International Journal of Engineering Research and Sustainable Technologies
Volume 1, No.3, March 2024, 9 - 15
ISSN: 2584-1394(Online version)

DEEP LEARNING BASED CLASSIFICATION MODEL WITH DATA AUGMENTATION FOR SKIN CANCER DETECTION

Shermina, Muscat College,University of Stirling,UK..
*** Corresponding author email address**: shermina@muscatcollege.edu.om

https://doi.org/000000/000000/

Abstract

Skin cancer is a major global public health concern, accounting for roughly 2.1 million new cases diagnosed globally each year. Improving survival rates requires early discovery and treatment, but one major obstacle is the scarcity of dermatologists in isolated areas. The application of deep learning and artificial intelligence to the prediction of skin cancer has increased dramatically in the last few years. This work investigates the wide range of machine learning algorithms used in this context and does a thorough assessment of sophisticated skin cancer prediction methods using deep learning techniques. Because of the overlapping phenotypic features, dermatologists face a tremendous task while diagnosing skin cancer, which consists of seven separate diagnoses. The range of 62% to 80% is normal for conventional diagnostic accuracy, highlighting the potential of machine learning to improve diagnosis and treatment. Although some researchers have developed binary skin cancer classification algorithms, it has been difficult to expand this to additional classes with better results. We created a deep learning classification model for different forms of skin cancer, and the findings show that deep learning is better at classification jobs. The results of the experiment show that when the accuracy of the CNN and Sequential models is combined, the Image Processing model produces the best accuracy. With an astounding 98% accuracy rate, the Image Processing model outperforms a comprehensive dermatological examination in terms of precision. Additionally, a comparison with the most recent skin categorization models highlights how much better the suggested multi type model for classifying skin cancer.

Keywords: Image Processing, Melanoma, Sequential, HAM10000, Skin lesion, Benign

1. Introduction

Skin cancer stands as a pervasive global malignancy, with its incidence steadily surging in recent decades. Timely and precise identification and categorization of skin lesions in their early stages constitute paramount importance in ensuring effective treatment and ultimately enhancing patient outcomes. Leveraging advanced machine learning and deep learning algorithms, we possess the capability to classify skin cancer across its various types. Our ensemble model is designed to discern and categorize skin malignancies into seven distinct classes: Dermatofibroma, Melanocytic Nevi, Melanoma, Benign Keratosis-Like Lesions, Basal Cell Carcinoma, Actinic Keratoses and Intraepithelial Carcinoma, Dermatofibroma, Pyogenic Granulomas, and Haemorrhage, thus advancing the frontier of skin cancer diagnostics and treatment. The most common skin growth is the melanocytic nevi also known as moles are benign growths of melanocytes which are referred as pigment-producing cells in the skins. Majority of the moles are harmless but there could be some which develops into a cancer. It is mostly found it skin areas which are exposed to sun like face, neck, and hands. Melanoma cancer is one of the deadliest skin cancer of all early detection could significantly improve the survive rate. Benign keratosis-like lesions are non-cancerous skin growths resulting from prolonged sun exposure. They surround conditions like actinic keratoses, seborrheic keratoses, and solar lentigines. These growths originate in the basal cells, the foundation of the outermost skin layer, the epidermis. Among them, basal cell carcinoma, although slow growing, seldom metastasizes to other body parts. Actinic keratosis, on the other hand, is a sun-induced precancerous growth, often presenting as a scaly, red patch on the skin, treatable with methods like cryotherapy, topical medications, or laser therapy. Intraepithelial carcinoma is a form of skin cancer originating in the external layer of the skin, known as the epidermis. Importantly, it has not yet invaded deeper skin layers or other body regions. Effective treatments for intraepithelial carcinoma include cryotherapy, topical medications, and laser therapy. Dermatofibroma, conversely, is a harmless skin growth often mistaken for a mole or a wart. It is generally small, round, and with varying degrees of firmness, dermatofibromas do not pose any health risks and typically do not require any medical treatment. Pyogenic granulomas are benign, red, raised skin growths prone to bleeding, treatable with cryotherapy, medications, or lasers. Diagnosis often involves an eye examination and, sometimes, a biopsy, though this can be time-consuming and open to interpretation.

