ethical implications of artificial intelligence in skin cancer diagnostics: use-case analyses" (2024) found that existing skin cancer analysis solutions require a great deal of transparency and collaboration to avoid potential ethical problems and misuse [31].

To resolve the ethical issues, the research community must show awareness when handling such sensitive data. In the paper "Ethical considerations for artificial intelligence in dermatology: a scoping review," Emily R. Gordon (2024) identified key principles to follow to avoid ethics-related risks like fairness, inclusivity, transparency, accountability, etc. [32].

Additional Clinical Criteria. While lesion images serve as the primary source of information, various studies suggest that including more contextual data can positively impact classification performance. Esteva et al. (2017) noted that including clinical data could further enhance performance [4]. Haenssle et al. (2018) demonstrated that combining dermatoscopic images with patient information improved melanoma detection rates [33]. Pacheco and Krohling (2019) found that integrating clinical metadata with images in DL models led to higher accuracy in skin lesion classification [34]. These additional criteria may vary: patient age is important because certain age groups are more inclined to have some types of skin cancer than others, evolution over time is one of the key characteristics that experts use to make a decision on lesion state, as well as the size of the lesion. Embedding mathematical measures derived from the image, such as fractal dimension, may also contain insights useful for correct classification [35].

Incorporating these features into a classification model is a new challenge, as image and contextual data normally have different modalities. Making a model more complex to handle this may cause unintended overfitting. Therefore, the challenge is to balance model complexity while integrating features effectively.

Preprocessing Stage. In the previous section the key data challenges were identified. Many of them such as class imbalance, ethnic groups underrepresentation, restricted data, unlabeled data, inconsistency in the data gathering tools and noise can be addressed on a data preprocessing stage. In this section the key steps of data preparation are discussed.

Balancing Data. Some datasets do not suffer from class imbalance issues. For example, Kaggle dataset for malignant melanoma classification [36] contains approximately 5000 samples of each class which makes it a good option for fine tuning lightweight melanoma classifiers for exploration and education purposes.

Nevertheless, most of the available data samples are severely imbalanced with the largest amount of lesions being benign. For instance, HAM1000 dataset contains around 60% of benign nevus samples. Provided that this dataset is used for multi class classification, it makes it very imbalanced

Data Augmentation. A typical method for handling imbalanced data involves performing augmentations on the training data. Shen (2022) proposes an effective way of augmenting data for DL skin cancer classification with a significantly reduced search space of 60 possible transformations, compared to existing methods like AutoAugment and RandAugment [37]. Himel et al. (2024) applied data augmentation to 3,295 malignant images in the HAM10000 dataset, increasing the number to 5,000 [38]. These augmentations included rotation, flipping, and zooming. However, their approach may seem somewhat straightforward, as they simultaneously disregarded more than 1,500 benign lesion images to balance the data. This was also risky, as it essentially balanced the data through data loss. Polat et al. (2020) augmented images of the same dataset with noise, scaling, and rotation, increasing the total number of images to more than 33,000 [39]. Walker et al. (2019) used cropping in addition to the aforementioned methods while working with the ISIC 2017 Challenge [5], [40]. Ali et al. (2021) applied color-shifting using principal component analysis to create augmented images [41].

Another way of augmenting images is application of Generative adversarial network (GAN) architecture. Wu et al. (2020) did a review of GAN application for augmenting skin cancer images to solve data imbalance problem [42]. In article they note that various GAN architectures, including DCGANs [43], style-based GANs [44], TED-GAN [45], SPGGAN [46], and conditional GANs (CGANs) [47], have been employed to generate high-quality, diverse synthetic images. Enhancements such as artifact removal, attention modules, informative noise vectors, and stability improvements like the Two-Timescale Update Rule have further optimized GAN performance. Wu state that these methods have consistently improved classification metrics — including accuracy, sensitivity, specificity, and F1 scores — by providing richer training data and more reliable image generation, ultimately enhancing the effectiveness of skin cancer classification models.

While this approach offers benefits, it also comes with potential risks and drawbacks. Excessive noise, extreme scaling, or unusual rotation angles can distort images to the point where essential features are obscured or altered. If the augmentation methods produce images that are too similar to each other, the model might overfit to these synthetic variations rather than learning generalized features. This reduces the model's ability to perform well on truly unseen data. Adding noise can sometimes degrade image quality to a level where the model's ability to extract meaningful features is compromised. Drawbacks also include increased computational cost of creating and processing larger amount of data.

Fixing Ethnic Representation Issues. To mitigate the issue of underrepresentation, several strategies can be employed to improve both the diversity of data and the fairness of DL models.

- Color jittering, which adjusts the brightness, contrast, and saturation of images, can simulate a broader range of skin tones.
- Histogram equalization can improve the visibility of features across different skin tones.
- Style transfer can modify images to appear as if they belong to varied skin types.

Pope et al. (2024) compared training model on imbalanced and sampled datasets [48]. However, their results show that, albeit the sampled model tends to be less biased towards dark-toned skin, the overall accuracy decreases.

Beyond augmentation, synthetic data generation can further expand datasets, particularly for underrepresented skin tones. GANs are useful for generating synthetic images that mimic lesions on darker skin, while domain adaptation techniques can align feature representations to ensure models perform well across different skin tones. Rezk et al. (2022) composed a skin cancer dataset from existing open access data and applied style transferring to make it more diverse in terms of skin tone [49].

There also were instances of less traditional approach to solve this problem. Continuing their study of 2019, Walker et al. (2024) used sonification of skin cancer data to decrease influence of skin tone on classification [50]. The analysis demonstrated high and comparable diagnostic accuracy for both fair skin (FS) and darker skin (DS), with ROC curve AUCs of 0.858 and 0.856, respectively, indicating no significant difference between the two groups. Sensitivity and specificity values were similar (around 80–85%), confirming the model's consistent performance in detecting true positives and negatives across diverse skin tones. The results suggest that the classifier maintains equivalent diagnostic reliability for both FS and DS, even with the limitations of smartphone-based imaging.

