IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.53

Volume 4, Issue 5, April 2024

Pixel Based Multi Class Skin Cancer Classification using Convolutional Neural Network

P. R. S. S.Venkatapathi Raju[1], Matcha Manoj Kumar[2], Mugada Raju[3], Kotla Yashwanth[4], Kilaparthi Gokul Karthik[5]

Assistant Professor, Department of Information Technology[1]

Students, Department of Information Technology[2,3,4,5]

S.R.K.R. Engineering College, Bhimavaram, Andhra Pradesh, India

prssvraju2@gmail.com, matchamanojkumar77@gmail.com

mugadaraju485@gmail.com, yashwanthkotla1@gmail.com, gokulkarthikkilaparthi@gmail.com

Abstract: *Skin cancer poses a significant societal concern, as pigments responsible for skin colour can become carcinogenic, leading to the development of the disease. Detecting lesions early is vital for effective skin cancer treatment, yet diagnosis is challenging due to the similarity of many pigments. To aid dermatologists, substantial advancements have been achieved in developing automated tools utilizing artificial intelligence. One notable tool is a Computer-Aided Diagnosis (CAD) system employing a Convolutional Neural Network (CNN) to classify seven types of skin cancer from atypical lesion images.*

In recent years, there has been significant progress in the development of automated cancer classification systems utilizing deep learning techniques. However, these systems often exhibit bias towards fair skin tones due to datasets skewed towards lighter skin complexions. This project addresses this challenge by proposing a model comparing among Convolution Neural Network and Multi - Layer Perceptron deep learning architecture specifically designed for fair skin cancer classification. The proposed model incorporates with a pre - trained model on an imbalanced pixel dataset to mitigate bias on skin colour and improve classification accuracy for fair skin cancers. The model is evaluated on a HAM10000 pixel dataset for fair skin cancer classification, demonstrating competitive performance while promoting fairness in cancer diagnosis using minimal resources..

Keywords: Skin cancer classification, Convolutional Neural Network (CNN), Multi – Layer Perceptron, Pixels, Image Classification

I. INTRODUCTION

Skin cancer poses a significant global health concern, impacting individuals' quality of life. While melanoma constitutes the least common form of skin cancer, it is alarmingly responsible for 75% of skin cancer-related deaths due to its rapid spread if not detected early. Initiatives like the International Skin Imaging Collaboration (ISIC) play a pivotal role in leveraging skin images to mitigate melanoma mortality rates. Early diagnosis and treatment significantly improve melanoma prognosis, highlighting the critical importance of timely intervention.

The utilization of digital skin lesion images holds promise for the development of automated tele dermatology systems, facilitating clinical decision-making processes. Early detection and prevention remain paramount in combating skin cancer. Vigilance toward any new or changing skin lesions, especially those exhibiting unusual characteristics, is imperative. Prompt evaluation by healthcare professionals is essential for monitoring any progressive changes in lesion appearance, including alterations in size, shape, or colour.

The emergence of deep learning methodologies has revolutionized skin cancer detection, enabling the classification of lesions into seven diagnostic categories. This advancement offers a more sophisticated approach to diagnosis, enhancing the accuracy and efficiency of skin cancer detection and treatment. By harnessing the power of deep learning, healthcare practitioners can effectively leverage technology to improve patient outcomes and reduce mortality rates associated with skin cancer.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, April 2024

