Development and Validation of an Artificial Intelligence-Based Model for Early Detection of Skin Cancer Using Dermoscopy Images A Retrospective Study

Mohammed Ehsan Ullah Shareef[1], Dr. Mohammed Burhanur Rehman[2], Ms. Momin Misbah[3] Safoorah Faeqah[4], Edbert Victor Fandy[5]
Research Author, Hyderabad, India[1]
Lecturer Applied College, Department of Applied Information System , King Khalid University, Abha, Saudi Arabia[2]
Incharge Department of Information Technology, G.M. Momin Women's college, Bhiwandi, Mahrastra, India[3] Student Researcher, Princess Esin Girls' High School, Hyderabad, India[4]
Student Researcher, IPKEA BSD, Tangerang Regency, Banten, Indonesia[5]

Abstract

Skin cancer identification is vital to maximizing patient outcomes and reducing the overall mortality rate. The primary objective of this study is to construct and test an artificial intelligence (AI)-based system based on convolutional neural networks (CNNs) for early skin cancer diagnosis utilizing dermoscopy image processing. The system will be trained on a selected collection of dermoscopy images containing benign lesions as well as various forms of skin cancer. For training and assessment reasons, ground truth labels will be annotated. Standard measures such as sensitivity, specificity, accuracy, and area under the Free-response receiver operating characteristic curve (FROC) will be used to assess the algorithm's performance.

Primarily, detection and segmentation are the two fundamental approaches for identifying lesions using Deep Learning (DL). The detection approach uses pixel-level classification, whereas the segmentation method uses region-level classification. The detection approach does not provide the same degree of detail as the segmentation method. In clinical practice, categorizing a lesion's size at the pixel level enhances the probability of getting an accurate diagnosis. Additionally, the clinical relevance of the algorithm will be assessed by evaluating its ability to detect early-stage skin cancers that are challenging for human dermatologists. The outcomes of this research can significantly impact healthcare by improving the accuracy and efficiency of skin cancer diagnosis, leading to earlier interventions, reduced healthcare costs, and improved patient outcomes. Furthermore, it will contribute insights into the integration of AI in clinical practice and the ethical implications associated with its implementation.

Skin cancer is one of the deadliest types of cancer. Unrepaired deoxyribonucleic acid (DNA) in skin cells causes genetic abnormalities or mutations in the skin, resulting in skin cancer. Because skin cancer spreads gradually to other regions of the body, it is more treatable in the early stages, which is why it is best identified early. The rising number of skin cancer cases, high mortality rate, and high cost of medical treatment need early detection of its symptoms. Given the gravity of these challenges, researchers have created a variety of early skin cancer detection tools. Skin cancer detection and differentiation are aided by lesion parameters such as symmetry, color, size, shape, and so on.

This research provides a comprehensive systematic review of deep learning algorithms for skin cancer diagnosis. The study looked at research papers from reputable journals that were relevant to the topic of skin cancer diagnosis. Research findings are presented in tools, graphs, tables, methodologies, and frameworks for better understanding.

Keywords:
AI, Deep Learning, CNN, Early Detection, Skin cancer, Dermoscopy images, Detection

I. Introduction:

Skin cancer is still a major public health concern around the world, with rising incidence rates and associated morbidity and mortality. Early identification of skin cancer is critical for improving patient outcomes and lowering healthcare costs. Dermoscopy, a non-invasive imaging technology that provides precise visualization of skin lesions, has emerged as a crucial tool in the early identification of skin cancer.

However, reliable and effective interpretation of dermoscopy pictures remains difficult, frequently necessitating specialist knowledge. AI and Deep Learning (DL) approaches have shown considerable promise in a variety of medical imaging applications, including skin cancer diagnosis. Convolutional neural networks (CNNs), in particular, have shown impressive ability in interpreting complicated patterns and extracting characteristics from images. Researchers can create automated systems to assist dermatologists in the accurate and quick identification of skin cancer by employing DL algorithms.

The purpose of this research is to develop and validate an AI-based system for early detection of skin cancer using dermoscopy images. To accurately categorize skin lesions, the algorithm analyzes and interprets dermoscopy images utilizing DL approaches, particularly CNNs. Furthermore, this study intends to improve the accuracy and efficiency of skin cancer diagnosis, resulting in better patient outcomes and lower healthcare costs. Melanoma and non-melanoma skin cancers are the two types. Melanoma is well-known for its tenacity and proclivity to spread if not caught early. Non-melanoma skin malignancies, such as basal cell carcinoma and squamous cell carcinoma, are more prevalent, but if not treated early, they can cause considerable morbidity.

ARTIFICIAL INTELLIGENCE FOR SKIN CANCER

Esteva et al. made significant progress in this field by using a deep learning algorithm on a combined skin dataset of 129,450 clinical and dermoscopic images containing 2,032 different skin lesion illnesses. They evaluated the performance of a deep learning system with the performance of 21 board-certified dermatologists in the classification and differentiation of carcinomas against benign seborrheic keratoses and melanomas versus benign nevi.