To ensure fairness in model performance, algorithmic adjustments can be utilized. Reweighting samples can help balance training by assigning higher weights to underrepresented classes, while bias correction layers can be integrated into models to correct inherent biases.

Data Cleaning. Data cleaning is a crucial step in DL, involving the preparation and preprocessing of data to improve its quality for model training [51]. This process includes various data processing techniques that enable more effective feature extraction, ultimately enhancing model performance.

Normalization. To address inconsistencies in the conditions under which images of skin lesions are captured, data normalization and denoising are essential. Normalization techniques, such as Gray World, Shades of Gray, or max-RGB algorithms, adjust the color balance of images to compensate for lighting differences [41], [52]. Global or adaptive histogram equalization methods, like Contrast Limited Adaptive Histogram Equalization (CLAHE), can be applied to improve contrast and standardize the intensity distribution across images [53], [54]. Gamma correction is another option, adjusting non-linear luminance or color values to standardize brightness and contrast.

Standard normalization methods, such as scaling pixel values to have zero mean and unit variance (z-score normalization) or scaling between 0 and 1 (min-

max normalization), are commonly used in image preprocessing [55]. For example, 0–1 normalization was utilized in [41]. These steps ensure that images fed into DL models are consistent and of high quality.

Noise Reduction. Removing noise from images is critical to enhance quality and improve model accuracy. Common strategies include applying filters like median or Gaussian filters to smooth images and reduce noise. Midasala et al. (2024) applied a top-hat transform to remove thick hairs, while filters effectively eliminated noise and thin hairs, albeit with limitations on contrast-based histogram equalization [56]. Morphological operations, such as opening and closing, assist in removing small artifacts.

Body hair presents a specific challenge in analyzing skin images, as it can obscure important lesion features. Algorithms like the DullRazor detect hair pixels using edge detection and inpaint these regions to remove hair from the image [57]. Thresholding methods identify hair regions so that they can be cleaned out.

Another approach to denoising is the use of autoencoders. These models learn to create a compressed representation of the input and reconstruct it, effectively removing noise in the process. Maurya et al. (2024) used autoencoders for denoizing and feature extraction [58].

Segmentation. Segmentation is another crucial preprocessing step, involving the selection of the region of interest (ROI) from the image—in this case, the skin lesion. Accurate segmentation focuses analysis on the lesion and removes irrelevant background information. This topic is well-researched, and researchers often utilize large pretrained models.

For instance, Himel et al. (2024) [38] used Meta's Segment Anything Model [59] to perform segmentation on skin cancer images from the HAM10000 dataset. They then used the segmented images to pass only the ROI to feature extraction and classification models, achieving 96% accuracy using Google's Vision Transformer (ViT) [60]. This approach demonstrates the effectiveness of combining advanced segmentation models with powerful classification architectures.

TRAINING AND OPTIMIZATION

Model training and optimization are pivotal in developing robust DL models for skin cancer classification. These processes involve selecting suitable architectures, optimizing learning algorithms, fine-tuning hyperparameters, and employing strategies to enhance model performance while addressing overfitting.

Transfer Learning. Transfer learning is extensively utilized in medical image analysis due to the scarcity of labeled datasets [11]. By leveraging models pre-trained on large-scale datasets like ImageNet, researchers can fine-tune these models for specific tasks such as skin cancer classification [4]. This approach not only accelerates convergence but also often yields superior performance compared to training models from scratch.

In skin cancer classification, the data available is limited for many reasons. Therefore, usage of pretrained models is popular approach. Ali et al. (2021) [41], while developing a custom NN for skin cancer classification, applied transfer learning to the task with pre trained ResNet [61], AlexNet [62], VGG-16 [63], DenseNet [64] and MobileNet [65] achieving top accuracy of 86.09. Another example is using ViT in [60] with 0.96 accuracy result.

Optimization Techniques. *Learning Rate Scheduling.* Adjusting the learning rate during training can significantly influence model convergence. Techniques such as Step Decay, Exponential Decay, and more advanced methods like Cyclical Learning Rates and Warm Restarts are employed to optimize training efficiency [66, 67].

Optimizers. Selecting an appropriate optimization algorithm is crucial for training deep neural networks. Optimizers like Stochastic Gradient Descent (SGD) with momentum, Adam, and RMSProp are widely used. Adaptive optimizers like Adam combine the advantages of AdaGrad and RMSProp, providing efficient training for deep models [68]. In skin cancer classification tasks, Adam is often preferred for its ability to handle sparse gradients and noisy data (for instance in [69]).

Regularization Methods. Regularization techniques prevent overfitting by adding constraints to the model. L1 and L2 regularization add penalties to the loss function based on the magnitude of weights. Dropout randomly deactivates neurons during training, reducing interdependent learning among neurons. Batch normalization normalizes layer inputs, accelerating training and improving generalization.

It is also worth mentioning that on this stage is possible to address class imbalance problem if it was not solved completely on data preprocessing stage. Le et al. (2020) used focal loss to train on imbalanced data of HAM10000 and received 0.94 top accuracy with pretrained EfficientNetB1 [70].

Hyperparameter Tuning. Hyperparameters such as learning rate, batch size, network depth, and activation functions significantly impact model performance. Systematic methods like Grid Search and Random Search explore combinations of hyperparameters, while Bayesian Optimization offers a more efficient search by modeling the performance as a probabilistic function.

Same time hyperparameter tuning can always be costly no matter the technique chosen. In [69] it was done on only 10% of training dataset which allowed to select appropriate values more efficiently.

Early Stopping. Early stopping halts training when the validation loss stops decreasing, preventing the model from overfitting to the training data. This method is especially useful when training deep networks on limited datasets, as is common in medical imaging.