- **Melanocytic Nevus**: Typically benign, melanocytic nevus, also known as birthmarks or moles, originates from irregularities in pigment-producing skin cells, or melanocytes. These moles often manifest as larger lesions.
- **Melanoma**: Recognized as the deadliest form of skin cancer, melanoma arises when melanocytes, the cells responsible for skin pigmentation, undergo malignant transformation.
- **Benign Keratosis**: Characterized by waxy brown, black, or tan growths, benign keratosis is a non-cancerous skin condition commonly observed in older individuals. While solitary occurrences are possible, multiple growths are more prevalent.
- **Basal Cell Carcinoma**: Originating in basal cells, which generate new skin cells, basal cell carcinoma constitutes a common type of skin cancer. Sun exposure plays a significant role in its development, emphasizing the importance of sun protection.
- **Actinic Keratosis**: Resulting from prolonged sun exposure, actinic keratosis presents as rough, scaly patches on the skin, predominantly affecting older adults. Minimizing sun exposure is key to mitigating risk.
- **Vascular Lesion**: Vascular lesions, often referred to as birthmarks, encompass various skin and underlying tissue irregularities.
- **Dermatofibroma**: Dermatofibroma, a frequently encountered cutaneous nodule primarily affecting women, typically emerges on extremities, particularly the lower legs. While its exact cause remains unknown, dermatofibroma is generally asymptomatic.

II. LITERATURE SURVEY

Naqvi and colleagues (2023) developed deep learning algorithms for real-time skin cancer detection, aiming to support dermatologists by comparing various methods in skin cancer classification while considering their performance and computational costs. Mishra et al. (2019) suggested improvements for deep learning-based dermoscopic classification and dataset creation. Esteva et al. (2017) utilized a CNN framework to train a large-scale skin disease dataset, achieving results comparable to dermatologists and tailored for mobile devices. Lopez et al. (2017) reduced model training time using transfer learning while maintaining high sensitivity and precision. Walker et al. (2019) introduced sonification to enhance model sensitivity in diagnosing skin cancer lesions. Perez et al. (2019) systematically evaluated factors influencing CNN structure choice by analysing 13 factors from 9 models. Polat et al. (2020) proposed a method combining CNN with one-versus-all (OVA) for skin disease classification. Rahman et al. (2021) employed a grid search strategy to identify the best ensemble learning methods for skin cancer classification. Kawahara et al. (2019) designed a multi-task network to classify the seven-point checklist and skin disease diagnosis, utilizing different loss functions to handle various input modalities. Lastly, Hameed et al. (2020) introduced an expert system "i-Rash" based on SqueezeNet to classify four skin diseases.

III. EXISTING SYSTEMS

The utilization of deep learning in the classification of skin cancer marks a significant breakthrough in the field of medical image analysis. Deep learning methodologies, particularly convolutional neural networks (CNNs), have showcased impressive capabilities in accurately discerning and categorizing skin lesions. By leveraging extensive datasets comprising annotated skin images, these systems empower neural networks to discern intricate patterns and features associated with diverse forms of skin cancer.

Initiatives led by organizations such as IEEE and Semantic Scholar are actively engaged in the development of various models aimed at classifying and detecting skin cancer through the analysis of skin lesion images, achieving notable accuracy rates, often reaching around 86%. However, a notable challenge lies in the potential bias observed in these systems towards individuals with fair skin tones. This bias stems from datasets that predominantly feature lighter skin complexions, highlighting the need for greater diversity and inclusivity in dataset curation to ensure robustness and fairness in skin cancer classification algorithms.

IV. PROBLEM STATEMENT

The objective of skin cancer classification lies in the development of a machine learning model capable of categorizing skin lesions into specific types of skin cancer with high accuracy, while utilizing minimal system resources and a reduced number of layers. Timely and precise diagnosis is critical for effective treatment, given the prevalence of skin cancer as the most common form of cancer. The three primary types of skin cancer—basal cell carcinoma, squamous cell carcinoma, and melanoma—require distinct classification approaches.

The issue of bias in AI systems towards individuals with fair skin tones is evident, stemming from imbalanced datasets that predominantly feature lighter complexions. To address this, diverse learning models incorporating numerous neurons are employed for skin lesion image classification. Through extensive training on comprehensive datasets, the objective is to develop models capable of accurately classifying lesions into specific types of skin cancer, with a particular emphasis on identifying malignant cases, all while optimizing resource utilization.

Accurate skin cancer classification models play a vital role in early detection and treatment, potentially saving lives and improving patient outcomes. By efficiently leveraging machine learning techniques, these models contribute to the advancement of medical diagnostics and ultimately enhance the quality of healthcare delivery.