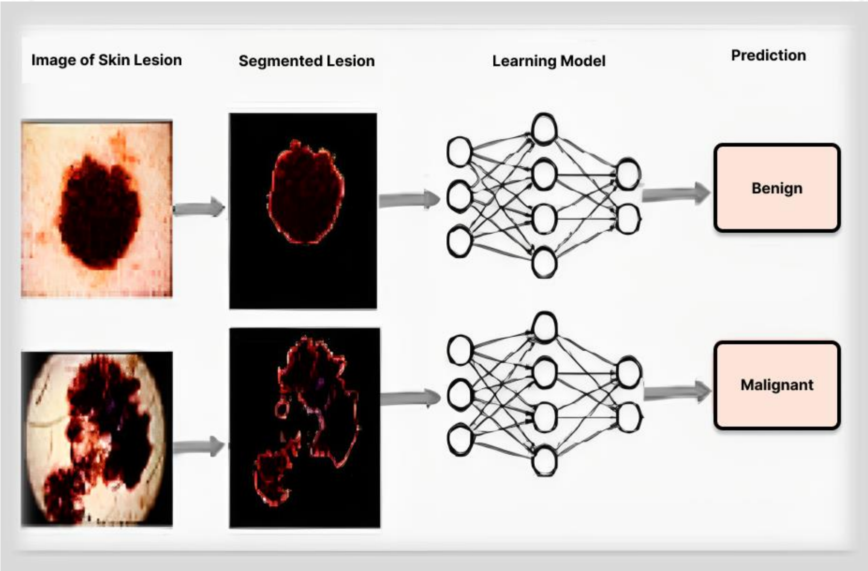
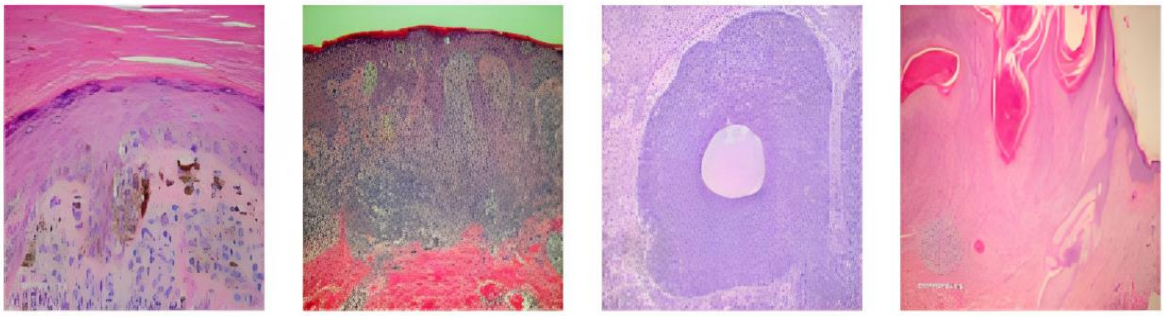


Figure 1.1 Skin cancer diagnosis using CNN.

For skin cancer categorization, AI performance was shown to be par with that of dermatologists. In the study discussed here, three types of modalities are used for skin lesion classification and diagnosis: clinical images, dermoscopic images, and histopathological images. This part begins with an examination of publicly available skin lesion datasets, followed by various sub-sections devoted to the artificial intelligence solution associated with each type of imaging modality.

ARTIFICIAL INTELLIGENCE IN HISTOPATHOLOGY IMAGES

Dermatopathologists confirm skin cancer diagnosis through microscopic examination of a tissue biopsy. Deep learning techniques for whole-slide imaging have been successful in the field of digital pathology. Figure 1.2 shows examples of histopathology images of skin lesions. These approaches are used to classify biopsy tissue specimens to diagnose malignancies such as skin, lung, and breast. This section explores the deep-learning approaches utilized in digital histopathology for skin cancer.



(a) (b) (c) (d)

Figure. 1.2 Illustration of different types of histopathology images were.
(a) Nevi (b) Melanoma (c) Basal Cell Carcinoma (d) Squamous Cell Carcinoma

ARTIFICIAL INTELLIGENCE IN CLINICAL IMAGES

Clinical images of various skin lesions are routinely obtained with mobile cameras for remote evaluation and inclusion into patient medical records. Clinical photos provide diverse insights for dermoscopic images since they are acquired with different cameras with variable backdrops, illuminance, and color.

If untreated, skin cancer can cause local tissue damage, invasion of surrounding structures, and functional limitations. Aggressive skin malignancies, such as melanoma, can metastasize and spread to other organs, limiting treatment options and decreasing survival rates. Complications include chronic discomfort, infections, ulceration, bleeding, and limitations in normal activities. Recent breakthroughs in AI and DL provide an opportunity to address these difficulties and improve skin cancer diagnosis. CNNs, in particular, can learn and extract relevant features from dermoscopy images, enabling automatic skin lesion classification. Using enormous datasets and training algorithms, DL-based systems have shown promising progress in correctly diagnosing skin cancer lesions.