Ensemble Learning. Ensembling combines predictions from multiple models to improve robustness and accuracy. Techniques like averaging, majority voting, or stacking can enhance performance in skin cancer classification. Ghosh et al (2024) [71] utilized ensemble learning with DCNN [72], Caps-Net [73] and ViT [60].

Non Supervised Learning. *Semi-supervised learning*. Semi-supervised learning combines a small amount of labeled data with a large amount of unlabeled data during training. This approach is particularly beneficial in medical imaging, where labeling is expensive and time-consuming. Liu et al. (2020) employed a semi-supervised learning approach using a mean teacher model to leverage unlabeled skin lesion images, improving classification performance [7]. Techniques:

- **Pseudo-Labeling**: Assigning labels to unlabeled data using the model's own predictions and then retraining the model with this expanded dataset. This iterative process can improve performance but may propagate errors if the initial model is not accurate.

- **Consistency Regularization**: Encouraging the model to produce similar outputs when input images are perturbed or augmented, leveraging unlabeled data to learn robust features.
- **Mean Teacher Model**: Utilizing a student-teacher framework where the teacher model generates targets for the student model, which learns from both labeled and unlabeled data. A noisy stutent algorithm was successfully used in ISIC Kaggle competition in melanoma classification challenge [74].
- **Graph-Based Methods**: Modeling data as graphs where nodes represent samples and edges represent similarities, propagating labels through the graph to infer labels for unlabeled data.

Self-Supervised Learning. Self-supervised learning is a form of unsupervised learning where the model learns representations by solving pretext tasks created from unlabeled data. This approach is gaining traction in medical imaging. Chaitanya et al. (2020) showed that self-supervised learning on unlabeled medical images significantly improves model performance on downstream tasks with limited labeled data [75]. Techniques:

- **Contrastive Learning**: Learning representations by maximizing agreement between differently augmented views of the same data sample. The key idea is to develop a non-trivial network that preserves similar semantic structure for two (somewhat modified) versions of the same image and keeps two different images distinguishable [76]. Models like SimCLR have been adapted for medical images to learn robust features from unlabeled data [77].
- **Pretext Tasks**: Designing tasks such as predicting image rotations, solving jigsaw puzzles, or reconstructing distorted images to force the model to learn meaningful features. Haggerty and Chandra (2024) showed that models pre-trained using SSL (Barlow Twins) significantly outperformed those pre-trained with SL on ImageNet in scenarios with limited labeled data specifically using skin cancer images. Moreover, by applying additional SSL pre-training on smaller, task-specific datasets (like skin lesion images), SL-pre-trained models could achieve performance equivalent to SSL models. This demonstrates that even minimal further SSL pre-training can be as effective as extensive pre-training on large datasets [78].

EVALUATION AND VALIDATION

Next important step in the DL pipeline is assessing the model's effectiveness. It involves measuring the model's performance on previously unseen data. Testing dataset is distinct from the validation dataset used during training. The validation dataset helps evaluate the model's progress and detect overfitting, enabling techniques like early stopping to be applied.

To accurately identify model quality, an appropriate evaluation strategy and metrics must be chosen. For instance, when working with relatively small training and testing datasets, it's crucial to focus on validating whether observed improvements stem from the new approach rather than statistical noise.

Selecting suitable metrics is equally important. In medical diagnostics, for example, both type I and type II errors carry significant consequences. However, it is generally considered "better" to classify a healthy patient as sick than to miss diagnosing an ill patient. Moreover, medical datasets, such as those related to skin cancer, are often highly imbalanced. In such cases, relying solely on accuracy as the primary metric can be misleading.

Validation Strategies. Selecting fitting strategy involves deep understanding of the data. It normally involves investigating the variability of model results so that we can distinguish a "luck" from "an actual improvement". However, in medical imaging, datasets are often limited in size, making this method susceptible to high variance in performance estimates.

Cross-Validation. **K-fold cross-validation (CV)** divides the dataset into k subsets (folds). The model is trained on k-1 folds and validated on the remaining fold, a process repeated k times. The results are averaged to produce a performance estimate [79].

Stratified K-fold ensures that each fold maintains the same class distribution as the original dataset, which is crucial for imbalanced datasets common in skin cancer classification. Mahesh et al. (2023) employed stratified K-Fold CV to handle class imbalance problem in [80].

Nested cross-validation addresses the bias in performance estimation due to hyperparameter tuning by having an inner loop for model selection and an outer loop for model assessment.

External validation. This approach involves testing the model on entirely independent datasets from different institutions or populations. This approach provides a robust assessment of the model's generalizability [7]. Brinker et al. (2019) performed an external validation of a DL model for melanoma detection across different populations, emphasizing the necessity of external validation for assessing generalizability [81].

Evaluation Metrics. Accuracy measures the proportion of correct predictions but can be misleading with imbalanced datasets. Precision shows how many identified class samples were identified correctly. Recall (Sensitivity) is how many samples of a class the model was able to identify. Specificity measures the proportion of true negatives among all actual negatives

F1-Score is harmonic mean of precision and recall:

$$F1 = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}}.$$

Balanced accuracy accounts for class imbalance by averaging the recall obtained on each class:

$$Balanced\ Accuracy = \frac{Precision + Recall}{2}.$$

These metrics provide a balanced view of performance, especially important in medical diagnosis where false negatives and false positives have different implications. They are recommended to be used together to get a different perspectives in performance.

Another set of effective metrics is Receiver Operating Characteristic (ROC) Curve and Area Under the Curve (AUC). Receiver Operating Characteristic is a graph with a y-axis representing Sensitivity and an x-axis representing 1 – Specificity. It represents relation of classification threshold and correctly classified samples. In order to easily compare different models AUC is employed — an area under ROC Curve.

