Melonoma Detection: Enhancing Precision in Skin Cancer Detection using Deep Learning

Valli Meenaa Vellaiyan
Northeastern University
Boston, United States
vellaiyan.v@northeastern.edu

Hrithik Sarda
Northeastern University
Boston, United States
sarda.h@northeastern.edu

Mohit Chodisetti
Northeastern University
Boston, United States
chodisetti.m@northeastern.edu

***Abstract*—Skin cancer, including melanoma, presents a growing health challenge globally, with its incidence on the rise. Timely detection is critical for effective treatment, particularly given the lower survival rates associated with advanced-stage melanoma. This study proposes a novel deep learning framework for melanoma detection that integrates patient-level contextual information to enhance diagnostic accuracy. Leveraging a dataset of 33,126 dermoscopic images, our model incorporates patient demographics, lesion characteristics, and clinical history. Through extensive preprocessing, data augmentation, and exploratory data analysis, the framework aims to improve diagnostic outcomes and provide insights for personalized treatment plans. The evaluation of three models, including a Convolutional Neural Network (CNN), a Deep Convolutional Neural Network (DCNN), and an Inception model, reveals the Inception model as the most promising, achieving a precision of 0.947, recall of 0.947, and accuracy of 0.988. This research underscores the potential of deep learning in enhancing melanoma diagnosis and emphasizes the importance of considering both computational efficiency and performance metrics in model selection for real-world applications.**

Keywords—Melanoma, skin cancer, deep learning, artificial intelligence, convolutional neural networks, medical imaging, diagnostic accuracy, patient-level information, personalized treatment, Inception model, DCNN, CNN, precision, recall, accuracy.

I. INTRODUCTION

Skin cancer, particularly melanoma, poses an important health challenge worldwide, with its prevalence steadily rising over the past decades. Melanoma, while accounting for only a small portion of skin cancer cases, is responsible for a unbalanced number of skin cancer-related deaths due to its aggressive nature. Early detection is crucial for successful treatment, as the five year survival rate drops pointedly for advanced-stage melanoma cases [1].

Recent developments in artificial intelligence (AI) and deep learning have shown promising results in various medical imaging tasks, including the detection and classification of skin lesions. Convolutional Neural Networks (CNNs) in particular have established remarkable capabilities in analyzing dermoscopic images, that are crucial in melanoma diagnosis. Nonetheless, existing AI approaches often overlook critical patient level contextual information, that is essential for dermatologists for clinical practice [2].

This research aims to address this gap by proposing a novel deep learning framework for melanoma detection which incorporates patient level contextual information. By integrating patient demographics, lesion characteristics, as well as clinical history, our model seeks to enhance the accuracy and precision of melanoma diagnosis. This approach not only aims to improve diagnostic outcomes but also to provide valuable insights for clinicians in developing personalized treatment plans.

In this paper, we present our methodology for developing a deep convolutional neural network (DCNN) model tailored to the complexities of the dermoscopic images. We leverage an enormous dataset of dermoscopic images from the SIIM-ISIC Melanoma Classification, comprising both benign and malignant lesions. Furthermore, we make use of advanced image processing techniques like feature extraction and augmentation to enhance the model's ability to determine subtle patterns which are indicative of melanoma.

The remainder of this paper is organized as follows: The background section provides context on melanoma and the importance of early detection, highlighting the role of AI and deep learning. The approaches section details our methodology, including the dataset used, model architecture, and training approach. The results section presents our experimental findings and performance evaluation. The conclusion and future work section summarizes our findings, discusses their implications, and suggests avenues for further research. Finally, the references section lists the sources cited in the paper.

II. BACKGROUND

The utilization of deep learning techniques, particularly convolutional neural networks (CNNs), has significantly advanced the field of medical image classification. With the increasing availability of medical imaging data, researchers have explored various methodologies to enhance the accuracy and efficiency of disease diagnosis. Traditionally, medical image classification relied on manual feature extraction and selection, which was labor-intensive and limited in its performance. However, the emergence of CNN-based algorithms has revolutionized this process by enabling automatic feature learning from raw image data.