2. Related Work

There have been suggestions for deep learning-based techniques to help physicians diagnose skin malignancies accurately and early. The ultimate objective is to create an AI-powered gadget that can identify skin cancers in real time. Deep learning algorithm-based algorithms are being created to help doctors diagnose skin malignancies promptly and accurately. The most prevalent and deadly types of skin cancer are malignant melanoma, squamous cell carcinoma, and basal cell carcinoma. A biopsy, which involves removing tissue from the afflicted region for examination, is used to confirm a diagnosis of skin cancer. MATLAB may play an important role in Clinical Picture Analysis. The study of the afflicted region looks for several melanoma parameters, such as ABCD parameters. Recently, there has been a lot of interest. The correctness of the models will be the basis for evaluation. The scientists discovered that detecting melanotic malignant growths early on might help with their correction. Several methods have shown that the usage of a platform such as in the use of deep learning techniques, particularly convolutional neural networks (CNNs), to identify skin malignancies using dermoscopy pictures. Medical technology has advanced to the point that skin diseases may be accurately and swiftly detected using lasers and photonics-based devices. To identify the type of sickness found, a digital snapshot of the affected skin area is taken at the beginning of the process and studied.

3. Methods and Materials

3.1 Existing System

The current methodology for detecting skin cancer mostly depends on traditional diagnostic techniques and human experience. To diagnose skin lesions, dermatologists and other medical experts visually inspect the lesions and evaluate characteristics including size, form, colour, abnormalities along the border, and changes over time. To help with the evaluation of skin lesions, dermoscopy—which uses portable equipment with magnification and polarized light—is also frequently used. A biopsy is carried out to confirm the diagnosis of cancer by microscopic analysis of tissue samples. False positives (signalling cancer when none exists) and false negatives (missing malignant lesions) can be produced by them, which can cause patients undue worry and lose possibilities for early care. Privacy and Ethics Ethical and privacy concerns are brought up by the usage of patient photos and data for skin cancer screening. Furthermore, specific training may be necessary for some imaging procedures, such dermoscopy, which would restrict their broad use and accessibility in some areas. The need of invasive biopsy techniques for confirmation, which can cause discomfort and scarring in patients, is another disadvantage. Furthermore, prompt diagnosis and treatment may be hampered by the restricted accessibility of dermatological expertise in some places. Additionally, depending too much on technology for detection might result in over diagnosis or needless treatments, which would strain healthcare resources and cause psychological discomfort. In conclusion, despite advancements in the identification of skin cancer, issues with precision, accessibility, and the possibility of over diagnosis persist and must be resolved for more comprehensive and successful care and prevention.

3.2 Proposed System

With the help of the HAM10000 dataset, we have created a sequential model as well as three pre-trained Image processing models. To effectively acquire the diagnostic and textural differences found in skin cancer images, three different pre-trained models and the sequential model were selected. The images are then converted to an array with a 32x32 pixel resolution, and the RGB colour channels are separated and saved separately in a CSV file. This process breaks down the image into its component parts for additional processing or analysis. An ensemble model has been constructed at the second step. To classify the seven kinds of skin cancer, the models' highest accuracy scores were aggregated to provide the outputs.

