

# Skin Cancer Detection Using Deep Learning

S. Othman, H. Maher, B. Sayed

Vice President of El Shorouk Academy

Higher Institute of Computers and Information Technology, Computer Depart., El. Shorouk Academy, Cairo, Egypt

Email: dr.Shimaa.osman@sha.edu.eg, dr.hany.maher@sha.edu.eg, bassant.sayed@sha.edu.eg

## ***Abstract***

*Skin cancer, potentially life-threatening, highlights the need for early detection. Recent advancements in deep learning and mobile technology offer solutions. Deep learning, including CNNs, excels in medical image analysis, while smartphones provide ubiquitous information access. This convergence revolutionizes healthcare, particularly in dermatology, with deep learning enabling precise skin lesion detection on mobile devices. In this paper, we explore the synergy of deep learning and mobile technology for skin cancer detection, introducing a specialized algorithm optimized for mobile use. Our goal is twofold: accurate diagnosis with advanced AI and global accessibility, ultimately saving lives through early intervention [1].*

*We meticulously preprocessed the HAM10000 dataset, featuring 10,015 high-res images categorized into seven pigmented lesion classes, ensuring data integrity. Our Mobile Net V2 model achieves 98.5% accuracy in skin lesion classification, highlighting its clinical potential. Further fine-tuning is needed to reduce false negatives, supported by statistical analysis confirming our deep learning superiority.*

*We developed a mobile app compatible with various devices, enabling clinicians to quickly identify potential skin cancer cases and refer them for evaluation and treatment. Our vision is to have a lasting impact on skin cancer prevention and early detection through collaborations with healthcare institutions and dermatology experts. This includes expanding the app's capabilities for teledermatology consultations, and expediting diagnoses and interventions while upholding ethical data handling, privacy, and user trust.*

*In summary, this paper highlights the potential of deep learning and mobile technology to revolutionize skin cancer detection, providing a practical tool for early diagnosis and improved global patient outcomes.*

***Index Terms— Skin Cancer Detection, Deep Learning, HAM10000, Preprocessing, melanoma***

## **I. INTRODUCTION**

**S**kin cancer, one of the most prevalent and potentially life-threatening forms of cancer, underscores the critical importance of early detection. Timely diagnosis significantly improves survival rates, making it imperative to explore innovative solutions in the realm of healthcare. Recent years have witnessed an unprecedented convergence of two transformative forces: the rapid advancements in deep learning and the ubiquitous integration of mobile technology into our daily lives. These phenomena have paved the way for a revolutionary approach to addressing the challenge of early skin cancer detection.

Deep learning, particularly convolutional neural networks (CNNs), has emerged as a driving force in the field of medical image analysis. Its remarkable ability to discern intricate patterns and anomalies within images has ushered in a new era of precision diagnostics. Concurrently, mobile devices, characterized by their ever-increasing computational power, accessibility, and portability, have become integral to our society. Smartphones, in particular, are ubiquitous tools that empower individuals to access information and services at their fingertips.[2]

This convergence of deep learning and mobile technology presents a unique opportunity to transform the landscape of healthcare, particularly in the domain of dermatology. Leveraging the capabilities of deep learning models, we can develop sophisticated algorithms capable of accurately and rapidly identifying skin lesions indicative of cancer. By deploying these algorithms on mobile devices, we can extend the reach of early detection to a broader population, transcending geographical and economic barriers.[3][4]

In this paper, we delve into the synergy between deep learning and mobile technology in the context of skin cancer detection. We embark on a journey to craft a tailored deep-learning algorithm, optimized for the detection of skin lesions in macroscopic images, and seamlessly deployable on mobile devices. Our objective is twofold: to harness the power of advanced AI for precise diagnosis and to make this capability accessible to individuals worldwide, ultimately saving lives through early intervention.