V. PROPOSED SYSTEM

This project addresses this challenge by proposing a **Convolutional Neural Network** deep learning architecture specifically designed for fair skin cancer classification. The proposed model incorporates with a pre-trained model on a balanced dataset to mitigate bias and improve classification accuracy for fair skin cancers. The model is evaluated on a **HAM10000_28_28_RGBPIXELdataset** which contains 10015 images pixels of 7 types of skin cancer for fair skin cancer classification, demonstrating competitive performance while promoting fairness in cancer diagnosis.

We aim to enhance accessibility by leveraging an existing model and improving the current system. To achieve this, we will develop a user-friendly website where users or dermatologists can upload patient skin lesion images. The model will swiftly analyze the data and provide results instantaneously. Through extensive testing with various architectures and methodologies, including comparing multiple neural network models, we achieved an accuracy of over 75% using the HAM10000 dataset. Additionally, we tested the model with randomly generated augmented images and achieved nearly identical accuracy and precision.

VI. ARCHITECTURE

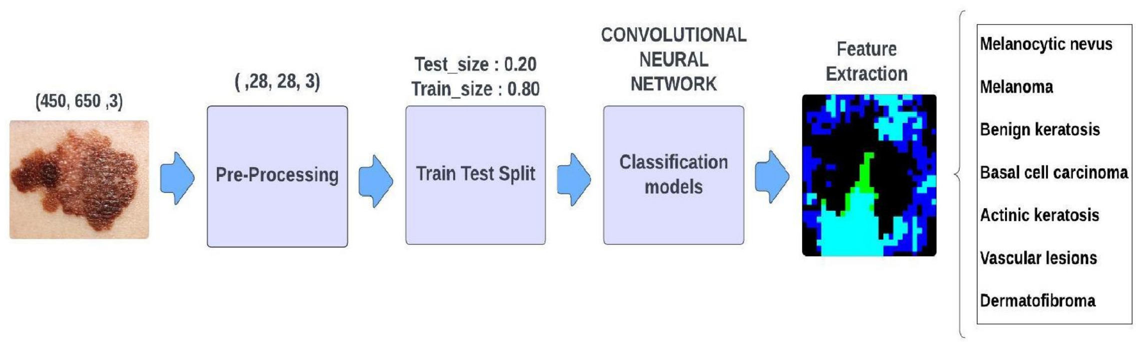

Figure 1 System Architecture

VII. METHODOLOGY

Dataset : https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000?select=hmnist_28_28_RGB.csv

The HAM10000_28_28_RGB ("Human Against Machine with 10015 training images pixels") dataset, accessible through the ISIC archive, comprises a vast collection of skin lesion images. With thousands of images available, this dataset serves as a valuable resource for training machine learning models aimed at automating the analysis of skin lesions. The primary objective is to aid in the early detection of skin cancer by leveraging these images.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, April 2024

Through the analysis of this dataset, machine learning models can acquire the capability to distinguish between benign and malignant lesions. This differentiation is pivotal in assisting dermatologists during the diagnostic process, potentially enhancing the accuracy and efficiency of skin cancer diagnosis. By harnessing the wealth of information contained within these images, these models contribute to advancements in medical technology and the improvement of patient care.