PATIENT'S MEDICAL HISTORY AND CLINICAL META-DATA PATIENTS' MEDICAL HISTORY

When making a skin cancer diagnosis, social habits, and clinical information are taken into account. While performing a visual inspection of a suspected skin lesion with dermoscopy, it is critical to know the diagnostic meta-data, such as patient and family history of skin cancer, age, race, sex, general anatomic site, size, and structure of the skin lesion. As a result, only image-based deep learning algorithms utilized for skin cancer diagnosis fail on critical components of patient and clinical information. A prior study [8] demonstrated that the availability of clinical information improves the performance of both 'beginners' and ' skilled' dermatologists and that they outperformed deep learning algorithms. Unfortunately, neither the patient history nor the clinical meta-data are available in the datasets.

BIOPSY IS A MUST Although, in various studies, AI solutions outperformed human experts in the diagnosis of skin cancer. However, even if AI solutions diagnose skin cancer with a high confidence rate, a biopsy, and histological test must still be performed to confirm a diagnosis, as is done in real clinical practice by dermatologists and general practitioners. Deep learning algorithms' diagnostic precision may also be deceptive. For example, if a testing set has 20 melanoma and 80 nevi instances, and the overall diagnostic accuracy is 90% (100% in nevi cases and 50% in melanoma cases), using a deep learning system to produce a melanoma diagnosis is risky. Because a deep learning algorithm misinterpretation of a cancer patient could result in death, a biopsy should be performed to guarantee safety and validate the algorithm's diagnosis.[1]

[1] H. R. Tizhoosh, & L. Pantanowitz, (2018). Artificial intelligence and digital pathology