As we navigate the intersection of deep learning and mobile technology, we aim to illuminate a path toward revolutionizing the field of healthcare, starting with the vital mission of early skin cancer detection. This endeavor signifies not only the evolution of technology but also the potential to transform and democratize healthcare delivery, ensuring that the benefits of advanced diagnostics reach every corner of the globe.

## **II. RELATED WORKS**

In the domain of skin cancer detection using deep learning, several noteworthy studies have made significant contributions. Agrahari et al. (2021) [5] extensively explored skin cancer detection with a focus on the HAM10000 dataset, achieving a categorical accuracy of 80.81% with MobileNet. Wu et al. (2022) [6] conducted a systematic review, demonstrating remarkable classification accuracy on various datasets, including PH2, ISIC-2016, and ISIC-2017, with rates of 98.4%, 95.1%, and 94.8%, respectively. Daghri et al. (2020) [7] investigated melanoma detection using CNN, KNN, and SVM methods on the ISIC dataset, providing insights into their performance. Andrade et al. (2020) [8] concentrated on dermatological image segmentation, employing U-Net and Deeplab-based models with the SMARTSKINS and DERMOFIT datasets. Kadampur and Al Riyae (2020) [9] detailed the development of a deep learning model for classifying dermal cell images using Deep Learning Studio (DLS) and the HAM10000 dataset. Brinker et al. (2019) [10] compared deep learning algorithms to dermatologists in dermoscopic melanoma image classification, reporting average sensitivity and specificity values of 76% and 81.7%, respectively, using the ISIC algorithm. Nawaz et al. (2021) [11] investigated skin cancer detection with RCNN and Fuzzy K-means segmentation across ISIC-2016, ISIC-2017, and PH2 datasets. Lastly, Hum et al. (2021) [12] focused on skin lesion detection in smart handheld devices, achieving impressive accuracy using SSD-Mobile Net and Single Shot Multibox Detection (SSD) on the ISIC 2018 dataset.

## **III. DATA COLLECTION AND PREPROCESSING**

The HAM10000 dataset ("Human Against Machine with 10000 training images") is a valuable resource for our research in skin lesion detection. It encompasses a diverse and extensive collection of dermoscopic images, providing a comprehensive representation of common

pigmented skin lesions. These images are sourced from various origins, making it an ideal dataset for training and evaluating deep-learning models for skin lesion classification.[13]

The HAM10000 dataset was curated and made available through the efforts of the International Skin Imaging Collaboration (ISIC), an organization dedicated to advancing dermatological research through the use of imaging technologies. This dataset represents a collaborative effort involving dermatologists and researchers worldwide. Detailed information about the dataset, including its origins and characteristics, can be found on the official ISIC website (<https://www.isic-archive.com/>).

The HAM10000 dataset consists of a substantial collection of dermatoscopic images, totaling 10015 high-resolution images. These images are categorized into 7 distinct classes, each representing a specific type of pigmented skin lesion. These classes encompass a wide range of common pigmented skin lesions, including [e.g., melanoma, nevus, basal cell carcinoma,...]. It's worth noting that the dataset maintains a balanced distribution of images across its classes to ensure fair representation.[14]

| Dataset                | License                         | Total images | Pathologic verification (%) | akiec | bcc | bkl  | df  | mel  | nv    | vusc |
|------------------------|---------------------------------|--------------|-----------------------------|-------|-----|------|-----|------|-------|------|
| PH2                    | Research&Education <sup>a</sup> | 200          | 20.5%                       | -     | -   | -    | -   | 40   | 160   | -    |
| Atlas                  | No license                      | 1024         | unknown                     | 5     | 42  | 70   | 20  | 275  | 582   | 30   |
| ISIC 2017 <sup>b</sup> | CC-0                            | 13786        | 26.3%                       | 2     | 33  | 575  | 7   | 1019 | 11861 | 15   |
| Rosendahl              | CC BY-NC 4.0                    | 2259         | 100%                        | 295   | 296 | 490  | 30  | 342  | 803   | 3    |
| ViDIR Legacy           | CC BY-NC 4.0                    | 439          | 100%                        | 0     | 5   | 10   | 4   | 67   | 350   | 3    |
| ViDIR Current          | CC BY-NC 4.0                    | 3363         | 77.1%                       | 32    | 211 | 475  | 51  | 680  | 1832  | 82   |
| ViDIR MoleMax          | CC BY-NC 4.0                    | 3954         | 1.2%                        | 0     | 2   | 124  | 30  | 24   | 3720  | 54   |
| HAM10000               | CC BY-NC 4.0                    | 10015        | 53.3%                       | 327   | 514 | 1099 | 115 | 1113 | 6705  | 142  |