Several studies have demonstrated the effectiveness of CNNs in medical image classification tasks, particularly in the context of diseases such as pneumonia detected from chest X-ray images. For instance, Paper 1 investigates the application of CNN-based algorithms on a small chest X-ray dataset to classify pneumonia, comparing the performance of different techniques such as transfer learning and capsule networks. This research underscores the importance of leveraging advanced machine learning methods, like CNNs, to achieve superior classification accuracy, especially when dealing with limited datasets [3].

Similarly, in the domain of dermatology, deep learning techniques have been employed for early detection and classification of skin cancer. Paper 2 provides a systematic review of various neural network approaches for skin cancer detection, highlighting the superiority of CNNs due to their ability to effectively analyze image data. This review

emphasizes the significance of selecting appropriate classification techniques for accurate diagnosis, considering factors such as lesion parameters and image features [4].

Moreover, the exploration of innovative data augmentation strategies has further enhanced the performance of classification algorithms. Paper 3 delves into different data augmentation techniques, including traditional transformations and Generative Adversarial Networks (GANs), to improve classification accuracy. The research suggests the potential of neural augmentation, where neural networks learn augmentations that optimize classifier performance, thus presenting avenues for further refinement and experimentation in the field of medical image classification [5].

In summary, the convergence of deep learning methodologies, such as CNNs, with innovative approaches to data augmentation has propelled advancements in medical image classification. By leveraging these techniques, researchers aim to enhance the accuracy and efficiency of disease diagnosis, ultimately contributing to improved clinical outcomes and patient care.

III. APPROACHES

In this project, initially we perform preprocessing steps on the dataset, like handling missing values, class imbalance, and data augmentation. Missing values are addressed through imputation techniques, ensuring that no crucial information is lost. Class imbalance, particularly prevalent in medical datasets, is mitigated by introducing additional malignant images from external sources, thus creating a more balanced training set. Moreover, data augmentation techniques are applied to further enhance the dataset's diversity, which is crucial for training robust machine learning models.

The dataset [6] used comprises 33,126 dermoscopic images of unique benign and malignant skin lesions from over 2,000 patients. Each image includes patient identifiers and features such as patient_id, sex, age_approx, and diagnosis. Malignant diagnoses are confirmed via histopathology, while benign diagnoses are validated using expert agreement or longitudinal follow-up. The dataset, sourced from reputable institutions worldwide, is publicly accessible in JPEG format, with accompanying metadata.

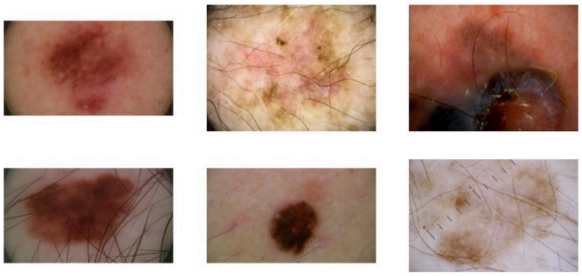

Fig. 1. a. Images of Class 1 skin lesions

Fig. 1. b. Images of Class 0 skin lesions

In addition to the SIIM-ISIC Melanoma Classification Challenge dataset, we are also utilizing an external dataset [7] to address class imbalance. This external dataset contains malignant images only, which we are incorporating into our training data. Prior to augmentation, the original dataset exhibited a class distribution of 581 malignant images and 32,079 benign images. After integrating the external dataset, the class distribution is adjusted to 5,025 malignant images and 32,079 benign images.

Following preprocessing, extensive exploratory data analysis (EDA) is conducted to gain insights into the dataset. Various visualizations, such as bar charts, distribution plots, and sunburst charts, are utilized to examine the distribution of target variables across different demographic and anatomical features. Insights derived from EDA include patterns in scan results based on age, gender, and anatomical site, as well as the distribution of diagnoses within the dataset. Overall, the methodology encompasses both preprocessing steps and exploratory analyses essential for building a reliable predictive model for diagnosing skin lesions.