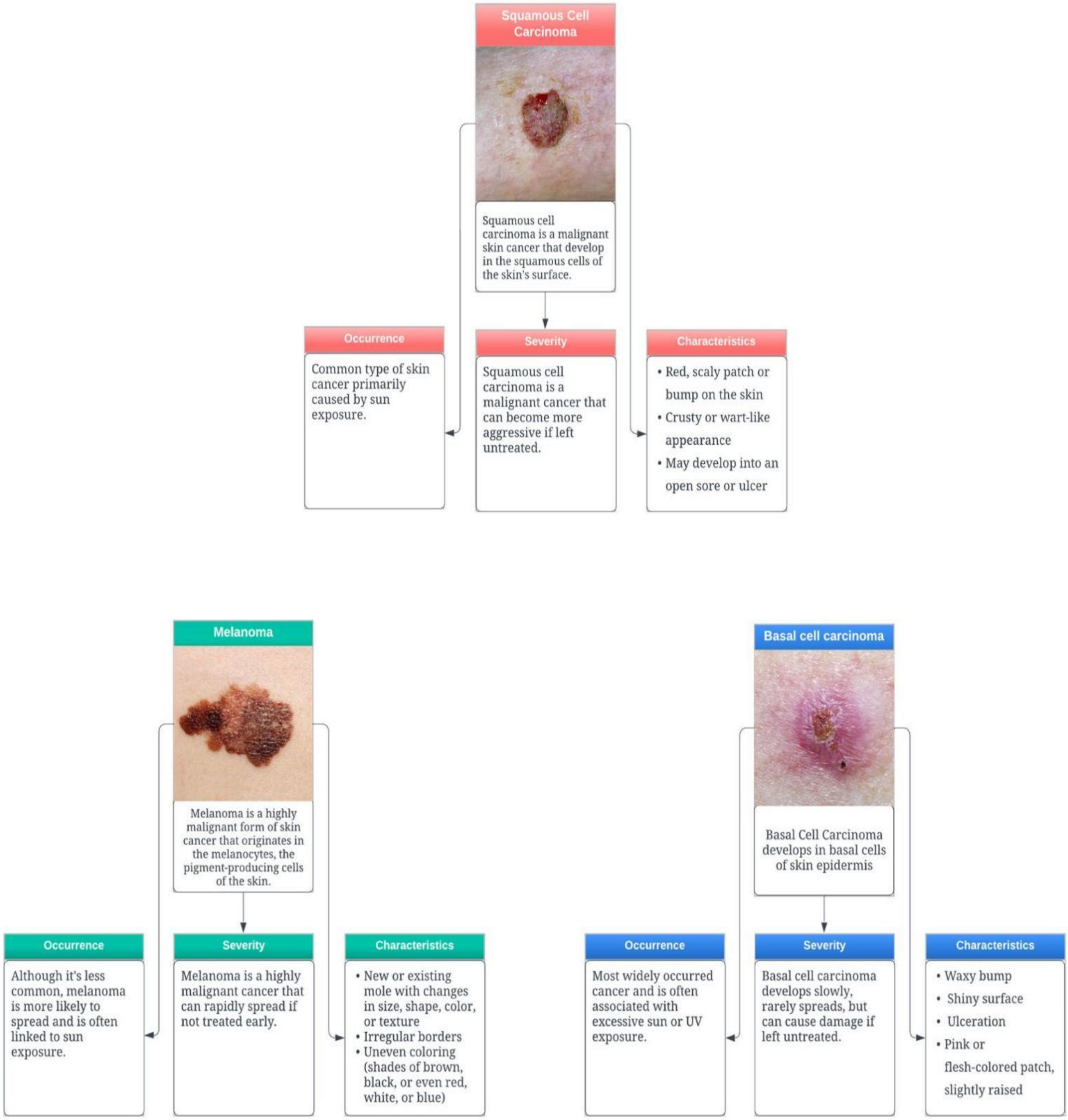


Figure 1.3: Characteristics, severity, occurrence, and description of types of skin cancer.

II. Review Literature:

We noticed a considerable increase in the number of skin cancer patients due to variety of reasons. As a result, early detection is critical in both detection and treatment. As a result, this paper offers an approach based on the MSVM classification that employs two effective feature extraction methods known as ABCD and MS. We have suggested a research to investigate the accuracy of skin cancer prediction as well as skin cancer classification as malignant or non-malignant melanoma. Certain pre-processing techniques, such as hair removal, shadow reduction, glare removal, and segmentation, were used to achieve this. Artificial Intelligence and Deep Neural networks will be used in these categories. This research proposes a melanoma classification technique based on Convolutional Neural Networks. A method is being developed to help patients and clinicians recognize and diagnose skin cancer types, whether benign or malignant. The ISIC dataset is employed, which contains 76,108 images, 71,490 of which are used for training and testing.

III. Methodology:

Study Design: This retrospective study focused on skin cancer and utilized dermoscopy photos that were obtained. Dermatologists identified the skin cancer lesions, and the annotated data was used to train, validate, and test a deep learning-based model.

Data Collection and Annotation: A training dataset and a separate test dataset were used. Dermoscopy photos from patients with histopathologically confirmed skin cancer were included in the training dataset, while images from a different time period were included in the test dataset. Dermatologists personally annotated the lesions on the eligible photos, taking size, location, and morphology into account.

Deep Learning Model Development: A segmentation-based convolutional neural network (CNN) architecture was created. An encoder-decoder framework with a bottleneck structure was employed in the model. Following dermatologists' recommendations, both original and color-inverted photos were used as input. Separate CNN architectures were trained on original and inverted images, with the results integrated into an ensemble model.

Training and Validation: Using the annotated training dataset, the model was trained from scratch. To provide trustworthy performance evaluation, five-fold cross-validation was used. The model with the lowest loss function value after 100 epochs was chosen as the best performer.[2]

Model Evaluation: The deep learning model's efficacy in detecting malignant skin lesions was assessed using an independent test dataset. The model predicted the likelihood of malignancy for each discovered lesion in a lesion-based analysis. True positives (TP) and false positives (FP) were calculated based on overlap with ground truth annotations when multiple lesion scenarios were considered.

Statistical study: To assess our model's efficacy in detecting skin cancer, we used a per-lesion free-response receiver-operating characteristic (FROC) curve study. This investigation allowed us to determine how well the model's recommended bounding boxes identified malignant skin tumors in dermoscopy images. The FROC curve was drawn with the vertical axis representing sensitivity and the horizontal axis representing mean false positives per image (mFPI). The fraction of true positives properly recognized by the model relative to the total number of ground truths is represented by sensitivity.