Han et al. (2020) utilized ensemble learning and evaluated their model using ROC-AUC, precision-recall curves, and conducted statistical significance testing to demonstrate improvements over previous methods [82].

OPPORTUNITIES FOR INNOVATION

In the domain of DL-based skin cancer classification, there are several promising avenues for innovation that can significantly enhance diagnostic accuracy, patient outcomes, and system scalability. As the field continues to mature, the integration of cutting-edge technologies can address existing challenges, optimize model performance, and broaden access to dermatological diagnostics.

Multimodal Data Integration. Integrating clinical data (e.g., patient demographics, lesion history, and genetic markers) with imaging data has the potential to improve diagnostic accuracy. While current models primarily rely on dermoscopic images, including non-visual patient information can provide additional context, leading to more accurate predictions. For instance, factors like age, lesion location, and personal/family history of skin cancer are crucial in assessing melanoma risk. Models that combine imaging with clinical data have demonstrated improved sensitivity and specificity, particularly for complex cases [83]. Moldovanu et al. (2021) used surface fractal dimensions and statistical color cluster features to improve model quality. Future research should focus on developing architectures capable of effectively fusing diverse data types. Nikitin et al. utilised fractal dimension with ViT focusing on different ways of integration of the metric [84]. Future research should focus on developing architectures capable of effectively fusing diverse data types.

Explainable AI. Building trust in AI-driven diagnostics is essential for clinical adoption. Techniques like Gradient-weighted Class Activation Mapping (Grad-CAM) [85] and Layer-wise Relevance Propagation (LRP) [86] offer explainability by highlighting regions in dermoscopic images that contributed to the model's decision. These visual explanations can help clinicians understand the model's reasoning, enabling them to verify the output and diagnose more confidently. Explainable AI can also aid in identifying potential biases in the model, particularly concerning underrepresented skin tones, thereby addressing disparities in diagnostic outcomes. Future work should explore enhancing the interpretability of DL models while maintaining high classification accuracy, especially in high-stakes settings like oncology.

Edge Computing. Deploying lightweight DL models on edge devices, such as smartphones, can facilitate real-time skin cancer detection, especially in resource-constrained settings. Advances in model optimization, such as pruning, quantization, and using architectures like MobileNet and EfficientNet, can reduce the computational load while maintaining accuracy. This is particularly relevant for underserved regions where access to dermatologists is limited. Edge-based AI systems can provide preliminary assessments, encouraging users to seek medical consultation if a lesion is flagged as suspicious. Research in this area should prioritize developing robust, efficient models that can handle diverse image quality and environmental conditions typical of mobile device usage.

GUIDANCE FOR EMERGING RESEARCHERS

The field of DL for skin cancer classification is both challenging and rewarding, offering ample opportunities for innovation. However, newcomers to the field may face a steep learning curve due to the complexity of data, algorithms, and clinical considerations. Here are some practical tips for emerging researchers to navigate this evolving landscape:

Guidance for Emerging Researchers. The field of DL for skin cancer classification is both challenging and rewarding, offering ample opportunities for innovation. However, newcomers to the field may face a steep learning curve due to the complexity of data, algorithms, and clinical considerations. Here are some practical tips for emerging researchers to navigate this evolving landscape:

1. **Focus on Data Quality and Preprocessing**. One of the biggest hurdles in dermatological AI research is access to high-quality, annotated datasets. Begin by familiarizing yourself with widely used datasets like ISIC and HAM10000. Pay special attention to data preprocessing, including normalization, augmentation, and segmentation techniques, to optimize model performance. Addressing challenges like class imbalance and image noise is crucial for achieving reliable results. However, the real experiments to achieve top metrics results must be conducted on larger data since it is available — as of now ISIC Archive includes more than 400,000 images available to use.

2. **Start with Transfer Learning and Fine-Tunin**g. Given the limited availability of labeled medical data (ISIC Archive is not nearly as big as ImageNet), leveraging transfer learning from pre-trained models on large-scale datasets like ImageNet can accelerate progress. Fine-tuning these models on skin lesion datasets can yield competitive results with relatively less training time. Explore architectures such as ResNet, EfficientNet, and ViT to identify what works best for your specific use case. Also, empirical results show that ensemble models do great in the task.

3. **Embrace Explainability from the Start**. Building interpretability into your models is essential, especially in healthcare applications where clinicians require transparent and understandable outputs. Experiment with tools like Grad-CAM and SHAP [87] to visualize your model's decision-making process. Prioritizing explainability will not only help you gain the trust of the medical community but also ensure that your models can be safely deployed in clinical settings.

4. **Keep Ethics and Privacy at the Forefront**. When dealing with sensitive medical data, ethical considerations and compliance with regulations like GDPR are paramount. Consider approaches like federated learning and differential privacy to ensure patient confidentiality. Being aware of these considerations early on will help you design ethically responsible research projects that can be more easily translated into real-world applications.

5. **Stay Updated on Emerging Trends**. The field of AI in healthcare is rapidly evolving, with new techniques like self-supervised learning, multimodal models, and quantum ML showing potential. Regularly reviewing the latest research, participating in conferences, and engaging with collaborative research communities like the ISIC Challenge can keep you at the forefront of the field. Additionally, exploring adjacent fields like personalized medicine and bioinformatics can open up interdisciplinary opportunities.

By focusing on these areas, emerging researchers can build a strong foundation and contribute meaningfully to the development of AI-driven skin cancer diagnostics.

CONCLUSION

DL has revolutionized the field of skin cancer diagnostics, offering tools that can potentially match, or even surpass, dermatologist-level performance. However, the journey from research to clinical application is fraught with challenges. Our analysis highlights the importance of high-quality data, rigorous preprocessing, and robust model evaluation in developing reliable diagnostic systems. Address-

ing issues like data privacy, class imbalance, and underrepresentation of diverse skin tones remains critical to ensuring equitable healthcare outcomes.