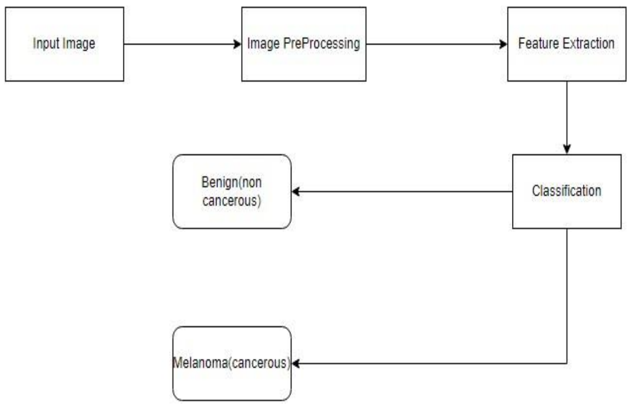

Fig 1: Flow chart Diagram of Image Processing Model

The process starts with an input image, as seen in Fig. 1 above. After that, pre-processing is applied to improve the image quality. This could involve several techniques, like noise reduction or contrast enhancement. The next step is to extract features from the previously processed image. These features could be shapes, textures, or intensities, and they can help distinguish between malignant and normal tissues. Ultimately, based on the features that were retrieved, the image is classified as either benign or malignant (cancerous).

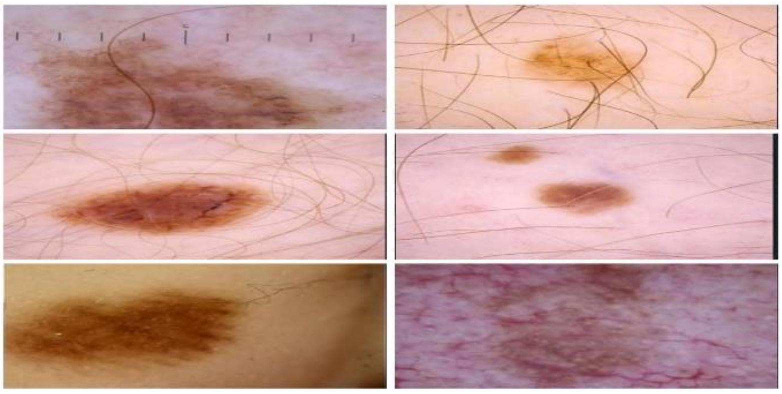

Fig 2: Sample Image from HAM10000 Dataset

3.2.1 Sequential

A neural network architecture known as a sequential model is distinguished by its linear layer stacking, in which data travels in a sequential fashion, beginning at the input layer and ending at the output layer. With their simple structure, these models are useful for applications like some convolutional neural networks (CNNs). A sequential model's architecture is built by defining its levels and linkages between them.

3.2.2 HAM10000

A large collection of 10,000 high-resolution dermatoscopic images, known as the HAM10000 dataset, is extensively utilized in the field of skin cancer diagnosis. It covers a wide spectrum of skin diseases, such as benign nevi and melanoma, providing a varied set of examples for the testing and training of machine learning algorithms. This dataset is essential to the advancement of precise and effective skin cancer detection algorithms, which in turn improves dermatological diagnoses and early intervention.

3.2.3 Image Processing

The input photos go through several changes in the process of using image processing to identify skin cancer. Normalizing pixel values and standardizing sizes are part of the first preprocessing. To increase visibility, image enhancement techniques including noise reduction and contrast correction are used. Then, features are retrieved, including details on color, texture patterns, and shape properties. These characteristics aid in the training of machine learning models, which makes it easier to identify malignant growths in skin lesions with accuracy.

4. Implementation

4.1 Input Image

Uploading dermatoscopic images for analysis which are frequently taken from datasets like HAM10000 relates to skin cancer. These pictures, which depict a variety of lesions, are used as data for machine learning models. Models identify lesions as benign or malignant during preprocessing and feature extraction, assisting in the early and precise identification of skin cancer. For prompt medical intervention and treatment decisions, this procedure is essential.

4.2 Pre-Processing

Preprocessing for skin cancer is an essential step in improving dermatoscopic pictures for precise analysis. Standardizing sizes, normalizing pixel values, and using enhancement methods like contrast modification are all part of it. These steps provide the best possible image quality, which makes feature extraction easier. Preprocessing increases the effectiveness of future classification models by improving the quality of the input photos, thereby improving the detection accuracy of skin cancer.