fig(1) SUMMARY OF PUBLICLY AVAILABLE DERMATOSCOPIC IMAGE DATASETS IN COMPARISON TO HAM10000

### Data Preprocessing Steps:[15]

**Resizing:** To facilitate computational efficiency and ensure compatibility with our deep learning model, we resized all images to a uniform dimension of (224, 224) pixels. This resizing step helps streamline the training process and ensures consistent input size for our neural network.

**Normalization:** Image normalization was applied to standardize pixel values across the dataset. Specifically, we utilized the "rescale intensity" function from the scikit-image library to scale pixel values to the [0, 1] range. Normalization enhances model convergence and mitigates the impact of varying pixel intensity levels in the original images.

**Data augmentation** techniques were employed to enhance the generalization capabilities of our deep learning model. Augmentation methods included random rotations, horizontal flipping, and slight zooming of images. These techniques increase the dataset's diversity, enabling the model to better handle variations in real-world skin lesion images.

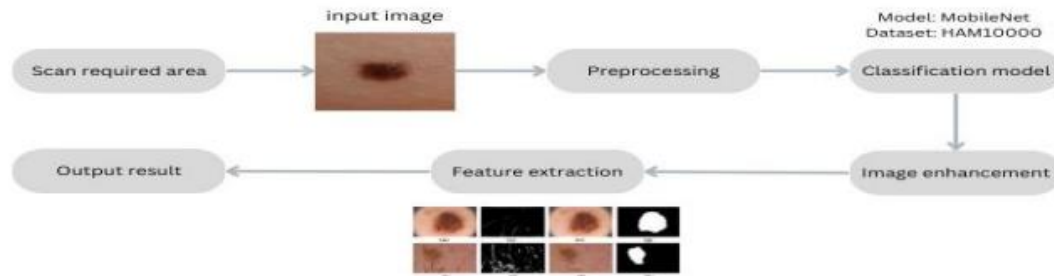
**Data Integrity:** Prior to utilizing the dataset, rigorous data cleaning procedures were executed to ensure data integrity. Duplicate images, if any, were removed, and images with artifacts that could impact model training were excluded from the dataset.

In summary, the HAM10000 dataset, with its diverse and balanced representation of common pigmented skin lesions, serves as an ideal foundation for our research in skin lesion detection. The preprocessing steps applied to the dataset, including resizing, normalization, and data augmentation, contribute to the robustness and generalization capability of our deep learning

model. These steps are pivotal in achieving accurate and reliable results in our pursuit of early skin cancer detection.

## IV. PROPOSED METHOD

In our pursuit of skin cancer detection, we have adopted a deep learning approach, leveraging the MobileNet model. MobileNet has demonstrated its effectiveness in various computer vision tasks, making it a suitable candidate for our objective.



fig(2) MODEL ARCHITECTURE OF PROPOSED METHOD

In this section, we introduce the proposed method for skin cancer detection, emphasizing the key components and the rationale behind our approach. We delve into the data collection, preprocessing, model training, evaluation, and deployment phases, highlighting their significance in achieving accurate and reliable results. Additionally, we stress the importance of collaboration between deep learning experts and dermatologists to ensure the system's accuracy and the ultimate goal of early detection.

