Deep Learning Models for Early Detection of Skin Cancer Using Images

Author: Nurbol Agybetov, Yermek Khaknazar, Baglan Yessenkeldi, Tokhtar Nuralin

Instructor: Seitov Altynbek

Part 1: Introduction and Methodology

1.1. Importance of the Topic

Skin cancer is among the most common forms of cancer worldwide. Early diagnosis, particularly of melanoma, is critical for successful treatment outcomes and survival in patients. Traditional methods of diagnosis, based on visual observation of the dermatologist, contain an element of subjectivity and depend on the professional skills of the clinician. Recent developments in technologies of Deep Learning (DL) have demonstrated remarkable potential in the analysis of medical images, achieving outcomes superior to or comparable to human observers. The development of automated computer algorithms based on DL of dermoscopic images may increase the accuracy, speed, and availability of skin cancer diagnosis, thus validating the topic of research presented here in current medical and healthcare practice.

1.2. Scope and Boundaries of the Review

This review of the literature includes academic work that addresses the application of deep learning algorithms in skin cancer detection and skin cancer classification because of the employment of digital images.

Integrated Elements:

- Time period: We primarily use articles between the time period 2016-2025 in an attempt to include the most influential work since the beginning of DL in this field.
- Methods: All of these Convolutional Neural Network (CNN) models are investigated, such as Transfer Learning-based (e.g., VGG, ResNet, Inception, DenseNet), ensemble, hybrid, and mobile application quick models.
- Extensive focus is placed on investigations with publically available dermoscopic datasets, such as the ISIC Archive, HAM10000, and PH2.
- The tasks in the binary and multi-class skin lesion classification are investigated.

Excluded Factors:

- It does not include citation of work performed exclusively on classical machine learning methods without involving deep learning.
- The review does not include papers addressing other medical image modalities (e.g., ultrasounds or confocal microscopies) aside from normal and dermoscopic photography.
- Model implementation details and computer codes are not included in this review.

1.3. Report Structure

Thematic structure was selected for structuring this report rather than structuring it chronologically, as it would facilitate more in-depth analysis and synthesis of principal research directions, would facilitate intercomparison of varied methodologies for addressing the same challenges (e.g., imbalance of data or interpretability of models), and would facilitate identification of general tendencies of the development of the field. Three principal parts constitute the report:

- **Introduction and Methodology:** Explains the significance of the subject, the extent of the review, and gives a table of sources summarized.

- **Thematic Literature Review:** Critically examines and synthesizes the literature on major themes: architectural solutions, data strategies, increasing clinical value, and comparing model performance with doctors.|

- **Conclusion:** It outlines the key findings, highlights significant research gaps, and rationalizes the recommendation for a future work.

1.4. Conceptual Map and Search Process

To present the makeup of the research field, a conceptual map was drawn. It indicates the major theme and how it relates to major sub-topics, which are model architectures, data processing techniques, integration with clinics, and performance measures.

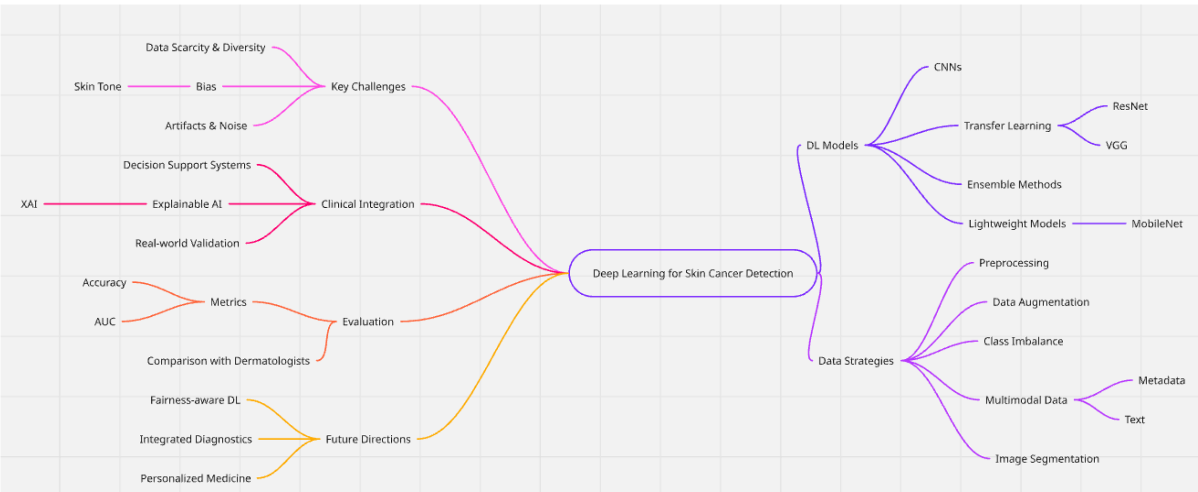

Figure 1. Conceptual map of the research area.