The integration of clinical metadata with imaging data, along with techniques such as federated learning and edge computing, presents promising avenues to enhance model performance and broaden access to diagnostics, particularly in resource-constrained regions. Additionally, incorporating explainable AI methodologies can help gain the trust of clinicians, paving the way for real-world adoption.

For emerging researchers, focusing on ethical considerations, leveraging transfer learning, and embracing advancements in multimodal integration are vital steps toward impactful contributions in this domain. As AI continues to evolve, its role in skin cancer diagnostics will likely expand, enabling more accurate, accessible, and personalized care. Future research should aim at bridging the gap between technological capabilities and clinical needs, ultimately transforming the landscape of dermatological diagnostics and improving patient outcomes.

REFERENCES

1. *Cancer Research UK*. Available: https://www.cancerresearchuk.org/
2. *Health workforce: The health workforce crisis.* Available: https://www.who.int/news-room/questions-and-answers/item/q-a-on-the-health-workforce-crisis
3. L.G. Wich et al., "Impact of Socioeconomic Status and Sociodemographic Factors on Melanoma Presentation Among Ethnic Minorities," *Journal of Community Health*, vol. 36, no. 3, pp. 461–468, Nov. 2010. doi: 10.1007/s10900-010-9328-4.
4. A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, Jan. 2017. doi: 10.1038/nature21056.
5. *ISIC International Skin Imaging Collaboration*. Available: https://www.isic-archive.com/
6. "Can artificial intelligence detect melanoma?" *Memorial Sloan Kettering Cancer Center*, Jul. 24, 2024. Available: https://www.mskcc.org/news/can-artificial-intelligence-detect-melanoma
7. Y. Liu et al., "A deep learning system for differential diagnosis of skin diseases," *Nature Medicine*, vol. 26, no. 6, pp. 900–908, May 2020. doi: 10.1038/s41591-020-0842-3.
8. A. Yilmaz, M. Kalebasi, Y. Samoylenko, M.E. Guvenilir, H. Uvet, "Benchmarking of Lightweight Deep Learning Architectures for Skin Cancer Classification using ISIC 2017 Dataset," *arXiv* (*Cornell University*), Jan. 2021. doi: 10.48550/arxiv.2110.12270.
9. M. Baygin, T. Tuncer, S. Dogan, "New pyramidal hybrid textural and deep features based automatic skin cancer classification model: Ensemble DarkNet and textural feature extractor," *arXiv (Cornell University)*, Jan. 2022. doi: 10.48550/arxiv.2203.15090.
10. K. Agarwal, T. Singh, "Classification of Skin Cancer Images using Convolutional Neural Networks," *arXiv (Cornell University)*, Jan. 2022. doi: 10.48550/arxiv.2202.00678.
11. D. Wen et al., "Characteristics of publicly available skin cancer image datasets: a systematic review," *The Lancet Digital Health*, vol. 4, no. 1, pp. e64–e74, Nov. 2021. doi: 10.1016/s2589-7500(21)00252-1.
12. "ISIC 2024 — Skin Cancer Detection with 3D-TBP," *Kaggle*. Available: https://www.kaggle.com/competitions/isic-2024-challenge
13. P. Tschandl, C. Rosendahl, H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific Data*, vol. 5, no. 1, Aug. 2018. doi: 10.1038/sdata.2018.161.
14. A.T. Priyeshkumar, G. Shyamala, T. Vasanth, Selvan V. Ponniyin, "Transforming Skin Cancer Diagnosis: A Deep Learning Approach with the Ham10000 Dataset," *Cancer Investigation*, vol. 42, no. 10, pp. 801–814, Nov. 2024. doi: 10.1080/07357907.2024.2422602.
15. T.M. Alam et al., "An efficient Deep Learning-Based skin cancer classifier for an imbalanced dataset," *Diagnostics*, vol. 12, no. 9, p. 2115, Aug. 2022. doi: 10.3390/diagnostics12092115.
16. H.A. Owida et al., "Improved deep learning architecture for skin cancer classification," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 36, no. 1, p. 501, Jul. 2024. doi: 10.11591/ijeecs.v36.i1.pp501-508.
17. A. Ameri, "A deep learning approach to skin cancer detection in dermoscopy images," *Journal of Biomedical Physics and Engineering*, vol. 10, no. 6, Nov. 2020. doi: 10.31661/jbpe.v0i0.2004-1107.