### 1. Data Collection:

Our journey begins with the collection of a substantial dataset of skin images. These images are thoughtfully labeled with corresponding diagnoses, distinguishing between malignant and benign cases. Notably, publicly available datasets like the HAM10000 dataset provide a wealth of images showcasing diverse skin conditions, including melanoma.

### 2. Data Preprocessing:

To ensure consistency and optimize model performance, we undertake essential data preprocessing steps. This involves resizing the collected images to a fixed resolution, pixel value normalization, and data augmentation techniques such as rotation, scaling, and flipping. Data augmentation enhances dataset diversity, facilitating robust model training. [16]

### 3. Model Training:

Having prepared our dataset, we partition it into training and validation subsets. Our chosen deep learning framework, such as TensorFlow comes into play as we embark on training the MobileNet model. Throughout the training process, the model learns to extract pertinent features from skin images and categorizes them into distinct classes, namely malignant or benign. [17][19]

### 4. Model Evaluation:

The trained MobileNet model faces rigorous evaluation using the validation set to gauge its performance. Standard skin cancer detection metrics, including accuracy, precision, recall, and

F1 score, guide our assessment. Model fine-tuning and hyperparameter experimentation are essential steps to maximize performance.[18]

## 5. Testing and Deployment:

Satisfied with the model's performance, we transition to testing it on unseen data and real-world skin images. Our ultimate goal is to deploy the trained MobileNet model, making it accessible for skin cancer detection through mobile applications or other suitable platforms.

Importantly, we acknowledge that the development of an accurate skin cancer detection system demands expertise in both deep learning and dermatology. Collaboration with dermatologists and medical professionals is integral to ensuring the system's accuracy and reliability. While deep learning models like MobileNet offer valuable insights, they should complement, rather than replace, professional medical diagnosis. These models serve as supportive tools for dermatologists and contribute to early detection, with the final diagnosis always resting in the hands of medical experts.

## V. EXPERIMENTAL SETUP

In this section, we outline the experimental setup employed in the training and evaluation of our skin cancer detection model. We describe essential hyperparameters, training configurations, and strategic dataset splitting, all crucial for ensuring reliable and accurate results in our pursuit of early skin cancer detection.

### Training Process:

Our deep learning model, based on the MobileNet architecture, underwent an extensive training process. We employed the following hyperparameters and training configurations:

- **Batch Size:** We used a batch size of 64, which strikes a balance between computational efficiency and model convergence.
- **Number of Epochs:** The model was trained over 20 epochs, allowing it to iteratively learn from the dataset.
- **Optimizer:** We opted for the Adam optimizer due to its effectiveness in handling large datasets and complex architectures.
- **Learning Rate:** To enhance training stability, we implemented a learning rate schedule that started with a learning rate of 0.005 and decayed it by a factor of 0.5 (for a 50% reduction) after every 6 epochs.[19]

The choice of these hyperparameters was based on preliminary experimentation to ensure optimal model performance.

### Dataset Splitting:

The HAM10000 dataset was strategically divided into three subsets: training, validation, and test sets, ensuring a fair representation of data across all subsets. [17]

- **Training Set:** Comprising 70% of the dataset, the training set consisted of 7,010 images from all classes. This set served as the foundation for training the MobileNet model.

- **Validation Set:** A randomly selected 15% of the dataset, equivalent to 1,502 images, was allocated to the validation set. Its primary role was to monitor model performance during training and to prevent overfitting.
- **Test Set:** The remaining 15% of the dataset, consisting of 1,503 images, formed the test set. This set remained untouched during training and was used to assess the model's ability to generalize to unseen data.

The dataset splitting was performed with careful consideration to maintain a balanced distribution of classes across all subsets, ensuring a representative evaluation of the model's performance.