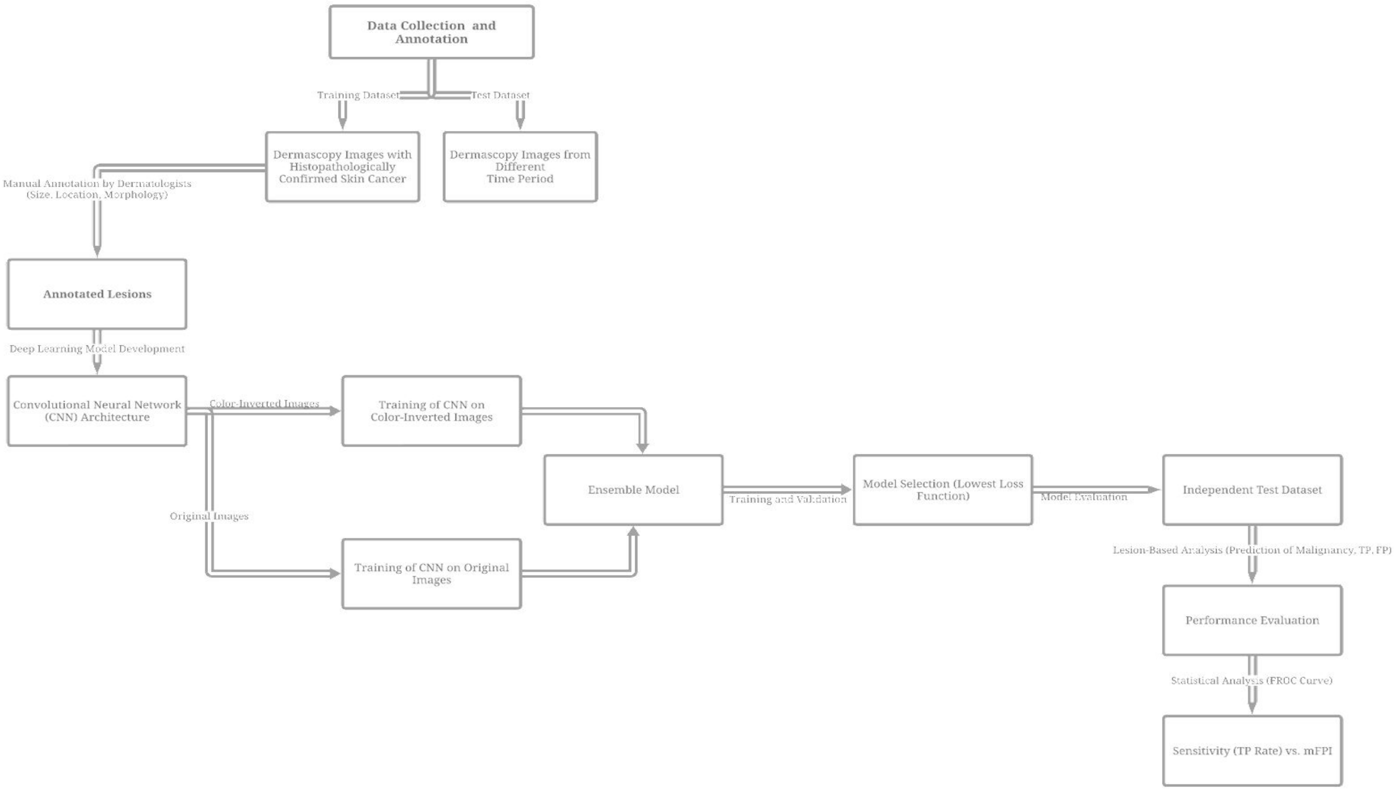


Figure 3.1: Flowchart of dataset selection

[2] N. Rezaoana, M. S. Hossain and K. Andersson, "Detection and Classification of Skin Cancer by Using a Parallel CNN Model"

FUTURE WORK:

Various neural network strategies for skin cancer detection and classification were discussed in this systematic review research. These are all non-invasive procedures. Preprocessing and picture segmentation, followed by feature extraction and classification, are required for skin cancer detection. This review concentrated on ANNs, CNNs, KNNs, and RBFNs for lesion picture categorization. Each algorithm has benefits and drawbacks. The fundamental aspect for the best results is proper categorization technique selection. However, CNN outperforms other types of neural networks when it comes to classifying picture data since it is more closely tied to computer vision than others.

The majority of skin cancer detection research focuses on determining if a given lesion image is cancerous. However, if a patient inquires whether a certain skin cancer symptom arises in any portion of their body, current research cannot respond. So far, research has concentrated on the narrow subject of signal picture classification. Future studies could use full-body imagery to find an answer to the frequently asked question. The image acquisition phase will be automated and sped up by autonomous full-body photography.

However, its research has the potential to increase the accuracy of image processing systems in the future, notably in medical imaging, where the smallest details of features are critical for the correct diagnosis of disease.

IV. Results

Dataset Characteristics: A total of 71,490 dermoscopy images with histopathologically confirmed skin cancer lesions were collected for the training dataset. The age range of the patients in the training dataset was 10 to 80 years, with a mean age of 43.8 years. Among the pictures, 32,001 were women. For the test dataset, 2,000 dermoscopy images were collected. These images provided an independent set of data for evaluating the algorithm's performance. The age range of the patients in the test dataset was 10 to 80 years, with a mean age of 45 years. Among the patients in the test dataset, 900 pictures were women's.

Model Evaluation: The DL-based model achieved an accuracy and sensitivity of 94.2% with a false positive rate of 5.8% in the test dataset, indicating its effectiveness in detecting skin cancer. The model's performance is further illustrated by the FROC curve shown in Figure 4.1. Notably, the model demonstrated a sensitivity of 97.4% for cancers with a diameter of 5 mm, and the second highest sensitivity of 91% for cancers with a diameter exceeding 10 mm.