18. S.S. Chaturvedi, J.V. Tembhurne, T. Diwan, "A multi-class skin Cancer classification using deep convolutional neural networks," *Multimedia Tools and Applications*, vol. 79, no. 39–40, pp. 28477–28498, Aug. 2020. doi: 10.1007/s11042-020-09388-2.
19. *ADDI - Automatic computer-based Diagnosis system for Dermoscopy Images*. Available: https://www.fc.up.pt/addi/ph2%20database.html)
20. I. Giotis, N. Molders, S. Land, M. Biehl, M.F. Jonkman, N. Petkov, "MED-NODE: A computer-assisted melanoma diagnosis system using non-dermoscopic images," *Expert Systems With Applications*, vol. 42, no. 19, pp. 6578–6585, May 2015. doi: 10.1016/j.eswa.2015.04.034.
21. Dermofit Image Library available from The University of Edinburgh. Available: https://licensing.edinburgh-innovations.ed.ac.uk/product/dermofit-image-library
22. X. Sun, J. Yang, M. Sun, K. Wang, "A benchmark for automatic visual classification of clinical skin disease images," in *Lecture notes in computer science*, 2016, pp. 206–222. doi: 10.1007/978-3-319-46466-4_13.
23. "Datasets used to train AI to detect skin cancer lack information on darker skin and often incomplete," *Haiku*, Nov. 10, 2021. Available: https://www.cancer.ox.ac.uk/ news/lack-of-data-available-to-detect-skin-cancer-in-darker-skin
24. S. Hu, R.M. Soza-Vento, D.F. Parker, R.S. Kirsner, "Comparison of stage at diagnosis of melanoma among Hispanic, Black, and white patients in Miami-Dade County, Florida," *Archives of Dermatology*, vol. 142, no. 6, Jun. 2006. doi: 10.1001/archderm.142.6.704.
25. H.M. Gloster, K. Neal, "Skin cancer in skin of color," *Journal of the American Academy of Dermatology*, vol. 55, no. 5, pp. 741–760, Oct. 2006. doi: 10.1016/j.jaad.2005.08.063.
26. "Health Insurance Portability and Accountability Act (HIPAA)", *U.S. Department of Health and Human Services*. Available: https://www.hhs.gov/hipaa/index.html
27. "General Data Protection Regulation (GDPR) – Legal Text," *General Data Protection Regulation (GDPR)*, Apr. 22, 2024. Available: https://gdpr-info.eu/
28. S. Khan, H. Ali, Z. Shah, "Identifying the role of vision transformer for skin cancer — A scoping review," *Frontiers in Artificial Intelligence*, vol. 6, Jul. 2023. doi: 10.3389/frai.2023.1202990.
29. P. Tschandl et al., "Human–computer collaboration for skin cancer recognition," *Nature Medicine*, vol. 26, no. 8, pp. 1229–1234, Jun. 2020. doi: 10.1038/s41591-020-0942-0.
30. T.J. Brinker et al., "Deep neural networks are superior to dermatologists in melanoma image classification," *European Journal of Cancer*, vol. 119, pp. 11–17, Aug. 2019. doi: 10.1016/j.ejca.2019.05.023.
31. S.F.H. Shah, D. Arecco, H. Draper, S. Tiribelli, E. Harriss, R.N. Matin, "Ethical implications of artificial intelligence in skin cancer diagnostics: use-case analyses," *British Journal of Dermatology*, Nov. 2024. doi: 10.1093/bjd/ljae434.
32. E.R. Gordon et al., "Ethical considerations for artificial intelligence in dermatology: a scoping review," *British Journal of Dermatology*, vol. 190, no. 6, pp. 789–797, Feb. 2024. doi: 10.1093/bjd/ljae040.
33. H.A. Haenssle et al., "Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists," *Annals of Oncology*, vol. 29, no. 8, pp. 1836–1842, May 2018. doi: 10.1093/annonc/mdy166.
34. A.G.C. Pacheco, R.A. Krohling, "Recent advances in deep learning applied to skin cancer detection," *arXiv (Cornell University)*, Jan. 2019. doi: 10.48550/arxiv.1912.03280.
35. P. Popecki, M. Kozakiewicz, M. Ziętek, K. Jurczyszyn, "Fractal Dimension Analysis of Melanocytic nevi and melanomas in Normal and Polarized Light—A Preliminary Report," *Life*, vol. 12, no. 7, p. 1008, Jul. 2022. doi: 10.3390/life12071008.
36. "Melanoma Skin Cancer Dataset of 10000 Images," *Kaggle*, Mar. 29, 2022. Available: https://www.kaggle.com/datasets/hasnainjaved/melanoma-skin-cancer-dataset-of-10000-images
37. S. Shen et al., "A Low-Cost High-Performance data augmentation for Deep Learning-Based skin lesion classification," *BME Frontiers*, 2022. doi: https://doi.org/10.34133/2022/9765307
38. G.M.S. Himel, Md.M. Islam, Kh.A. Al-Aff, S.I. Karim, Md.K.U. Sikder, "Skin Cancer Segmentation and Classification Using Vision Transformer for Automatic Analysis in Dermatoscopy-Based Noninvasive Digital System," *International Journal of Biomedical Imaging*, vol. 2024, pp. 1–18, Feb. 2024. doi: 10.1155/2024/3022192.