Search and choice of literature were carried out along a defined methodology, which is described in the following flowchart. This included keyword searching on high-ranking scientific databases, relevance and inclusion filters, and critical analysis of selected articles.

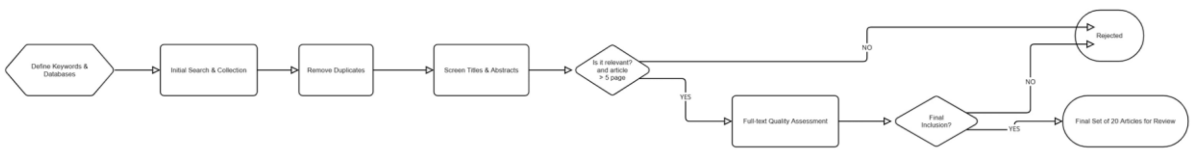

Figure 2. Flowchart of the literature search and selection process.

1.5. Table of Analyzed Academic Sources

Table 1. Summary Analysis of Key Academic Sources

Bibliography (APA Style)	**Method/Design**	**Data/Context**	**Key Findings**	**Limitations**	**Relevance to Research**
Rahi, M. M. I., et al. (2019). Detection of skin cancer using deep neural networks. *ICASERT*.	Comparative analysis: Custom CNN vs. VGG11, RESNET50, DENSENET 121.	HAM10000 (from ISIC).	ResNet50 and DenseNet 121 achieved 90% accuracy.	Only accuracy is reported.	Demonstrates the superiority of transfer learning.
Mijwil, M. M. (2021). Skin cancer disease images classification using deep learning solutions. *Multimedia Tools and Applications*.	Comparative analysis: Inception V3, ResNet, VGG19 for binary classification.	24,225 images from ISIC (2019-2020). Resampling for balancing.	Inception V3 was the best (86.90% accuracy).	Only binary classification.	Confirms the effectiveness of Inception V3 and the importance of data balancing.
Abunadi, I., & Senan, E. M. (2021). Deep Learning and	Hybrid approach: 1) Classical ML (ANN, FFNN) on handcrafted features. 2)	ISIC 2018 and PH2.	FFNN on hybrid features outperformed DL models (accuracy	CNN models might have been undertuned.	Shows the competitiveness of hybrid approaches.

Machine Learning Techniques... *Electronics*.	DL (ResNet-50, AlexNet).		up to 97.91%).		
Ahmadi Mehr, R., & Ameri, A. (2022). Skin Cancer Detection Based on Deep Learning. *JBPE*.	Model based on Inception-ResNet-v2 using patient metadata.	57,536 images from ISIC.	Including metadata increased accuracy by >5% (up to 94.5%).	Less effective on non-dermoscopic photos.	An innovative approach integrating clinical data.
Layode, O., et al. (2019). Deep Learning Based Integrated... System. *IEEE Big Data*.	Integrated system (DSS) with classification and retrieval (CBIR).	Dermofit (1300 images, 10 classes). Features from VGG-19, ResNet-50, Inception.	Best accuracy (85%) with an ensemble of fused features.	85% accuracy leaves room for improvement.	Presents a hybrid "doctor's assistant" system.
Jinnai, S., et al. (2020). The Development of a Skin Cancer Classification System... *Biomolecules*.	Faster R-CNN (FRCNN) model on clinical photos. Comparison with 20 dermatologists.	5846 clinical images (6 classes) from a single hospital.	FRCNN outperformed dermatologists (86.2% vs 79.5% accuracy).	Data from a single institution.	Proves that DL models can outperform doctors even on regular photos.