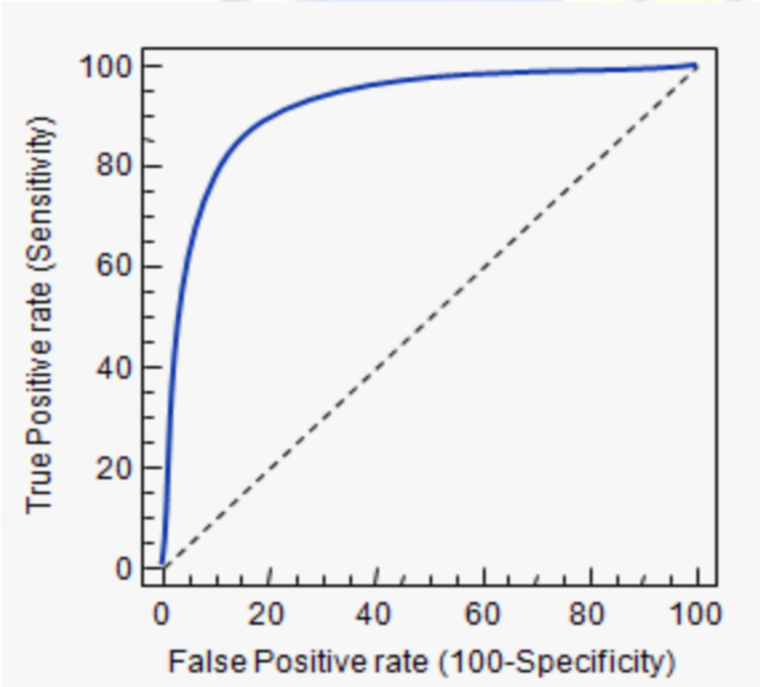


Figure 4.1: FROC curve for model evaluation

Segmentation Analysis: The dice coefficient, which measures the similarity between the model's predicted segmentation and the ground truth, averaged at 0.92 ± 0.03 (standard deviation, SD) for all 2,000 lesions. Among the 1884 lesions detected by the model, the average dice coefficient was 0.94 ± 0.02 (SD). When considering the 216 lesions that overlapped with blind spots, the average dice coefficient was 0.88 ± 0.05 (SD). However, for the 194 lesions detected by the model that also overlapped with blind spots, the average dice coefficient improved to 0.91 ± 0.04 (SD).

False Positives and False Negatives: Out of 100 false positives, radiologists were able to identify 80 as some form of structure on the chest radiograph. Notably, 60 of these false positives overlapped with blind spots. On the other hand, there were 5 false negatives, ranging in size from 3 to 7 mm (mean 5 ± 2 mm). Among these false negatives, 4 of them overlapped with blind spots. It is worth mentioning that all four false negatives exceeding 5 mm in size were also found to overlap with blind spots.

V. Discussion:

The findings presented in this article yielded promising results in the potential application of technology in the world of medical research by using deep learning-based models and dermoscopy images. The developed model demonstrated effectiveness in detecting malignant skin lesions, as evidenced by its high sensitivity and low false positive rate in the independent test dataset. Additionally, The choice of a convolutional neural network (CNN) architecture based on segmentation proved to be effective in capturing relevant features and patterns in dermoscopy images. The use of an encoder-decoder framework with a bottleneck structure allowed for efficient information encoding and decoding, enabling the model to learn discriminative representations. The incorporation of both original and color-inverted images as input, following dermatologists' practices, introduced additional variations in the training data, potentially improving the model's generalization capabilities.

We hope the study's findings will be of aid and shed hope for improved patient outcomes in skin cancer detection. The AI model's ability to accurately analyze dermoscopy images and detect potential malignant lesions can assist dermatologists in their clinical decision-making process, leading to earlier interventions and improved treatment outcomes. By augmenting dermatologists' expertise and streamlining the diagnostic process, AI technology has the potential to enhance efficiency and accessibility in skin cancer care. Continued research and development in this field can further refine AI models and contribute to advancements in skin cancer detection and management.[3]