39. K. Polat, K.O. Koc, "Detection of Skin Diseases from Dermoscopy Image Using the combination of Convolutional Neural Network and One-versus-All," *Journal of Artificial Intelligence and Systems*, vol. 2, no. 1, pp. 80–97, Jan. 2020. doi: 10.33969/ais. 2020.21006.
40. B. Walker et al., "Dermoscopy diagnosis of cancerous lesions utilizing dual deep learning algorithms via visual and audio (sonification) outputs: Laboratory and prospective observational studies," *EBioMedicine*, vol. 40, pp. 176–183, 2019. doi: https://doi.org/10.1016/j.ebiom.2019.01.028
41. M.S. Ali, M.S. Miah, J. Haque, M.M. Rahman, M.K. Islam, "An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models," *With Applications*, vol. 5, 100036, 2021. doi: https://doi.org/ 10.1016/j.mlwa.2021.100036
42. Y. Wu, B. Chen, A. Zeng, D. Pan, R. Wang, S. Zhao, "Skin Cancer Classification With Deep Learning: A Systematic Review," *Frontiers in Oncology*, vol. 12, Jul. 2022. doi: 10.3389/fonc.2022.893972.
43. D. Bisla, A. Choromanska, R.S. Berman, J.A. Stein, D. Polsky, "Towards Automated Melanoma Detection With Deep Learning: Data Purification and Augmentation," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Jun. 2019*. doi: 10.1109/cvprw.2019.00330.
44. Z. Qin, Z. Liu, P. Zhu, Y. Xue, "A GAN-based image synthesis method for skin lesion classification," *Computer Methods and Programs in Biomedicine*, vol. 195, p. 105568, May 2020. doi: 10.1016/j.cmpb.2020.105568.
45. B. Ahmad, S. Jun, V. Palade, Q. You, L. Mao, M. Zhongjie, "Improving Skin Cancer Classification Using Heavy-Tailed Student T-Distribution in Generative Adversarial Networks (TED-GAN)," *Diagnostics*, vol. 11, no. 11, p. 2147, Nov. 2021. doi: 10.3390/diagnostics11112147.
46. I.S.A. Abdelhalim, M.F. Mohamed, Y.B. Mahdy, "Data augmentation for skin lesion using self-attention based progressive generative adversarial network," *Expert Systems With Applications*, vol. 165, p. 113922, Sep. 2020. doi: 10.1016/j.eswa.2020.113922.
47. R. Kaur, H. Gholam Hosseini, R. Sinha, "Synthetic Images Generation Using Conditional Generative Adversarial Network for Skin Cancer Classification," *TENCON 2021 - 2021 IEEE Region 10 Conference (TENCON)*, vol. 27, pp. 381–386, Dec. 2021. doi: 10.1109/tencon54134.2021.9707291.
48. J. Pope, M. Hassanuzzaman, M. Sherpa, O. Emara, A. Joshi, N. Adhikari, "Skin Cancer Machine Learning Model Tone Bias," *arXiv (Cornell University)*, 2024. doi: https://doi.org/10.48550/arxiv.2410.06385
49. E. Rezk, M. Eltorki, W. El-Dakhakhni, "Improving Skin Color Diversity in Cancer Detection: Deep Learning Approach," *JMIR Dermatology*, 5(3), e39143, 2022. doi: https://doi.org/10.2196/39143
50. B.N. Walker et al., "Skin cancer detection in diverse skin tones by machine learning combining audio and visual convolutional neural networks," *Oncology*, 103(5), pp. 113–120, 2024. doi: https://doi.org/10.1159/000541573
51. J. Han, M. Kamber, J. Pei, *Data Mining: Concepts and Techniques*; 3rd edition, 2011. Available: https://www.scholartext.com/book/88809627?_locale=fr
52. T. Majtner, B. Bajić, S. Yildirim, J.Y. Hardeberg, J. Lindblad, N. Sladoje, "Ensemble of Convolutional Neural Networks for Dermoscopic Images Classification," *arXiv (Cornell University)*, 2018. doi: https://doi.org/10.48550/arxiv.1808.05071
53. S.M. Pizer et al., "Adaptive histogram equalization and its variations," *Computer Vision, Graphics, and Image Processing*, 39(3), pp. 355–368, 1987.
54. K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics gems IV*. Academic Press Professional, Inc., 1994, pp. 474–485.
55. C.M. Bishop, *Pattern recognition and machine learning*. Springer, 2006.
56. V.D. Midasala, B. Prabhakar, J.K. Chaitanya, K. Sirnivas, D. Eshwar, P.M. Kumar, "MFEUsLNet: Skin cancer detection and classification using integrated AI with multilevel feature extraction-based unsupervised learning," *Engineering Science and Technology an International Journal*, 51(10), 101632, 2024. doi: https://doi.org/ 10.1016/j.jestch.2024.101632
57. T.K. Lee et al., "Dullrazor®: A software approach to hair removal from images," *Computers in Biology and Medicine*, 27(6), pp. 533–543, 1997.
58. R. Maurya et al., "Skin cancer detection through attention guided dual autoencoder approach with extreme learning machine," *Scientific Reports*, vol. 14, no. 17785, Aug. 2024. doi: 10.1038/s41598-024-68749-1.

59. A. Kirillov et al., "Segment Anything," *arXiv (Cornell University)*, Apr. 2023. doi: https://doi.org/10.48550/arxiv.2304.02643
60. A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," *arXiv.org*, Jun. 03, 2021. Available: https://www.arxiv.org/abs/2010.11929v2
61. S. Targ, D. Almeida, K. Lyman, "Resnet in Resnet: Generalizing Residual Architectures," *arXiv (Cornell University)*, Jan. 2016ю doi: 10.48550/arxiv.1603.08029.
62. X. Han, Y. Zhong, L. Cao, L. Zhang, "Pre-Trained AlexNet Architecture with Pyramid Pooling and Supervision for High Spatial Resolution Remote Sensing Image Scene Classification," *Remote Sensing*, vol. 9, no. 8, p. 848, Aug. 2017. doi: 10.3390/rs9080848.
63. Q. Guan et al., "Deep convolutional neural network VGG-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: a pilot study," *Journal of Cancer*, vol. 10, no. 20, pp. 4876–4882, Jan. 2019. doi: 10.7150/jca.28769.
64. P. Carcagnì et al., "Classification of Skin Lesions by Combining Multilevel Learnings in a DenseNet Architecture," in *Lecture notes in computer science*, pp. 335–344, 2019. doi: 10.1007/978-3-030-30642-7_30.
65. D. Sinha, M. El-Sharkawy, "Thin MobileNet: An Enhanced MobileNet Architecture," *IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, pp. 0280–0285, Oct. 2019. doi: 10.1109/uemcon47517. 2019.8993089.
66. L.N. Smith, "Cyclical learning rates for training neural networks," in *2017 IEEE Winter Conference on Applications of Computer Visio*n, pp. 464–472.
67. I. Loshchilov, F. Hutter, "SGDR: Stochastic gradient descent with warm restarts," *arXiv preprint*, 2016. doi: https://doi.org/10.48550/arXiv.1608.03983
68. D.P. Kingma, J. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations (ICLR)*, 2015.
69. R. Ali, A. Manikandan, R. Lei, J. Xu, "A novel SpaSA based hyper-parameter optimized FCEDN with adaptive CNN classification for skin cancer detection," *Scientific Reports*, vol. 14, no. 1, Apr. 2024. doi: 10.1038/s41598-024-57393-4.
70. Duyen N.T. Le, Hieu X. Le, Lua T. Ngo, Hoan T. Ngo, "Transfer learning with class-weighted and focal loss function for automatic skin cancer classification," *arXiv (Cornell University)*, Jan. 2020. doi: https://doi.org/10.48550/arxiv.2009.05977
71. S. Ghosh, S. Dhar, Raktim Yoddha, S. Kumar, Abhinav Kumar Thakur, Nanda Dulal Jana, "Melanoma Skin Cancer Detection Using Ensemble of Machine Learning Models Considering Deep Feature Embeddings," *Procedia Computer Science*, vol. 235, pp. 3007–3015, Jan. 2024. doi: https://doi.org/10.1016/j.procs.2024.04.284
72. J. Atwood, D. Towsley, "Diffusion-Convolutional Neural Networks," *arXiv.org*, Jul. 08, 2016. Available: https://arxiv.org/abs/1511.02136
73. S. Sabour, N. Frosst, G.E. Hinton, "Dynamic Routing Between Capsules," *arXiv (Cornell University)*, Jan. 2017. doi: https://doi.org/10.48550/arxiv.1710.09829
74. "SIIM-ISIC Melanoma Classification," *Kaggle*, 2024. Accessed on: Dec. 28, 2024. Available: https://www.kaggle.com/competitions/siim-isic-melanoma-classification/ discussion/181830
75. K. Chaitanya, E. Erdil, N. Karani, E. Konukoglu "Contrastive learning of global and local features for medical image segmentation with limited annotations," in *Advances in Neural Information Processing Systems*, 2020.
76. T. Chen, S. Kornblith, M. Norouzi, G. Hinton, "A simple framework for contrastive learning of visual representations," in *International Conference on Machine Learning*, 2020.
77. S. Azizi et al., "Big self-supervised models advance medical image classification," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021.
78. H. Haggerty, R. Chandra, "Self-supervised learning for skin cancer diagnosis with limited training data," *arXiv (Cornell University)*, Jan. 2024. doi: https://doi.org/ 10.48550/arxiv.2401.00692
79. P. Refaeilzadeh, L. Tang, H. Liu, "Cross-validation," in *Encyclopedia of Database Systems*, pp. 532–538. Springer, 2009.
80. T.R. Mahesh, V.V. Kumar, D.V. Kumar, O. Geman, M. Margala, M. Guduri, "The stratified K-folds cross-validation and class-balancing methods with high-performance ensemble classifiers for breast cancer classification," *Healthcare Analytics*, vol. 4, p. 100247, Sep. 2023. doi: 10.1016/j.health.2023.100247.