Dildar, M., et al. (2021). Skin Cancer Detection: A Review... *IJERPH*.	Systematic literature review (51 articles) on DL techniques.	Review of articles from 2011-2021.	CNN is the most effective approach. Data imbalance issues identified.	Review article.	Provides a structured overview of the field and key challenges.
Naqvi, M., et al. (2023). Skin Cancer Detection Using Deep Learning—A Review. *Diagnostics*.	Literature review with a focus on articles from 2021-2022.	Review of articles mainly using ISIC and HAM10000.	Confirms the dominance of CNNs. Highlights issues of artifacts and data bias.	Review article.	Updates the state of the field, formulates the problem of data bias.
Akinrinade, O., & Du, C. (2025). Skin cancer detection using deep machine learning techniques. *Intelligence-Based Medicine*.	Review with a focus on class imbalance and small dataset problems.	Review of existing methods.	Transfer learning, augmentation, and GANs are key strategies for data scarcity.	Review article.	Focuses on practical solutions for common problems.
Ameri, A. (2020). A Deep Learning	Transfer learning using a pre-trained	HAM10000 (3400 images, balanced).	84% accuracy, 0.91 AUC.	A small subset of data was used.	Example of applying transfer learning for

Approach to Skin Cancer Detection.. *JBPE*.	AlexNet model.				binary classification.
Ghosh, H., et al. (2024). A-Study-on-the-Application-of-Machine-Learning...	Hybrid model combining VGG16 and ResNet50. Use of class weights.	3000 images, 9 classes.	The hybrid model outperforms individual ones (validation accuracy 97.50%).	Data source not specified.	Demonstrates the effectiveness of hybrid models.
Mahmud, M. A. A., et al. (2025). Explainable deep learning approaches... *Scientific Reports*.	A novel model based on Xception with advanced layers. Includes XAI techniques.	Two combined datasets. Extensive preprocessing.	High accuracy (up to 96.48%). XAI visualizes the "thought process".	Lack of diversity in datasets is noted.	Addresses the "black box" problem using XAI.
Vijayalakshmi, M. M. (2019). Melanoma Skin Cancer Detection.. *IJTSRD*.	Comparative analysis: NN, SVM, and CNN. Emphasis on preprocessing.	ISIC (1000-1500 images).	Achieved 85% accuracy.	Relatively small dataset size.	Highlights the importance of preprocessing.
Gururaj, H. L., et al. (2023). DeepSkin: A Deep Learning Approach.	Comparative analysis of DenseNet169 and ResNet50 using sampling.	MNIST: HAM10000 (10015 images, 7 classes).	DenseNet169 with undersampling showed the best	Sampling techniques were compared on different models.	Demonstrates the impact of sampling techniques on performance.

.. *IEEE Access*.			accuracy (91.2%).		
Haenssle, H. A., et al. (2018). Man against machine... *Annals of Oncology*.	Inception v4 CNN. Performance comparison with 58 dermatologists.	100 images for the test set.	The CNN outperformed most dermatologists (specificity 82.5% vs 75.7% at the same sensitivity).	Test set did not include all lesion types.	A key paper proving DL's superiority over a large group of experts.
Imran, A., et al. (2022). Skin Cancer Detection Using Combined Decision of Deep Learners. *IEEE Access*.	An ensemble of three DL models (VGG, CapsNet, ResNet) using majority voting.	ISIC (5800 images, binary classification).	The ensemble (93.5% accuracy) significantly outperformed individual models (69-79%).	Only binary classification.	Convincingly proves the effectiveness of ensemble approaches.
Wei, L., et al. (2020). Automatic Skin Cancer Detection.. based on Ensemble Lightweight Deep Learning Network. *IEEE Access*.	Ensemble of lightweight models (MobileNetV1, DenseNet-121) with a discriminator network.	ISBI 2016 challenge.	The lightweight model showed state-of-the-art results in the competition.	The model has a complex architecture.	Focuses on creating efficient models for mobile devices.