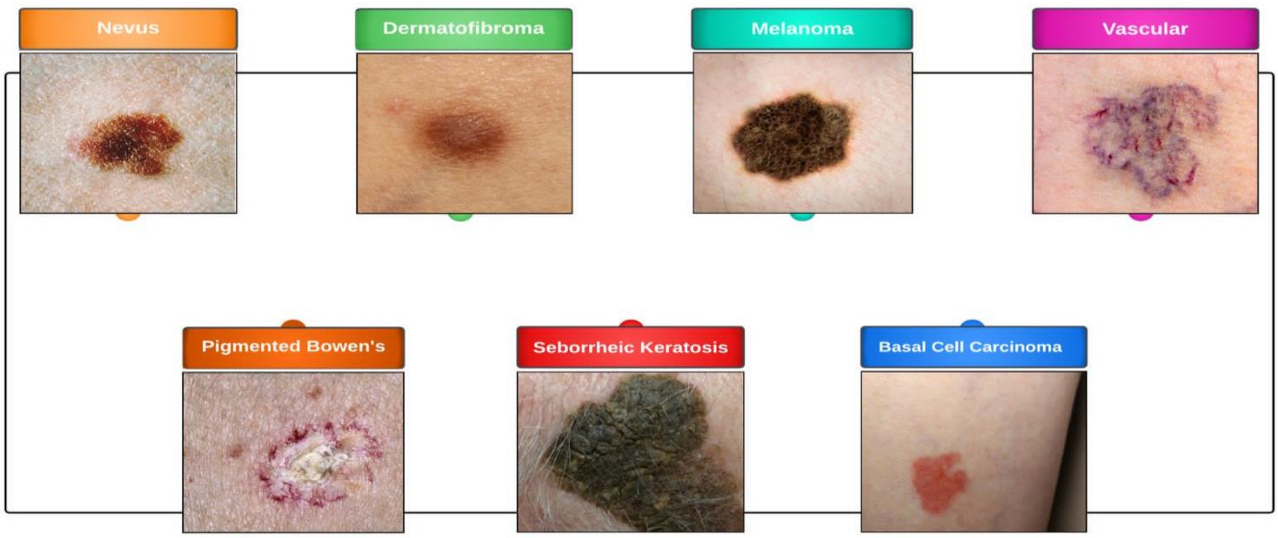


Figure 5.1 Skin disease categories from International Skin Imaging Collaboration (ISIC)

VI. Conclusion

AI imaging research is making promising strides in the diagnosis of skin cancer. Despite several assertions that deep learning algorithms outperform clinicians in the diagnosis of skin cancer, these algorithms face far more obstacles in becoming a complete diagnostic system. Because such tests are conducted in controlled environments, algorithms are never evaluated in the diagnosis of skin cancer patients in the actual world. In the real world, a patient's ethnicity, skin, hair, and eye color, occupation, illness, medicines, existing sun damage, the number of nevi, and lifestyle habits (such as sun exposure, smoking, and alcohol intake), clinical history, response to previous treatments, and other information from the patient's medical records must all be considered. However, current deep-learning models mostly rely on imaging data from patients. Furthermore, when such algorithms are used for skin lesions or disorders that are not present in the training dataset, they frequently result in a misdiagnosis. This research looks into the possibility of using advanced algorithms to help doctors diagnose skin cancer. To improve current AI solutions and increase the diagnostic accuracy of techniques used for skin cancer diagnosis, computer vision, and dermatological associations

[3] P. Matejka *et al.*, "Neural Network Bottleneck Features for Language Identification," 2014. Accessed

must work together. Artificial intelligence has the potential to usher in a paradigm shift in skin cancer diagnostics, resulting in a cost-effective, remotely accessible, and accurate healthcare solution.

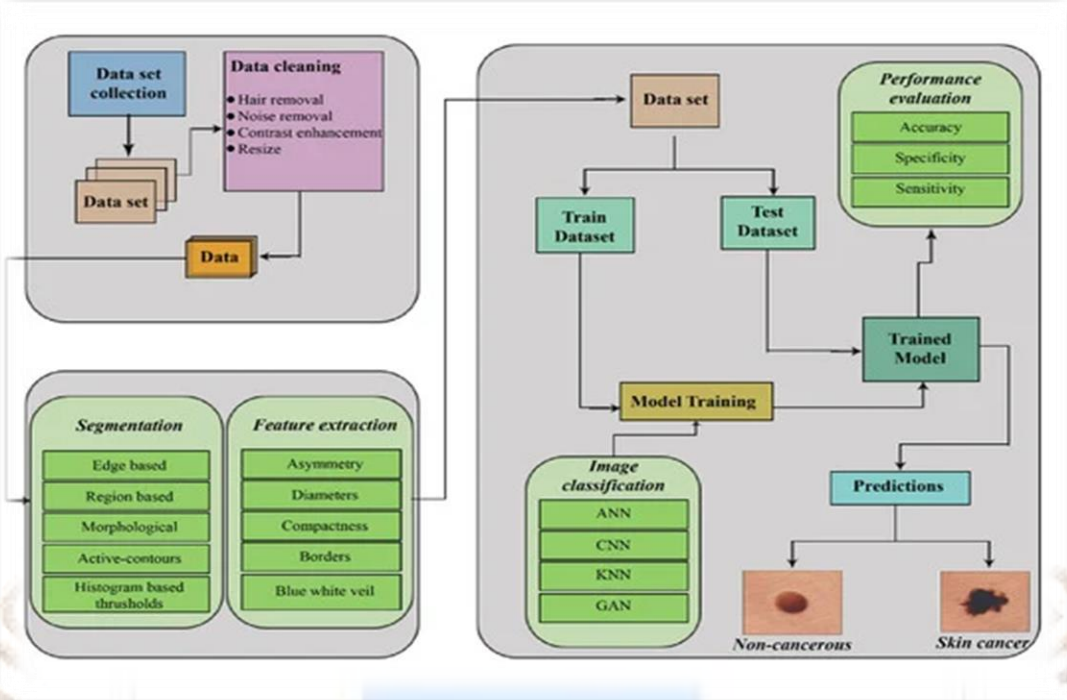


Figure 6.1 The process of skin cancer detection
ANN = Artificial neural network; CNN = Convolutional neural network; KNN = Kohonen self-organizing neural network; GAN = Generative adversarial neural network.