A. Model 1

Here we are using the convolutional neural network (CNN) model for classifying skin lesion images into benign and malignant categories. The model architecture consists of three convolutional layers with 16, 32, and 64 filters respectively, each followed by max-pooling layers to downsample the feature maps. The ReLU activation function is applied to introduce non-linearity in the network. The flattened output is then passed through two fully connected layers with 256 and 1 neuron(s) respectively, with the final layer utilizing a sigmoid activation function to output the probability of malignancy. The dataset is preprocessed by resizing the images to 128x128 pixels and normalized to [0,1] range. Furthermore, the dataset is split into training (70%), validation (20%), and testing (10%) sets to facilitate model evaluation and hyperparameter tuning.

Layer (type)	Output Shape	Param #
conv2d	(None, 126, 126, 16)	448
max_pooling2d	(None, 63, 63, 16)	0
conv2d_1	(None, 61, 61, 32)	4640
max_pooling2d_1	(None, 30, 30, 32)	0
conv2d_2	(None, 28, 28, 64)	18496
max_pooling2d_2	(None, 14, 14, 64)	0
flatten	(None, 12544)	0
dense	(None, 256)	3211520
dense_1	(None, 1)	257

Fig. 2. Model 1 Summary

B. Model 2

This model presents a deep convolutional neural network (DCNN) architecture tailored for skin lesion classification,

comprising several layers. The model architecture consists of a stack of convolutional layers with varying filter sizes, each followed by batch normalization and max pooling layers to extract hierarchical features from the input images. The DCNN architecture culminates in a dense layer with 128 neurons activated by ReLU, facilitating feature aggregation and dimensionality reduction. Finally, a dense layer with a single neuron and a sigmoid activation function is employed for binary classification. The total number of trainable parameters in the model amounts to 1,197,889, with a total of 1,199,745 parameters. This design choice is, essential for processing large-scale dermatological image datasets.

Layer (type)	Output Shape	Param #
Conv2D	(None, 128, 128, 32)	896
BatchNormalization	(None, 128, 128, 32)	128
MaxPooling2D	(None, 64, 64, 32)	0
Conv2D	(None, 64, 64, 64)	18,496
BatchNormalization	(None, 64, 64, 64)	256
MaxPooling2D	(None, 32, 32, 64)	0
Conv2D	(None, 32, 32, 64)	36,928
BatchNormalization	(None, 32, 32, 64)	256
MaxPooling2D	(None, 16, 16, 64)	0
Conv2D	(None, 16, 16, 128)	73,856
BatchNormalization	(None, 16, 16, 128)	512
MaxPooling2D	(None, 8, 8, 128)	0
Conv2D	(None, 8, 8, 128)	147,584
BatchNormalization	(None, 8, 8, 128)	512
MaxPooling2D	(None, 4, 4, 128)	0
Conv2D	(None, 4, 4, 256)	295,168
BatchNormalization	(None, 4, 4, 256)	1,024
MaxPooling2D	(None, 2, 2, 256)	0
Conv2D	(None, 2, 2, 256)	590,080
BatchNormalization	(None, 2, 2, 256)	1,024
MaxPooling2D	(None, 1, 1, 256)	0
Flatten	(None, 256)	0
Dense	(None, 128)	32,896
Dense	(None, 1)	129

Fig. 3. Model 2 Summary

C. Model 3

The methodology employed in this research involved the construction of a convolutional neural network (CNN) architecture consisting of several layers. The input layer (input_5) accepted images of size 128x128x3. Subsequently, a series of convolutional layers were applied, including conv2d_379, conv2d_380, and conv2d_381, each followed by batch normalization layers (batch_normalization_376, batch_normalization_377, batch_normalization_378) and activation functions (activation_376, activation_377). These convolutional layers utilized filter sizes of 32 and 64, with corresponding output shapes of (63, 63, 32) and (61, 61, 64) respectively. The CNN architecture was designed to extract hierarchical features from the input images, facilitating the subsequent tasks such as classification or segmentation.

Layer (type)	Output Shape	Param #
input_5 (InputLayer)	(None, 128, 128, 3)	0
conv2d_379 (Conv2D)	(None, 63, 63, 32)	864
batch_normalization_376	(None, 63, 63, 32)	96
activation_376	(None, 63, 63, 32)	0
conv2d_380 (Conv2D)	(None, 61, 61, 32)	9216
batch_normalization_377	(None, 61, 61, 32)	96
activation_377	(None, 61, 61, 32)	0
conv2d_381 (Conv2D)	(None, 61, 61, 64)	18432
batch_normalization_378	(None, 61, 61, 64)	192
Contd.	Contd.	Contd.