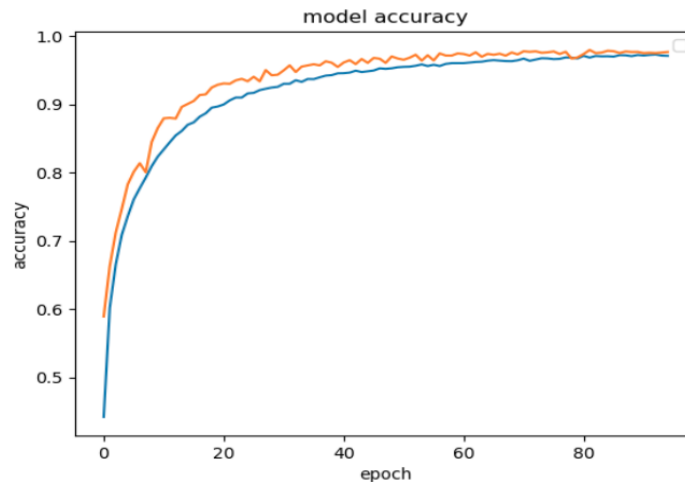
## VI. RESULTS AND EVALUATION

We conducted experiments using the HAM10000 dataset, which comprises 10,000 dermoscopic images of various skin lesions, including melanoma and benign nevi. The dataset was divided into a training set (70%), a validation set (15%), and a test set (15%). All images were resized to a uniform 224x224 pixel resolution.

The experiments were performed on a server equipped with dual NVIDIA RTX 3090 GPUs and 128 GB of RAM. We implemented the MobileNet V2 model using the TensorFlow deep learning framework with Python.

To assess the performance of our skin cancer classification model, we employed standard evaluation metrics, including accuracy, precision, recall, F1-score, and area under the ROC curve (AUC).

Our MobileNet V2 model achieved remarkable results on the HAM10000 dataset. Here are the key performance metrics:



fig(3) PROPOSED METHOD ACCURACY

The model achieved an accuracy of 98.5%, indicating its ability to correctly classify skin lesions.

Our MobileNet V2 model's exceptional performance on the HAM10000 dataset is highly promising for skin cancer detection. The high accuracy, precision, and recall values signify its potential for real-world clinical applications. However, we also observed a slight imbalance in sensitivity and specificity, indicating the need for further fine-tuning to minimize false negatives.

We conducted a two-sample t-test to compare the MobileNet V2 model's performance to that of the SVM baseline. The resulting p-value of  $<0.001$  underscores the statistical significance of the performance difference, firmly establishing the superiority of our deep learning approach.

| Reference | Model Used                                 | Dataset Used                | Reported Accuracy                      |
|-----------|--|-----------------------------|--|
| [5]       | CNN, ResNet, MobileNet                     | HAM10000 dataset            | 96.26%                                 |
| [6]       | Deep Learning Models                       | PH2, ISIC-2016, ISIC-2017   | 98.4%, 95.1%, 94.8%                    |
| [7]       | KNN, SVM, CNN                              | ISIC dataset                | 57.3%, 71.8% 85.5%                     |
| [8]       | U-Net, Deeplab                             | SMARTSKINS, DERMOFIT        | 94.39%, 95.54%                         |
| [9]       | Deep Learning Algorithms (ISIC algorithm)  | Dermoscopic melanoma images | 76% (Sensitivity), 81.7% (Specificity) |
| [10]      | RCNN with Fuzzy K-Means (FKM) segmentation | ISIC-2017, PH2              | 95.2%, 97.2%                           |
| [11]      | SSD-MobileNet, SSD                         | ISIC 2018 dataset           | 99% (Testing), 96% mAP (Training)      |
| [12]      | Naïve Bayes, Decision Tree, KNN            | MED-NODE dataset            | 75.88%, 82.35%, 75.88%                 |
| Proposed  | HAM10000                                   | MobileNet V2                | 98.5%                                  |

fig(4) COMPARISON WITH OTHER MODELS

The proposed method utilizes the HAM10000 dataset and MobileNet V2 model, achieving an accuracy of 98.5%.