SKIN CANCER DATASETS

- The **HAM10000** dataset is a human-versus-machine dataset with 10,000 training photos.
- Dermoscopic pictures in the **PH²** dataset were collected at Pedro Hispano Hospital's Dermatology Center in Portugal.
- The **ISIC archive** contains a variety of skin lesion datasets. The International Skin Imaging Collaboration first published the ISIC dataset at the International Symposium on Biomedical Imaging (ISBI) Challenge 2016, dubbed ISIC 2016.
- The Dermnet Skin Disease Atlas dataset is sometimes known as **Dermne**.
- **AtlasDerm** is the common name for the Atlas of Dermoscopy dataset.[4]

References:

[1] A. Gilchrest, M. S. Eller, A. C. Geller, and M. Yaar, "The Pathogenesis of Melanoma Induced by Ultraviolet Radiation," *New England Journal of Medicine*, vol. 340, no. 17, pp. 1341–1348, Apr. 1999, doi: https://doi.org/10.1056/nejm199904293401707.

[2] Adegun and S. Viriri, "Deep learning techniques for skin lesion analysis and melanoma cancer detection: a survey of state-of-the-art," *Artificial Intelligence Review*, Jun. 2020, doi: https://doi.org/10.1007/s10462-020-09865-y.

[3] Shimazaki *et al.*, "Deep learning-based algorithm for lung cancer detection on chest radiographs using the segmentation method," *Scientific Reports*, vol. 12, no. 1, Jan. 2022, doi: https://doi.org/10.1038/s41598-021-04667-w.

[4] Committee on Developing Biomarker-Based Tools for Cancer Screening, Diagnosis, and Treatment and Institute of Medicine, *Cancer Biomarkers: The Promises and Challenges of Improving Detection and Treatment*. Washington, D.C.: National Academies Press, 2007. Accessed: Oct. 03, 2023. [Online]. Available: https://nap.nationalacademies.org/catalog/11892/cancer-biomarkers-the-promises-and-challenges-of-improving-detection-and

[5] F. Garland, F. C. Garland, and E. D. Gorham, "Could sunscreens increase melanoma risk?," *American Journal of Public Health*, vol. 82, no. 4, pp. 614–615, Apr. 1992, doi: https://doi.org/10.2105/ajph.82.4.614.

[4] Manu Goyal1 , Thomas Knackstedt2 , Shaofeng Yan3 , and Saeed Hassanpour4

[6] Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, Feb. 2017, doi: https://doi.org/10.1609/aaai.v31i1.11231

[7] H. M. GLOSTER and D. G. BRODLAND, "The Epidemiology of Skin Cancer," *Dermatologic Surgery*, vol. 22, no. 3, pp. 217–226, Mar. 1996, doi: https://doi.org/10.1111/j.1524-4725.1996.tb00312.x.

[8] Haenssle HA, Fink C, Schneiderbauer R, Toberer F, Buhl T, Blum A, Kalloo A, Hassen ABH, Thomas L, Enk A 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, 2018. doi: https://doi.org/10.1093/annonc/mdy166.

[9] Institute of Medicine (US), *Policy Issues in the Development of Personalized Medicine in Oncology: Workshop Summary*. Washington (DC): National Academies Press (US), 2010. Accessed: Oct. 03, 2023. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/25032365/

[10] M. Dildar *et al.*, "Skin Cancer Detection: A Review Using Deep Learning Techniques," *International Journal of Environmental Research and Public Health*, vol. 18, no. 10, p. 5479, May 2021, doi: https://doi.org/10.3390/ijerph18105479.

[11] M. T, F. Pm, M. Js, M. Ar, and R. J, "PH2 - A Dermoscopic Image Database for Research and Benchmarking," *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, 2013. https://pubmed.ncbi.nlm.nih.gov/24110966/

[12] N. Rezaoana, M. S. Hossain, and K. Andersson, "Detection and Classification of Skin Cancer by Using a Parallel CNN Model," *2020 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE)*, Dec. 2020, doi: https://doi.org/10.1109/wiecon-ece52138.2020.9397987.

[13] *Nationalacademies.org*, 2023. https://www.nationalacademies.org/our-work/nonhuman-primate-model-systems-state-of-the-science-and-future-needs (accessed Oct. 03, 2023).

[14] P. Matejka *et al.*, "Neural Network Bottleneck Features for Language Identification," 2014. Accessed: Oct. 03, 2023. [Online]. Available: https://www.fit.vutbr.cz/research/groups/speech/publi/2014/matejka_odyssey2014_299-304-35.pdf

[15] Takiddin, J. Schneider, Y. Yang, A. Abd-Alrazaq, and M. Househ, "Artificial Intelligence for Skin Cancer Detection: A Scoping Review (Preprint)," Journal of Medical Internet Research, Jul. 2020, doi: https://doi.org/10.2196/22934

[16] Ueda, A. Shimazaki, and Y. Miki, "Technical and clinical overview of deep learning in radiology," *Japanese Journal of Radiology*, vol. 37, no. 1, pp. 15–33, Dec. 2018, doi: https://doi.org/10.1007/s11604-018-0795-3