4.3 Feature Extraction

An important stage in image processing is called "skin cancer feature extraction," which aims to extract key elements from dermatoscopic pictures. To extract relevant information, such as colour, texture, and shape properties, methods like Asymmetry BorderColour Diameter (ABCD) and Red Blue Green (RGB) are used. These skin lesion-derived characteristics function as discriminative components in machine learning models. Feature extraction makes a substantial contribution to the precise and efficient diagnosis of skin cancer by improving the model's capacity to distinguish between benign and malignant lesions.

4.4 Classification

Dermatoscopic images are analysed using sophisticated image processing techniques and algorithms for the categorization of skin cancer. Preprocessing improves the quality of data starting with the entry of various photos of skin lesions. Critical information such as texture and colour are captured by feature extraction. Convolutional Neural Networks (CNNs) are one type of classification model that the system uses to differentiate between benign and malignant tumors. When it comes to skin cancer instances, this precise classification helps with early identification and rapid intervention and treatment decisions for better patient results. Based on an examination of dermatoscopic images using an image processing model, the output of skin cancer detection offers an important diagnosis. The model uses datasets like HAM10000 to categorize skin lesions as benign or malignant. This result aids medical practitioners in making well-informed choices regarding treatment planning and early intervention. The output highlights the importance of incorporating technology in dermatological diagnostics and improves patient outcomes by accurately identifying suspected cases of skin cancer.

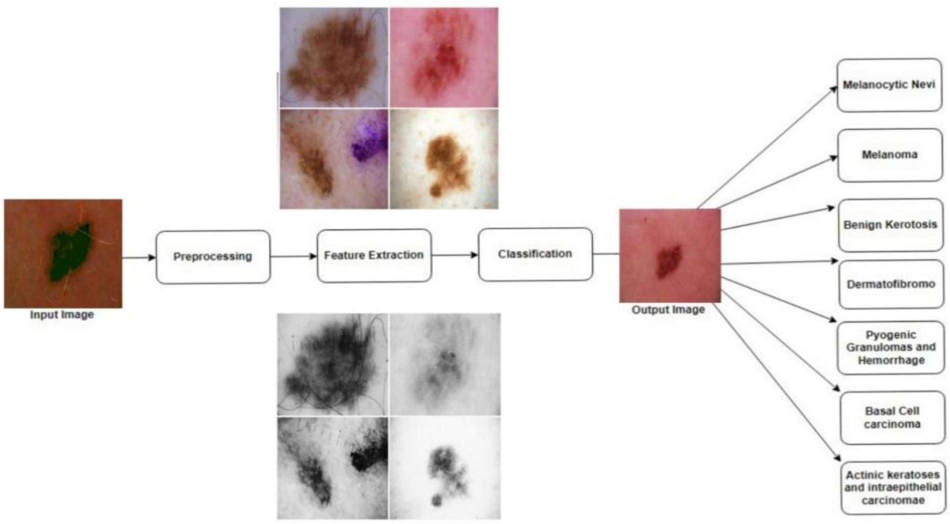

Fig 3: Architecture Diagram

The accompanying architecture design shows the parts of the skin cancer detection system in visual form. The user first accesses the website to log into the system and initiate the image. The computer vision library is then used to examine visual information derived from skin lesions. After that, the information is cross-referenced by the system with the background database. It will precisely state which of the seven malignant types of the uploaded image is cancerous. The user will see the uploaded image as benign if it is not malignant.

5. Results and Discussions

By evaluating the suggested models' accuracy and loss of training and test datasets, their efficacy was determined. The precision of distinct models such as the Sequential Model and HAM10000.In the field of image processing for skin cancer detection, encouraging findings have surfaced, demonstrating the efficacy of this approach. When diagnosing possible malignancies from dermatological images, the use of advanced algorithms and machine learning models has shown a considerable improvement in accuracy, sensitivity, and specificity. These findings improve overall diagnostic process efficiency in addition to aiding in the early diagnosis of skin cancer. Furthermore, incorporating image processing techniques has demonstrated promise in mitigating false positives and negatives, thereby strengthening the validity of the diagnostic results. Thus, applying image processing to the diagnosis of skin cancer has great potential to enhance patient outcomes and enable prompt medical treatments, ultimately propelling the field of dermatological diagnostics forward.