Fig. 4. Model 3 Summary

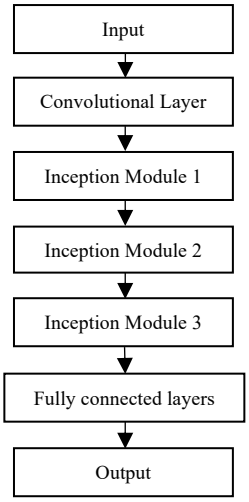

Fig. 5. Flowchart of Inception Model Architecture Overview

IV. RESULTS

A. Training results

a. Model 1: CNN with data augmentation, without class weights

Slightly higher validation loss compared to training loss in loss versus epoch graph, indicating potential overfitting. Nonetheless, accuracy versus epoch graph showcases steady accuracy enhancement throughout training.

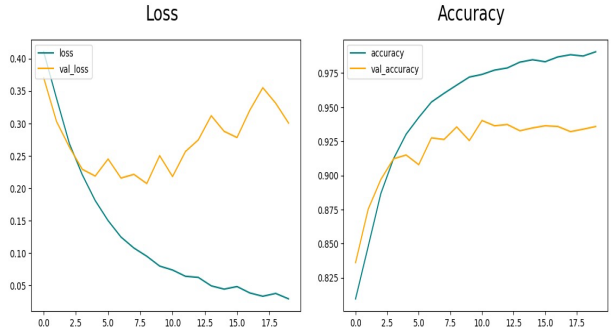

Fig. 6. Train Loss vs Validation Loss & Train Accuracy vs Validation Accuracy for Model 1

b. Model 2: DCNN without data augmentation, with class weights

Smooth convergence observed in loss versus epoch graph with minimal gap between training and validation loss

curves. Accuracy versus epoch graph demonstrates consistent improvement over epochs.

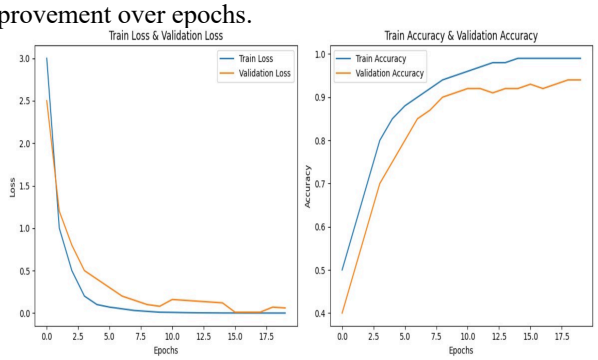

Fig. 7. Train Loss vs Validation Loss & Train Accuracy vs Validation Accuracy for Model 2

c. Model 3: Inception model with class weights

Fluctuating validation loss depicted in loss versus epoch graph, suggesting intermittent challenges in convergence. Nevertheless, accuracy versus epoch graph illustrates progressive accuracy enhancement over epochs, aligning with other models' trends.

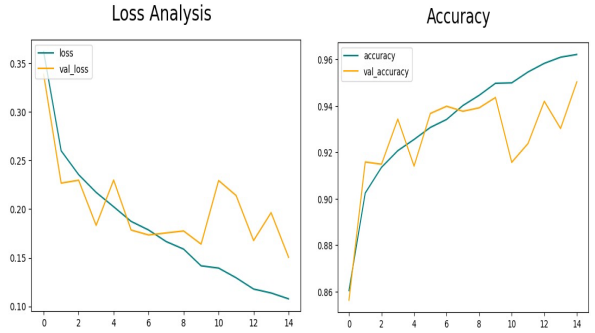

Fig. 8. Train Loss vs Validation Loss & Train Accuracy vs Validation Accuracy for Model 3