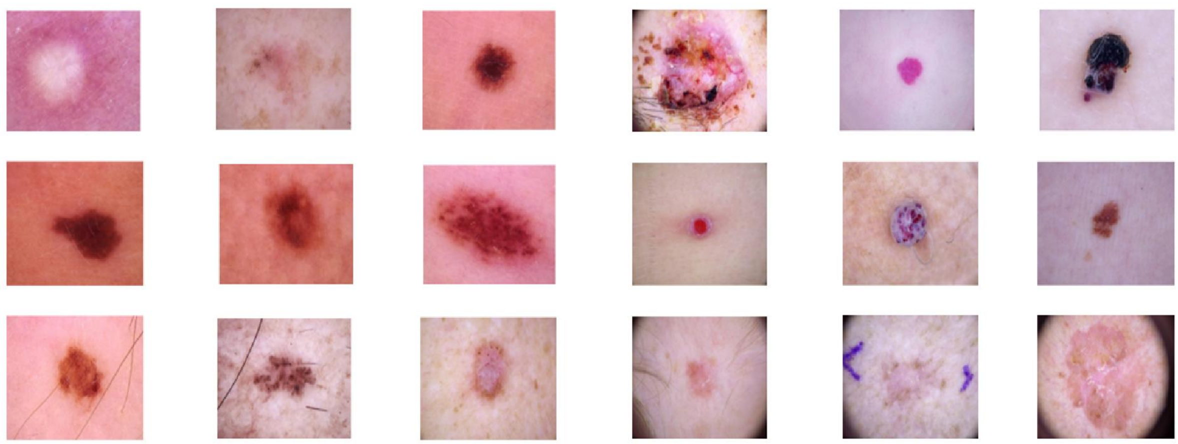

Figure 2 Sample Skin Lesions

Data Pre-processing :

Image Resizing

The original HAM10000 images, with dimensions of 450x600 pixels, can be quite large for some machine learning models, especially those with limited computational resources. So, we take Pixel representation of images. Which is ranges from 0 to 255. It sets the height and width variables to 28 pixels each. Then, it reshapes the training and test image datasets to have dimensions of (-1, 28, 28, 3), where -1 indicates that the number of samples remains unchanged, 28x28 represents the new height and width, and 3 indicates the number of colour channels (e.g., RGB). This reshaping prepares the data for further processing or analysis, such as training a machine learning model.

Normalization

Another key preprocessing step is normalization. Raw image pixel values in the HAM10000 dataset typically range from 0 to 255. Normalization involves dividing each pixel value by 255, effectively scaling all values to a range between 0 and 1. This creates a consistent data format for the model, improving its ability to learn patterns across the entire dataset. Normalization can lead to faster training times and potentially better model performance. In this model we use standard scaler for normalizing the data.

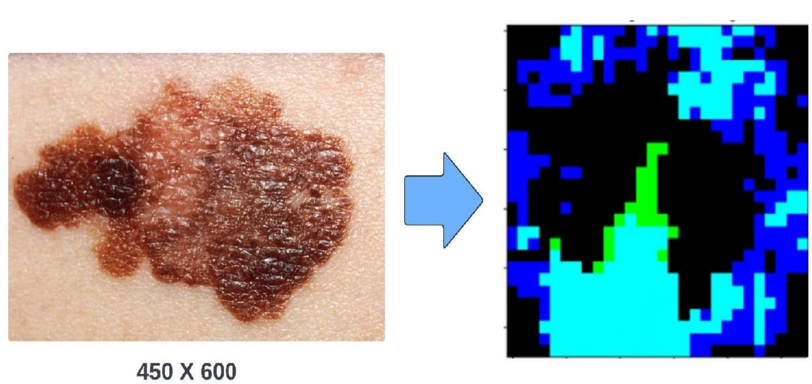

Figure 3 Feature extraction of input Image.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, April 2024

Feature Extraction

As most of the Datasets consists non coloured skin lesions, results in skewness of models towards lighter skin. Extracting features as pixels from the input images helps to classify the respected melanoma of any coloured skin. In this phase the system extracts the pixels of input image and then subjected to the model. This process results in significant results.

Data Splitting

Divide the dataset into training and testing sets to train the machine learning models and evaluate their performance.
Splitting the dataset into 80 : 20 ratio.
80% of data is taken for training set.
20% of data is taken for testing set.

Convolutional Neural Network

CNNs are widely employed in medical classification tasks, serving as efficient feature extractors that eliminate the need for complex feature engineering in analysing medical images. Their effectiveness lies in their ability to achieve high accuracy, especially with large datasets, making them invaluable for image recognition, classification, and computer vision tasks. Unlike traditional methods, CNNs learn object features directly through successive iterations, removing the necessity for manual feature extraction. This capability, along with the ability to retrain for new tasks and build on preexisting networks, expands the practical applications of CNNs without escalating computational complexities or expenses.