Yu, L., et al. (2017). Automated Melanoma Recognition... via Very Deep Residual Networks. *IEEE TMI*.	Two-stage framework: 1) Segmentation with FCRN (>50 layers), 2) Classification of the segmented area.	ISBI 2016 challenge.	Won 1st place in the classification competition. Proved that segmentation before classification improves results.	Very deep networks (>100 layers) showed worse results on limited data.	A fundamental paper proving the effectiveness of very deep networks and the "segmentation-first" approach.
Kalouche, S. (2016). Vision-Based Classification of Skin Cancer using Deep Learning.	Comparative analysis: Logistic Regression, custom NN, and fine-tuned VGG-16.	ISIC (1280 images).	Fine-tuned VGG-16 (78% accuracy) significantly outperformed other models (NN - 56%, LR - ~50%).	Small and imbalanced dataset.	An early example proving the effectiveness of transfer learning.
Bajwa, M. N., et al. (2020). Computer-Aided Diagnosis of Skin Diseases using Deep Neural Networks. *Applied Sciences*.	Ensemble of 4 models (ResNet, DenseNet, etc.). Use of disease taxonomy to improve classification.	DermNet (23 classes) and ISIC Archive (7 classes).	SOTA on DermNet (80% on 23 classes) and 93% on ISIC. Using disease hierarchy improved results.	Patient metadata was not used.	Demonstrates the scalability of DL to a large number of classes and introduces the idea of using taxonomy.

Part 2: Thematic Literature Analysis

2.1. Architectural Solutions: From Transfer Learning to Ensembles and Hybrid Systems

20 given sources analysis reveals a clear trajectory in the development of architectural strategies. Transfer learning is a well-established standard, yielding strong baseline accuracy. Kalouche (2016) and Rahi et al. (2019) studies prominently demonstrate that pre-trained networks (VGG, ResNet, DenseNet) significantly outperform scratch-trained models or conventional machine learning approaches [12, 17].

However, to achieve state-of-the-art performance, researchers are pushing further. One of the most well-known trends is the use of model ensembles. Instead of a single, albeit powerful, network, research such as Imran et al. (2022), Bajwa et al. (2020), and Ghosh et al. (2024) employs predictions from several different architectures (e.g., VGG, ResNet, CapsNet) [5, 7, 10]. This compensates for the shortcomings of individual models and improves the system's overall accuracy and credibility, as borne out by the results—the ensemble in Imran et al. (2022) was 93.5% accurate, an improvement of 14% on the best-performing individual model [10]. Further, research by Wei et al. (2020) indicates that the ensemble approach is also effective for light networks (MobileNet, DenseNet) and this is crucial in developing high-performing mobile applications [19].

2.2. Data Strategies: From Preprocessing to Taxonomy Utilization

It is a common consensus across all research that representation and quality of data are as important as the model architecture itself. Image preprocessing to remove artefacts such as hair, glare, and air bubbles is an indispensable step with special attention devoted to it in Vijayalakshmi (2019) and Mahmud et al. (2025) [14, 18].

Class imbalance is another universal problem. There are three main techniques authors employ to counter it: 1) Resampling (under- or oversampling), e.g., Gururaj et al. (2023) and Mijwil (2021) [8, 15]; 2) Class weights, e.g., Ghosh et al. (2024) [7]; and 3) Data augmentation, which also helps to avoid overfitting on small datasets [3, 19].

The most advanced techniques make use of data in non-trivial ways. Yu et al. (2017) illustrate the effectiveness of a two-stage approach: accurate segmentation of the lesion region by an FCRN network, and then classification of that region only [20]. This approach helped them to achieve first place in the ISBI 2016 challenge. Bajwa et al. (2020) propose one further intriguing idea: the use of disease taxonomy. They have developed the model by training it on 622 highly specialized subclasses to categorize 23 generalized classes more accurately, which improved the result [5].

2.3. Enhancing Clinical Value: Metadata and Explainable AI (XAI)

As a quest to render laboratory models close to real clinical practice, scientists have begun adding additional information and making the models open. Ahmadi Mehr & Ameri (2022) illustrated that adding patient metadata (age, sex, location) to the image before presenting it to the network enhances accuracy by more than 5% [2]. This is in line with what a physician does, where he always considers the background of the patient.

The "black box" problem is addressed using Explainable AI (XAI). Layode et al. (2019) created a system that, in addition to the diagnosis, visually displays the doctor with comparable cases from a database to enable an educated decision [13]. Mahmud et al. (2025) use Grad-CAM and Saliency Maps directly to image the areas of the image which the model used [14]. These are crucial in establishing physicians' trust in the technology.