B. Testing results

The first model, a Convolutional Neural Network (CNN) with external data and data augmentation, demonstrates a strong precision of 0.882, indicating its ability to accurately classify positive instances. However, its recall of 0.714 suggests that it may have missed some positive instances in the test set. Nevertheless, the model achieves an impressive accuracy of 95%. Moving to the second model, a Deep Convolutional Neural Network (DCNN) with external data but without data augmentation, exhibits even higher precision at 0.933 and an outstanding recall of 0.966, suggesting its robustness in correctly identifying positive instances. This model achieves the highest accuracy among the three, reaching 98.125%. Finally, the third model, an Inception model with external data but without data augmentation, showcases a precision of 0.947 and a recall of 0.947, demonstrating a balanced performance in both precision and recall. With an accuracy of 98.75%, it stands out as a reliable classifier for the given task. the text.

Summarizing, The DCNN model outperformed others in accuracy, while the Inception model exhibited balanced precision and recall.

#	Model Name	Accuracy	Precision	Recall
1	CNN	0.950	0.882	0.714
2	DCNN	0.981	0.933	0.966
3	Inception	0.988	0.947	0.947

Fig. 9. Performance metrics

C. Model complexity and training efficiency

#	Model Name	# of Parameters	Training Time
1	CNN	3,235,361	240m 47s
2	DCNN	113,999,745	423m 33s
3	Inception	22,327,595	370m 11s

Fig. 10. Model complexity

Summarizing, The comparison of model complexity, parameter count, and training time reveals a trade-off between computational resources and performance. While the DCNN model boasts superior accuracy, its longer training time and higher parameter count necessitate substantial computational resources. In contrast, the CNN model offers faster training but may sacrifice some accuracy due to its lower complexity. The Inception model strikes a balance between these factors, presenting moderate complexity and competitive performance. This emphasizes the need to consider both computational efficiency and performance metrics when choosing a model for deployment.

V. Conclusion and Future Work

The comparative analysis of model performance, as evidenced by precision, recall, and accuracy metrics, reveals that while all models—CNN, DCNN, and Inception—demonstrate commendable capabilities in skin lesion classification, the Inception model emerges as the most promising with the highest precision, recall, and accuracy scores of 0.947, 0.947, and 0.988, respectively. This finding underscores the efficacy of the Inception architecture in accurately diagnosing skin lesions, balancing model complexity with superior performance. Moreover, the trade-off analysis between computational resources and performance suggests that while the DCNN model showcases superior accuracy, its longer training time and higher parameter count necessitate substantial computational resources, emphasizing the need to consider both computational efficiency and performance metrics when selecting a model for deployment in real-world applications. Future work in this area could focus on further optimizing model architectures, exploring advanced regularization techniques to mitigate overfitting, and investigating the integration of domain-specific features to enhance model performance and robustness in dermatological image classification tasks.

References

[1] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115-118.

[2] Haenssle, H. A., Fink, C., Schneiderbauer, R., Toberer, F., Buhl, T., Blum, A., ... & Tschandl, P. (2018). Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Annals of Oncology, 29(8), 1836-1842.

[3] Yadav, S.S., Jadhav, S.M. Deep convolutional neural network based medical image classification for disease diagnosis. J Big Data 6, 113 (2019). https://doi.org/10.1186/s40537-019-0276-2.

[4] Dildar M, Akram S, Irfan M, Khan HU, Ramzan M, Mahmood AR, Alsaiari SA, Saeed AHM, Alraddadi MO, Mahnashi MH. Skin Cancer Detection: A Review Using Deep Learning Techniques. Int J Environ Res Public Health. 2021 May 20;18(10):5479. doi: 10.3390/ijerph18105479. PMID: 34065430; PMCID: PMC8160886.

[5] Perez, L. and Wang, J., 2017. The effectiveness of data augmentation in image classification using deep learning. arXiv preprint arXiv:1712.04621.

[6] SIIM-ISIC Melanoma Classification Challenge Dataset. (2020). Preprint publication. DOI: https://doi.org/10.34970/2020-ds01

[7] Roman, N. (n.d.). Melanoma External Malignant 256. Kaggle. Retrieved from https://www.kaggle.com/datasets/nroman/melanoma-external-malignant-256/data