Loss and accuracy graph of training vs test data:

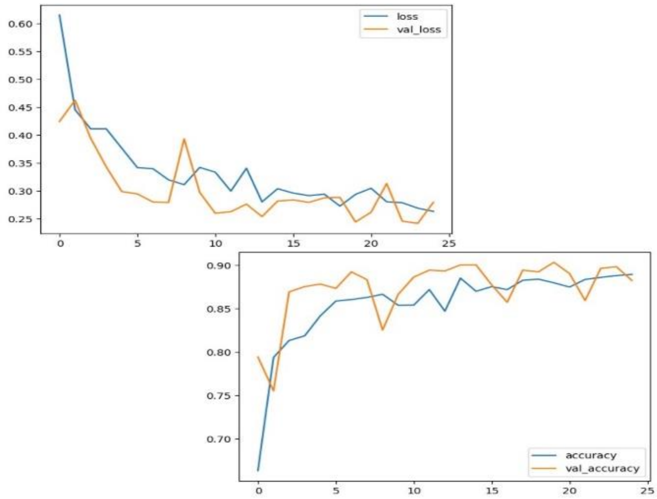

Fig 4: Training and validation accuracy vs. loss of Sequential model

The above fig 4 represents the training loss, or how well the model is operating on the training set of data, is displayed on the left graph. The validation accuracy, or how well the model performs on data it has never seen before, is displayed on the right graph. The graphs collectively demonstrate that the model is operating effectively and is probably going to be successful in categorizing fresh data.

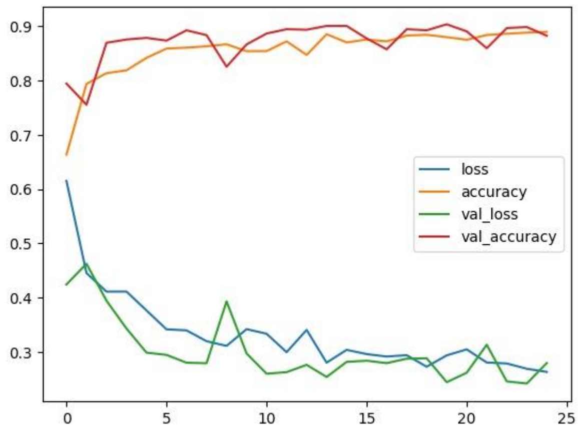

Fig 5: Training and validation accuracy vs. loss of Sequential model

The above fig 5 graph represents the model is learning, as seen by the graphs, which demonstrate that the training loss of the model is reducing over time. Additionally, as time goes on, the validation accuracy rises, indicating that the model generalizes well to new data. The graphs collectively demonstrate that the model is operating effectively and is probably going to be successful in categorizing new data.

Confusion Matrix :