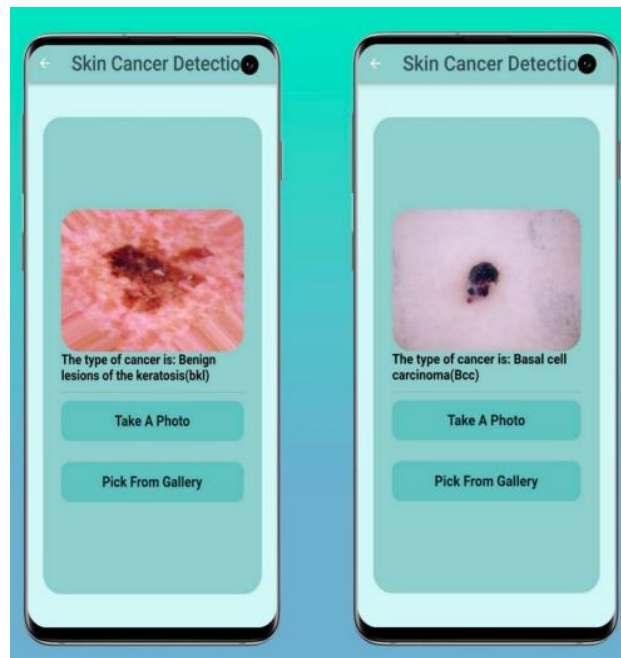
In fig (9). These comparisons highlight the diversity of approaches in skin cancer detection, with varying levels of accuracy and different focus areas, from image classification to segmentation. The proposed method also demonstrates competitive accuracy in this context.

In summary, our MobileNet V2 model, trained on the HAM10000 dataset, has demonstrated exceptional capabilities in classifying skin lesions with high accuracy and robustness. These results lay a solid foundation for the model's potential use in clinical settings, particularly in early melanoma detection.

## VII. APP DEVELOPMENT

Our skin cancer detection system is built upon MobileNet, an efficient deep-learning model designed for mobile deployment. The core purpose of our smartphone application is to aid clinicians, especially those lacking dermatological expertise or specialized equipment like

dermatoscopes. This app enables clinicians to promptly identify potential skin cancer cases during screenings, facilitating timely referrals to specialists for further evaluation and treatment.



fig(5) Mobile Application user interface

Deploying a deep convolutional neural network (CNN) model, such as MobileNet, on mobile devices presents two key challenges [20]:

- 1. Model Size:** Deep learning models often have large parameter sizes, resulting in substantial model weight files. These files can exceed the memory limits of many smartphones, making it impractical to store the model directly on the device.
- 2. Computational Resources:** Running a deep CNN model in real-time demands significant computational resources, which not all smartphones possess.

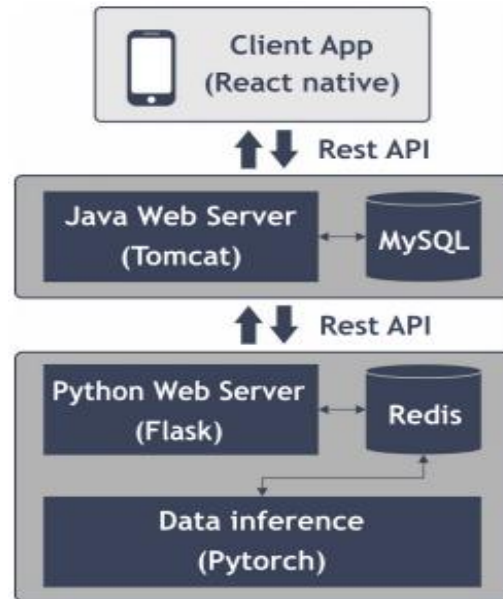
To address these challenges, we adopted a client-server architecture for our skin cancer detection application:

- 1. Model Deployment on Server:** We avoided burdening mobile devices with large model weights and instead deployed the MobileNet-based CNN model on a robust server with ample memory and computational resources. Clinicians send skin lesion images and clinical data via the mobile app to the server for diagnosis.
- 2. Client-Side Mobile Application:** On the client side, we developed a user-friendly, lightweight mobile application for clinicians. This app simplifies the process of capturing skin lesion images and entering clinical information, focusing on user interaction and data collection.
- 3. Server-Side Model Execution:** When a clinician submits an image via the mobile app, it's sent to the server for CNN model inference. The server, with its computational power, performs image analysis, ensuring that even smartphones with limited processing capabilities can effectively use the deep learning model.
- 4. Real-Time Diagnosis Feedback:** After the server completes the analysis, it generates a diagnosis prediction based on the skin lesion image and sends the result back to the mobile



application. The app displays the diagnosis prediction in real-time on the clinician's smartphone, facilitating immediate decision-making during patient screenings.

In adopting this client-server architecture, we effectively addressed the challenges related to model size and computational resources. This approach ensures that clinicians can harness the power of deep learning for skin cancer detection without being constrained by the limitations of their smartphones. In summary, our solution combines mobile technology and server-based deep learning to offer a practical and user-friendly tool for skin cancer detection, ultimately enhancing early diagnosis and patient outcomes.



fig(6) MOBILE APP ARCHITECTURE

## VIII. CONCLUSION

The development and implementation of our skin cancer application have illuminated its potential to profoundly impact the prevention, early detection, and overall management of skin cancer. Our unwavering commitment to user-friendly design, informative content, and cutting-edge technology has culminated in a powerful tool that empowers individuals to proactively manage their skin health and make informed decisions.

As we set our sights on the future, we envision evolving into an indispensable asset in the fight against skin cancer. Through collaborative efforts with healthcare institutions and dermatology experts, we are poised to expand the application's capabilities to include teledermatology consultations. This groundbreaking enhancement will enable users to directly connect with dermatologists for remote assessments, potentially expediting diagnoses and interventions, thus further improving patient outcomes.

Our unwavering commitment to ethical data handling and privacy remains at the core of our mission. We strictly adhere to stringent data protection standards, ensuring that user information is handled securely and in full compliance with relevant regulations. Informed consent stands as a cornerstone of our approach, and we prioritize user data privacy above all else, thereby building trust with our users.

Our innovative smartphone application, seamlessly integrated with the MobileNet deep learning model, stands as a testament to the potential of technology in healthcare. By offloading resource-intensive tasks to a server, we ensure accessibility across a wide spectrum of mobile devices.

This approach holds the potential to revolutionize early diagnosis rates and, as a result, significantly enhance patient outcomes through expedited referrals to specialists.

In summary, we stand ready to continue making a lasting impact in the battle against skin cancer. Together, with the collective efforts of individuals, healthcare providers, and public health organizations, we have the potential to significantly reduce the burden of this prevalent disease. Our vision is one where early detection becomes the norm, ultimately saving lives and improving patient outcomes.