2.4. Man vs. Machine Comparison

The final test of effectiveness is direct comparison with clinicians. Two of the seminal papers in the provided set are dedicated to this issue. Jinnai et al. (2020) compared their FRCNN model with 20 dermatologists on conventional clinical photographs and showed the machine's superiority [11]. Haenssle et al. (2018) went even larger in scope, comparing their Inception v4 CNN with 58 dermatologists from 17 nations. Their conclusion is straightforward: the CNN outperformed most of the experts with considerably higher specificity (i.e., fewer false-positive diagnoses) at the same sensitivity [9]. These results are definitive proof that DL systems will be an important tool in physicians' hands.

Part 3: Conclusion and Future Directions

3.1. Main Findings from the Literature

Following analysis of 20 scientific sources, one can conclude that deep learning has emerged as an efficient and proven method for the diagnosis of skin cancer. Modern research indicates that optimal performance is obtained not with single models, but with ensembles of different architectures [5, 7, 10]. The success of any model is entirely dependent on strict preprocessing of data to remove artifacts and the application of methods against class imbalance [7, 8, 15]. The latest methods enhance the clinical applicability of models by merging multimodal data (e.g., patient metadata or disease taxonomy) [2, 5] and applying Explainable AI (XAI) technology to enhance transparency and credibility [14]. Massive-scale studies have solidly confirmed that newer DL models perform better than large groups of skilled dermatologists in accuracy [9, 11].

3.2. Key Research Gaps

While much progress has been achieved, the following gaps remain relevant:

- **Fairness and Data Diversity:** It is the most debated and critical gap. Most public datasets consist of predominantly light-skinned patients' images, which limits the models' applicability and fairness to other races [3, 6, 16].

- **Multimodal Data Integration:** Although separate papers have proven the vast potential of metadata [2] and taxonomy [5] work, these are not yet standard procedures. Most models remain "purely visual," and most useful clinical information remains untapped
- **Clinical Validation and Implementation:** Almost all studies are conducted on retrospective curated data. There is a glaring shortage of prospective studies that would measure the performance of the models in real-time in actual clinical practice and their impact on doctors' workloads.
- **Explainability in Practice:** Although XAI technologies already exist [14], there is minimal research examining how physicians interact with such explanations, if indeed they do enhance diagnostic accuracy and confidence, or perhaps engender misleadings.

3.3. Justification for the Research Idea

These gaps, particularly in multimodal data fusion and model fairness, are the pillars of our study. Existing systems either train only on the image or do so on skin type-imbalanced data, and therefore they have constrained real clinical utility and ethics.

Our suggestion is to develop a Fair Multimodal Diagnostic System (FMDS). The system will be based on a hybrid architecture that will analyze in parallel: 1) the dermoscopic image (using EfficientNetV2), 2) structured clinical information (with a mandatory "Fitzpatrick skin type" field), and 3) textual information of the patient history (using a specialized language model like BioBERT).

The principal distinction of our solution will be that it will be heavily focused on fairness. We will actively use stratified sampling and loss weighting techniques to ensure the model trains as well on all skin types. An integrated XAI module will generate a combined explanation, showing not only relevant regions within the image (Grad-CAM) but also keywords from the patient's history and most contributing clinical features (SHAP values) that contributed towards the diagnosis. This will allow the creation of a more accurate, fair, and clinically relevant system, ready for the next step—prospective validation.

References

1. Abunadi, I., & Senan, E. M. (2021). Deep Learning and Machine Learning Techniques of Diagnosis Dermoscopy Images for Early Detection of Skin Diseases. *Electronics, 10*(24), 3158. https://doi.org/10.3390/electronics10243158
2. Ahmadi Mehr, R., & Ameri, A. (2022). Skin Cancer Detection Based on Deep Learning. *Journal of Biomedical Physics & Engineering, 12*(6), 559-568. https://doi.org/10.31661/jbpe.v0i0.2207-1517
3. Akinrinade, O., & Du, C. (2025). Skin cancer detection using deep machine learning techniques. *Intelligence-Based Medicine, 11*, 100191. https://doi.org/10.1016/j.ibmed.2024.100191