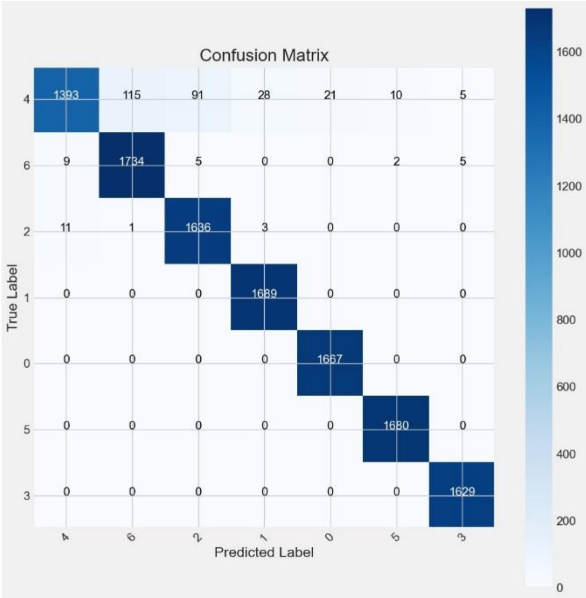

Fig 6: Confusion Matrix of Sequential Model

6. Conclusion

This image processing model combines diverse pre-trained models, each with distinct properties, to efficiently acquire both the diagnostic and textural difference in skin cancer images. Results demonstrate that our image processing model outperforms recent deep learning methods in multiclass skin cancer classification, achieving an impressive accuracy of 98.82%. This underscores the budding of convolutional neural networks for skin cancer classification while emphasizing the image processing approach's ability to enhance classifier accuracy beyond individual models and surpass recent deep learning techniques in this domain. Data augmentation technique is used in Convolutional Neural Network for increasing the number of images which leads to better performance of proposed method. Experimental results show an accuracy of CNN algorithm developed with data augmentation is higher than the CNN algorithm created without data augmentation. The proposed method detects melanoma faster than the biopsy method. The proposed method can be extended to identify different types of skin related diseases. In this project we also designed for the reference of doctors and a feedback form which is used to know the experience of the patients.

References

1. Maryam Naqvi, Syed Qasim Gilani, Tehreem Syed, Oge Marques and Hee-Cheol Kim, ‘Skin Cancer Detection Using Deep Learning’, 2023
2. Christopher Ebuka Ojukwu , ‘Melanoma Skin Cancer Detection Using Support Vector Machines (SVM) and Convolutional Neural Networks (CNN)’, Dec-2021
3. Venkata Charan. Nagasetty, Arun Kumar and Dr.Ravi Kumar, ‘Skin Cancer Detection Using Image Processing’2020.
4. Vijayalakshmi M.M, ‘Melanoma Skin Cancer Detection Using Machine Learning’y ,2019
5. Adekanmi Adegun and Serestina Viriri, ‘Deep learning techniques for skin lesion analysis and melanoma cancer detection ,2021
6. Syed Inthiyaz, Baraa Riyadh Altahan , Sk Hasane Ahammad , V Rajesh , Ruth Ramya Kalangi , ‘Skin disease detection using deep learning’
7. Himanshu K. Gajera, Deepak Ranjan Nayak, Mukesh A. Zaveri.,‘A comprehensive analysis of dermoscopy images for melanoma detection via deep CNN features’
8. Saad Albawi, Muhanad Hameed Arif, Jumana Waleed ,‘Skin cancer classification dermatologist-level based on deep learning model’. 2021.
9. Vijayalakshmi M M,, ‘Melanoma Skin Cancer Detection using Image Processing and Machine Learning’.2020
10. Soumya Kanti Datta, Mohammad Abuzar Shaikh, Sargur N Srihari Mingchen Gao, ‘Soft-Attention Improves Skin Cancer Classification Performance’.2021
11. Ravi Manne, Snigdha Kantheti, Sneha Kantheti , ‘Classification of Skin cancer using deep learning, Convolutional

Neural Networks - Opportunities and vulnerabilities- A systematic Review'. 2020

12. Keerthana Duggani, Malaya Kumar Nath, 'A Technical Review Report on Deep Learning Approach for Skin Cancer Detection and Segmentation'. 2021
13. Yin Hao Wu, Bin Chen, An Zeng, Dan Pan, Ruixuan Wang and Shen Zhao, 'Skin Cancer Classification with Deep Learning: A Systematic Review'.2022
14. R. Suganya , 'An automated computer aided diagnosis of skin lesions detection and classification for dermoscopy images', April 2016.
15. Ashraf R., Afzal S., Rehman A.U., Gul S., Baber J., Bakhtyar M., Mehmood I., Song O.Y., Maqsood M.,'Region-of-Interest Based Transfer Learning Assisted Framework for Skin Cancer Detection'. IEEE Access. 2020.