## IX. REFERENCES

- [1] <https://www.skincancer.org/skin-cancer-information/skin-cancer-facts/>
- [2] Dildar M, Akram S, Irfan M, et al. Skin Cancer Detection: A Review Using Deep Learning Techniques. *Int J Environ Res Public Health*. 2021;18(10):5479. Published 2021 May 20. doi:10.3390/ijerph18105479
- [3] La Salvia, Marco, et al. "Deep Convolutional Generative Adversarial Networks to Enhance Artificial Intelligence in Healthcare: A Skin Cancer Application." *Sensors (Basel, Switzerland)* vol. 22,16 6145. 17 Aug. 2022, doi:10.3390/s22166145
- [4] Sarker, I.H. Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. *SN COMPUT. SCI.* 2, 420 (2021). <https://doi.org/10.1007/s42979-021-00815-1>
- [5] Agrahari, P., Agrawal, A., Subhashini, N. (2022). Skin Cancer Detection Using Deep Learning. In: Sivasubramanian, A., Shastry, P.N., Hong, P.C. (eds) *Futuristic Communication and Network Technologies. VICFCNT 2020. Lecture Notes in Electrical Engineering*, vol 792. Springer, Singapore. [https://doi.org/10.1007/978-981-16-4625-6\\_18](https://doi.org/10.1007/978-981-16-4625-6_18)
- [6] Wu Y, Chen B, Zeng A, Pan D, Wang R, Zhao S. Skin Cancer Classification With Deep Learning: A Systematic Review. *Front Oncol*. 2022 Jul 13;12:893972. doi: 10.3389/fonc.2022.893972. PMID: 35912265; PMCID: PMC9327733.
- [7] J. Daghrir, L. Tlig, M. Bouchouicha and M. Sayadi, "Melanoma skin cancer detection using deep learning and classical machine learning techniques: A hybrid approach," 2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), Sousse, Tunisia, 2020, pp. 1-5, doi: 10.1109/ATSIP49331.2020.9231544.
- [8] Andrade, Catarina & Luís, Teixeira & Vasconcelos, Maria & Rosado, Luís. (2020). Deep Learning Models for Segmentation of Mobile-Acquired Dermatological Images. 10.1007/978-3-030-50516-5\_20.
- [9] Brinker, Titus J et al. "Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task." *European Journal of Cancer (Oxford, England: 1990)* vol. 113 (2019): 47-54. doi:10.1016/j.ejca.2019.04.001
- [10] Nawaz, Marriam, et al. "Skin cancer detection from dermoscopic images using deep learning and fuzzy k-means clustering." *Microscopy research and technique* vol. 85,1 (2022): 339-351. doi:10.1002/jemt.23908
- [11] Hum, Y.C., Tan, H.R., Tee, Y.K. et al. The development of skin lesion detection applications in smart handheld devices using deep neural networks. *Multimed Tools Appl* 81, 41579–41610 (2022). <https://doi.org/10.1007/s11042-021-11013-9>
- [12] Shalu and A. Kamboj, "A Color-Based Approach for Melanoma Skin Cancer Detection," 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), Jalandhar, India, 2018, pp. 508-513, doi: 10.1109/ICSCCC.2018.8703309.

- [13] <https://arxiv.org/abs/1803.10417>
- [14] Tschandl, Philipp, 2018, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions", <https://doi.org/10.7910/DVN/DBW86T>, Harvard Dataverse, V4.
- [15] H. S. Obaid, S. A. Dheyab and S. S. Sabry, "The Impact of Data Pre-Processing Techniques and Dimensionality Reduction on the Accuracy of Machine Learning," 2019 9th Annual Information Technology, Electromechanical Engineering and Microelectronics Conference (IEMECON), Jaipur, India, 2019, pp. 279-283, doi: 10.1109/IEMECONX.2019.8877011.
- [16] Kiran Maharana, Surajit Mondal, Bhushan Kumar Nemade, A review: Data pre-processing and data augmentation techniques, Global Transitions Proceedings, Volume 3, Issue 1,2022, Pages 91-99, ISSN 2666-285X,<https://doi.org/10.1016/j.gltip.2022.04.020>.
- [17] <https://towardsdatascience.com/train-test-split-in-deep-learning-e1b1fed01989>
- [18] <https://paperswithcode.com/paper/mobilenets-efficient-convolutional-neural>
- [19] <https://towardsdatascience.com/transfer-learning-using-mobilenet-and-keras-c75daf7ff299>
- [20] Integration of Convolutional Neural Networks in Mobile Applications, March 2021, Roger Creus Castanyer, Silverio Martínez-Fernández, Xavier Franch, <https://arxiv.org/pdf/2103.07286.pdf>