The model begins with Conv2D layers, employing adaptable filters to extract features from the image. MaxPooling2D layers follow, down sampling by selecting maximum values from neighbouring pixels to reduce complexity. Dropout regularization randomly deactivates nodes during training, combating overfitting. The 'relu' activation function introduces non-linearity, aiding in feature learning. The Flatten layer reshapes the output for fully connected layers, consolidating features. Two Dense layers handle classification, with the final layer using 'softmax' activation for class probability distribution. This architecture effectively captures both local and global features for accurate classification.

Model Architecture :

Convolutional Neural Network model architecture for Pixel Based Multi Class Skin Cancer Classification :

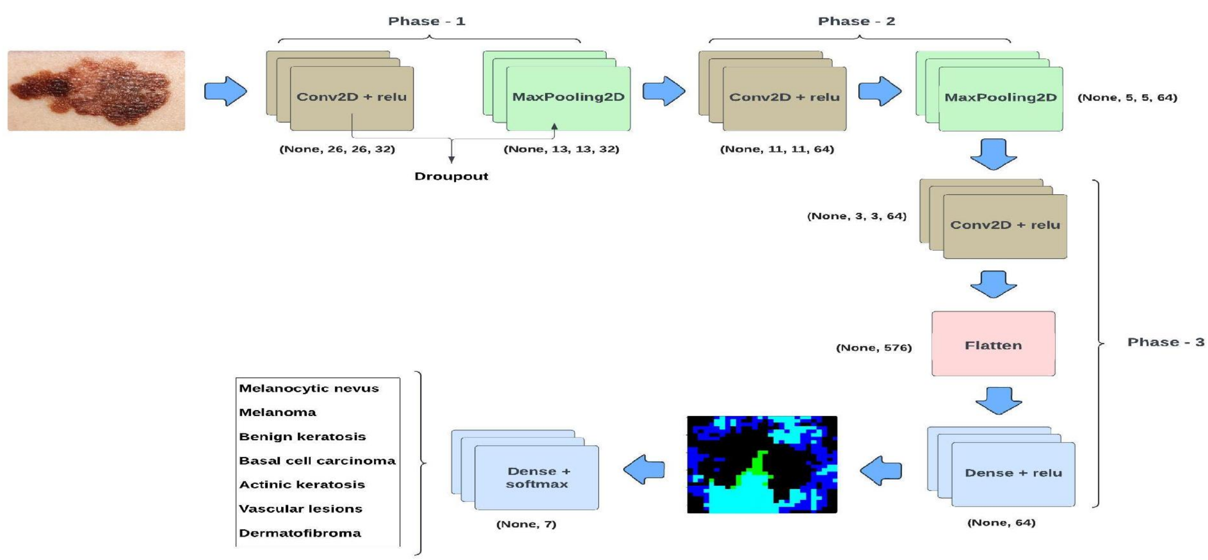

Figure 4 CNN Architecture

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, April 2024

VIII. RESULTS AND DISCUSSION

Graphical Representation of Accuracy and Loss

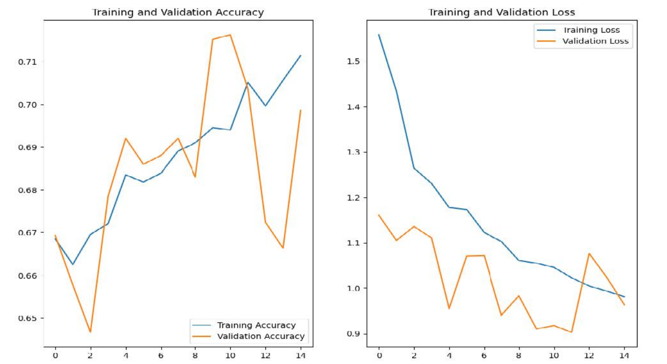

Figure 5 Accuracy and Loss curves.