81. T.J. Brinker et al., "Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task," *European Journal of Cancer*, vol. 113, pp. 47–54, 2019.
82. S.S. Han et al., "Deep neural networks show an equivalent and often superior performance to dermatologists in onychomycosis diagnosis: Automatic construction of onychomycosis datasets by region-based convolutional deep neural network," *PLOS One*, vol. 13, no. 1, p. e0191493, Jan. 2018. doi: https://doi.org/10.1371/ journal.pone.0191493
83. S. Moldovanu, F.A. Damian Michis, K.C. Biswas, A. Culea-Florescu, L. Moraru, "Skin Lesion Classification Based on Surface Fractal Dimensions and Statistical Color Cluster Features Using an Ensemble of Machine Learning Techniques," *Cancers*, vol. 13, no. 21, p. 5256, Oct. 2021. doi: https://doi.org/10.3390/cancers13215256
84. V. Nikitin, V. Danylov, "Integration of Fractal Dimension in Vision Transformer for Skin Cancer Classification," *Electronics and Control Systems*, vol. 2, no. 76, pp. 15–20, Jun. 2023. doi: https://doi.org/10.18372/1990-5548.76.17662
85. R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization," *International Journal of Computer Vision*, vol. 128, no. 2, pp. 336–359, Feb. 2020. doi: https://doi.org/10.1007/s11263-019-01228-7
86. A. Binder, G. Montavon, S. Bach, K.-R. Müller, and W. Samek, "Layer-wise Relevance Propagation for Neural Networks with Local Renormalization Layers," *arXiv (Cornell University)*, Apr. 2016. doi: https://doi.org/10.48550/arxiv.1604.00825
87. S. Lundberg, S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," *arXiv.org*, Nov. 24, 2017. Available: https://arxiv.org/abs/1705.07874v2

Received 09.01.2025

INFORMATION ON THE ARTICLE

Vladyslav O. Nikitin, ORCID: 0009-0001-9921-0213, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e mail: nvo63911@gmail.com

Valery Ya. Danilov, ORCID: 0000-0003-3389-3661, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e mail: danilov1950@ukr.net

ПОДОЛАННЯ ВИКЛИКІВ У ГЛИБОКОМУ НАВЧАННІ ДЛЯ ВИЯВЛЕННЯ РАКУ ШКІРИ / В.О. Нікітін, В.Я. Данилов

Анотація. Рак шкіри є одним із найпоширеніших злоякісних новоутворень у світі. Раннє виявлення є критичним фактором зниження рівня смертності. Це підкреслює необхідність доступних систем комп'ютерної діагностики. Нещодавні досягнення в глибокому навчанні показали великі перспективи у вирішенні цієї проблеми. Незважаючи на цей прогрес у галузі машинного навчання, дослідники стикаються із численними перешкодами, коли йдеться про класифікацію раку шкіри. Розглянуто сучасний стан діагностики раку шкіри на основі глибокого навчання, критичні аспекти розроблення системи, включно з попереднім обробленням даних, навчанням моделей та оцінкою продуктивності. Крім того, висвітлюються можливості для інновацій, які можуть значно просунути цю галузь. Надаючи вичерпний огляд, стаття має на меті допомогти дослідникам та практикам в оптимізації моделей глибокого навчання, усуненні існуючих обмежень та дослідженні нових тенденцій для підвищення точності та доступності діагностики.

Ключові слова: рак шкіри, глибоке навчання, класифікація, трансформатори, CNN, GAN, попереднє оброблення даних, доповнення даних.