4. Ameri, A. (2020). A Deep Learning Approach to Skin Cancer Detection in Dermoscopy Images. *Journal of Biomedical Physics & Engineering, 10*(6), 801-806. https://doi.org/10.31661/jbpe.v10i6.2004-1107
5. Bajwa, M. N., Muta, K., Malik, M. I., Siddiqui, S. A., Braun, S. A., Homey, B., Dengel, A., & Ahmed, S. (2020). Computer-Aided Diagnosis of Skin Diseases using Deep Neural Networks. *Applied Sciences, 10*(7), 2488. https://doi.org/10.3390/app10072488
6. Dildar, M., Akram, S., Irfan, M., Khan, H. U., Ramzan, M., Mahmood, A. R., Alsaiari, S. A., Saeed, A. H. M., Alraddadi, M. O., & Mahnashi, M. H. (2021). Skin Cancer Detection: A Review Using Deep Learning Techniques. *International Journal of Environmental Research and Public Health, 18*(10), 5479. https://doi.org/10.3390/ijerph18105479
7. Ghosh, H., Rahat, I. S., Mohanty, S. N., Ravindra, J. V. R., & Sobur, A. (2024). A Study on the Application of Machine Learning and Deep Learning Techniques for Skin Cancer Detection. *International Journal of Computer and Systems Engineering, 18*(1).
8. Gururaj, H. L., Manju, N., Nagarjun, A., Manjunath Aradhya, V. N., & Flammini, F. (2023). DeepSkin: A Deep Learning Approach for Skin Cancer Classification. *IEEE Access, 11*, 50205-50215. https://doi.org/10.1109/ACCESS.2023.3274848
9. Haenssle, H. A., Fink, C., Schneiderbauer, R., Toberer, F., Buhl, T., Blum, A., Kalloo, A., Ben Hadj Hassen, A., Thomas, L., Enk, A., & Uhlmann, L. (2018). Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Annals of Oncology, 29*(8), 1836–1842. https://doi.org/10.1093/annonc/mdy166
10. Imran, A., Nasir, A., Bilal, M., Sun, G., Alzahrani, A., & Almuhaimeed, A. (2022). Skin Cancer Detection Using Combined Decision of Deep Learners. *IEEE Access, 10*, 118198-118214. https://doi.org/10.1109/ACCESS.2022.3220329
11. Jinnai, S., Yamazaki, N., Hirano, Y., Sugawara, Y., Ohe, Y., & Hamamoto, R. (2020). The Development of a Skin Cancer Classification System for Pigmented Skin Lesions Using Deep Learning. *Biomolecules, 10*(8), 1123. https://doi.org/10.3390/biom10081123
12. Kalouche, S. (2016). *Vision-Based Classification of Skin Cancer using Deep Learning*. Stanford University.
13. Layode, O., Alam, T., & Rahman, M. M. (2019). Deep Learning Based Integrated Classification and Image Retrieval System for Early Skin Cancer Detection. *2019 IEEE International Conference on Big Data (Big Data)*, 5521-5526. https://doi.org/10.1109/BigData47090.2019.9006206
14. Mahmud, M. A. A., Afrin, S., Mridha, M. F., Alfarhood, S., Che, D., & Safran, M. (2025). Explainable deep learning approaches for high precision early melanoma detection using dermoscopic images. *Scientific Reports, 15*, 24533. https://doi.org/10.1038/s41598-025-09938-4
15. Mijwil, M. M. (2021). Skin cancer disease images classification using deep learning solutions. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-021-10952-7

16. Naqvi, M., Gilani, S. Q., Syed, T., Marques, O., & Kim, H.-C. (2023). Skin Cancer Detection Using Deep Learning—A Review. *Diagnostics, 13*(11), 1911. https://doi.org/10.3390/diagnostics13111911
17. Rahi, M. M. I., Ullah, A. K. M. A., Khan, F. T., Alam, M. G. R., Mahtab, M. T., & Alam, M. A. (2019). Detection of skin cancer using deep neural networks. *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*. https://doi.org/10.1109/ICASERT.2019.8934512
18. Vijayalakshmi, M. M. (2019). Melanoma Skin Cancer Detection using Image Processing and Machine Learning. *International Journal of Trend in Scientific Research and Development, 3*(4), 780-784.
19. Wei, L., Ding, K., & Hu, H. (2020). Automatic Skin Cancer Detection in Dermoscopy Images based on Ensemble Lightweight Deep Learning Network. *IEEE Access*. https://doi.org/10.1109/ACCESS.2020.2997710
20. Yu, L., Chen, H., Dou, Q., Qin, J., & Heng, P. A. (2017). Automated Melanoma Recognition in Dermoscopy Images via Very Deep Residual Networks. *IEEE Transactions on Medical Imaging, 36*(4), 994-1004. https://doi.org/10.1109/TMI.2016.2642839