The model initially achieves a training accuracy of 0.65, which increases to approximately 0.71 after 14 epochs, showing improved cancer image classification with more training data. Validation accuracy begins around 0.66 and rises to about 0.69, indicating better classification on unseen data. However, the validation accuracy consistently trails behind the training accuracy, hinting at possible overfitting. Training loss decreases from 1.4 to 1.1, reflecting improved model performance, while validation loss decreases from 1.35 to 1.2, suggesting better performance on unseen data. Nevertheless, the consistent gap between validation and training metrics suggests potential overfitting.

Confusion Matrix

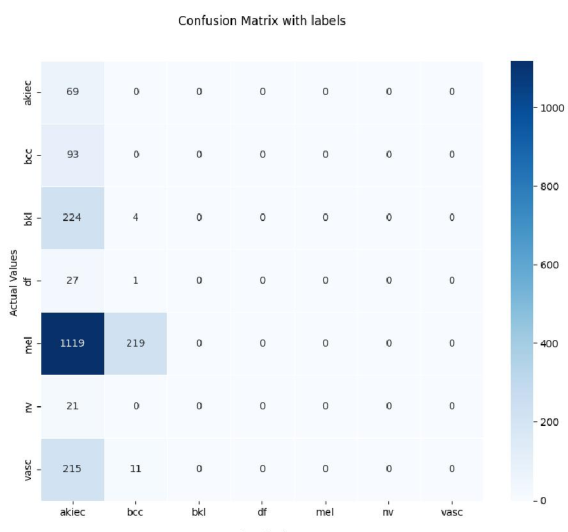

Figure 6 Confusion Matrix

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, April 2024

Classification Report

Accuracy 0.70

	precision	recall	f1-score	support
0	0.42	0.36	0.39	69
1	0.43	0.38	0.40	93
2	0.43	0.46	0.45	228
3	0.33	0.04	0.06	28
4	0.80	0.91	0.85	1338
5	0.18	0.90	0.30	21
6	0.00	0.00	0.00	226

Predictions

This is the user interface where the user upload respective skin lesion image in any format, The system interface processes the input lesion image and classify the type of cancer.

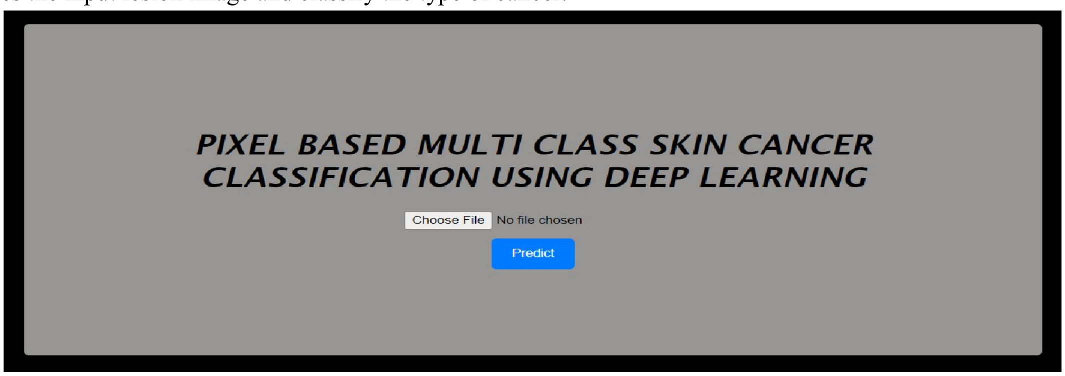

Figure 7 Webpage.

Basal cell carcinoma

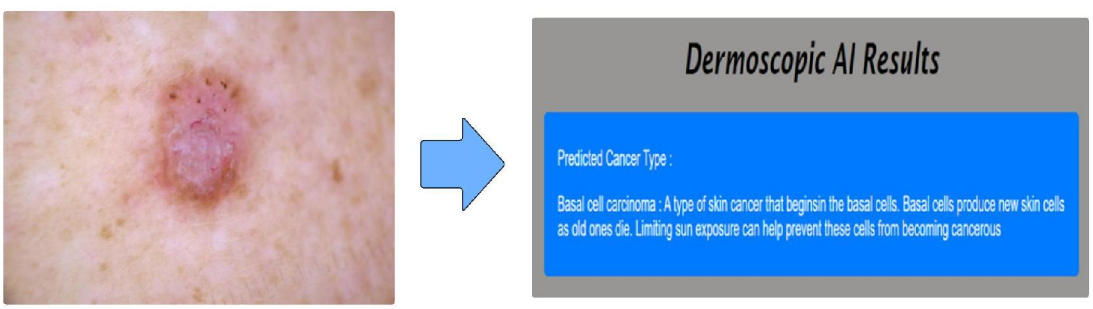

Figure 8Basal cell carcinoma.

IX. CONCLUSION

In summary, the project achieved the development of a skin cancer classification system utilizing Convolutional Neural Network (CNN) technology, incorporating debiasing methods to ensure fairness across demographic groups. Achieving an accuracy of 71%, CNN showcased significant potential in accurately categorizing skin lesions while addressing biases. These findings emphasize the significance of fairness considerations in healthcare AI systems to ensure equitable outcomes across various populations, demonstrating the capacity of CNN-based strategies to enhance diagnostic precision and fairness in skin cancer detection.

REFERENCES

[1] Naqvi, M.; Gilani, S.Q.; Syed, T.; Marques, O.; Kim, H.-C. Skin Cancer Detection Using Deep Learning—A Review. Diagnostics 2023, 13, 1911. https://doi.org/ 10.3390/diagnostics13111911

[2] Mishra S, Imaizumi H, Yamasaki T. Interpreting Fine-Grained Dermatological Classification by Deep Learning. *Proc IEEE/CVF Conf Comput Vision Pattern Recognit Workshops* (2019), 0–0. doi: 10.1109/CVPRW.2019.00331

[3] Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks. *Nature* (2017) 542:115–8. doi: 10.1038/nature21056

[4] Lopez AR, Giro-i Nieto X, Burdick J, Marques O. Skin Lesion Classification from Dermoscopic Images Using Deep Learning Techniques, in: 13th IASTED International Conference on Biomedical Engineering (BioMed), Manhattan, New York, U.S: Institute of Electrical and Electronics Engineers (IEEE) (2017). pp. 49–54.

[5] Walker B, Rehg J, Kalra A, Winters R, Drews P, Dascalu J, et al. Dermoscopy Diagnosis of Cancerous Lesions Utilizing Dual Deep Learning Algorithms via Visual and Audio (Sonification) Outputs: Laboratory and Prospective Observational Studies. *EBioMedicine* (2019) 40:176–83. doi: 10.1016/j.ebiom.2019.01.028

[6] Perez F, Avila S, Valle E. Solo or Ensemble? Choosing a CNN Architecture for Melanoma Classification. *Proc IEEE/CVF Conf Comput Vision Pattern Recognit Workshops* (2019), 0–0. doi: 10.1109/CVPRW.2019.00336

[7] Polat K, Koc KO. Detection of Skin Diseases From Dermoscopy Image Using the Combination of Convolutional Neural Network and One-Versus-All. *J Artif Intell Syst* (2020) 2:80–97. doi: 10.33969/AIS.2020.21006

[8] Rahman Z, Hossain MS, Islam MR, Hasan MM, Hridhee RA. An Approach for Multiclass Skin Lesion Classification Based on Ensemble Learning. *Inf Med Unlocked* (2021) 25:100659. doi: 10.1016/j.imu.2021.100659

[9] Kawahara J, Daneshvar S, Argenziano G, Hamarneh G. Seven-Point Checklist and Skin Lesion Classification Using Multitask Multimodal Neural Nets. *IEEE J Biomed Health Inf* (2019) 23:538–46. doi: 10.1109/JBHI.2018.2824327

[10] Hameed N, Shabut A, Hameed F, Cirstea S, Harriet S, Hossain A. Mobile Based Skin Lesions Classification Using Convolution Neural Network. *Ann Emerging Technol Comput (AETiC)* (2020) 4:26–37. doi: 10.33166/AETiC.2020.02.003