A Comparative Study for diagnosing Skin Cancer with Deep Learning Models of CNN and VGG16

Alwin Jackson Avaran
ajacksonavar2023@fau.edu
Florida Atlantic University
Boca Raton, Florida, USA

Saideva Goud Nathi
snathi2023@fau.edu
Florida Atlantic University
Boca Raton, Florida, USA

***Abstract*— Skin cancer, a serious public health concern, arises from the unchecked proliferation of skin cells. Timely and precise diagnosis is vital for effective treatment, especially in distinguishing between benign and malignant skin lesions. Traditional diagnostic methods, reliant on visual examinations by dermatologists, are time-consuming, typically ranging from five to fifteen minutes per case. This conventional approach, while effective, struggles to keep pace with the burgeoning incidence of skin cancer cases, necessitating a faster, yet accurate, diagnostic solution. In this study, we explore the potential of deep learning in advancing the diagnosis of skin cancer. Leveraging the computational environment of Google Colab and the deep learning libraries of TensorFlow and Keras, we implemented a dual-model framework that combines the strengths of Convolutional Neural Networks (CNNs) and the pre-trained VGG16 network, both well-regarded for their image recognition capabilities. Our models were trained and tested on a substantial Kaggle dataset containing 3,600 images of benign and malignant skin lesions, providing a solid basis for evaluation. The findings of our research reveal that the synergistic use of CNN and VGG16 models not only enhances the accuracy but also the efficiency of skin cancer classification, marking a significant improvement over traditional diagnostic method. The research underlines the transformative potential of such AI-driven tools in dermatology and their adaptability to a wider range of medical imaging tasks.**

Keywords— Skin cancer diagnosis, Deep learning, Convolutional Neural Networks (CNNs), VGG16 network , and Medical imaging

I. INTRODUCTION

Over nine thousand five hundred new instances of skin cancer are diagnosed every single day in the United States, and more than two people lose their lives to the disease every single hour[1][2]. There is a growing concern about the health of the general populace. The fact that more than 5.4 million cases were treated in 2012 alone is significant evidence of how widespread it is [1] . Due to the dangerously high incidence of occurrence, the economy is placed under a significant amount of pressure, with the expenses of treatment increasing to almost $8.1 billion annually[3]. One further data that demonstrates the significance of this issue is the fact that by the age of 70, one in every five people in the United States will have had skin cancer, and actinic keratosis is a common indicator of this condition [4]. In spite of the fact that melanomas are very hazardous, early detection is nevertheless a glimmer of hope since it significantly boosts the likelihood of survival. When a doctor is trying to diagnose a condition, the first thing they often do is examine any patches on the skin that seem to be worrisome. On the other hand, the accuracy of these tests is contingent upon the level of expertise possessed by the physician; in instances of melanoma, the accuracy of solo diagnostics ranges between 65 and 80 percent [5]. Increasing this degree of precision is accomplished via the use of dermatoscopic photography. The deeper layers of skin may be seen with the use of these specialized cameras, which are able to enlarge and capture high-resolution photographs under regulated lighting conditions. The diagnostic accuracy is improved by as much as 49% as a result of this, which results in a success rate of between 75% and 84% in locating melanoma [6][7]. The purpose of this research topic is to investigate ways in which skin cancer diagnosis may be made more precise, simplified, and expedited. In light of the fact that many kinds of lesions have a lot of similarities in subtle ways, we need more effective methods to detect them as soon as possible. The motivation for addressing this issue is not just to enhance the outcomes for patients, but also to minimize the financial and societal obstacles that are associated with skin cancer.

Before 2016, automatic skin lesion classification was done in several steps: images were pre-processed, lesions were segmented, defining features were extracted, and finally, the lesions were classified. Despite being organized, this method had a number of problems, yet can save from human life, effort and time [8]. Most importantly, correct feature extraction took a lot of time and needed a lot of subject knowledge. Also, mistakes in the beginning stages, like not segmenting correctly, could deteriorate further steps in modelling, making the total rating less accurate [9]. This chain reaction of mistakes shows how important it is to use a stronger and more automatic method to find skin cancer. In answer to these problems, our study suggests a new way to do things that uses the latest progress in AI and machine learning. We want to use Convolutional Neural Networks (CNNs) and the VGG16 model to speed up the diagnosis process. Since these models are good at spotting images[30], they seem like a good way to automatically group skin cancer into the right groups. The images let them learn and spot traits right away, so they don't have to be removed by manual, which is one of the main problems with normal ways. These models can also be added to mobile apps to make early skin cancer diagnosis easier for everyone. This will allow more people to get this important medical service. This research will test how well CNNs and the VGG16 model can sort skin sores into different groups. We will discuss about their design approach in more detail and show how these models can correctly tell the difference between spots that are healthy and ones that are dangerous. Not only will this help doctors get better at their jobs, but it will also make early skin cancer screening easier to get, especially in places where it's not available enough.

Fig.1. ProposedMethodology

II. Related Works

Dermatology has seen a big rise in the use of deep learning techniques to find skin cancer. These literature reviews are meant to give us an idea of the many different methods and formulas that have been created to help doctors find skin cancer more accurately and quickly. The study by Haghighi and coworkers [11], called "A Deep Convolutional Neural Network for Melanoma Recognition in Dermoscopy Cancer," came out in 2020. It was an attempt to solve the problems that come with automatic melanoma detection. They came up with a methodology even though they didn't have a lot of training data by mixing Convolutional Neural Networks (CNNs) with data enrichment. They also mixed CNN features with those from a Support Vector Machine (SVM) predictor. The combination method we used helped us tell the difference between melanoma and non-melanoma cancers with an accuracy of 89.52%. Researchers Cakmak et al. [12] looked into how the Nasnet Mobile deep neural network could be used to find cancer. Through their study, they brought attention to how important it is to add to data when dealing with class imbalances, especially in the HAM10000 dataset. These methods have a huge impact on how well models work, as shown by the fact that when the augmentation techniques were used, the accuracy went from 89.20% to an amazing 97.90%.

Brinker et al. [13] used the ResNet50 design, which is a model that has already been trained well, to tell the difference between nevi tumours and melanoma. They showed that their model had a sensitivity rate of 77.9% and a precision rate of 82.3%. This shows how useful current designs can be in the field of medical images classification by looking at how well they work. Han et al. [14] used the ResNet152 version to sort a lot of skin blemishes into groups. This is the same as the last case. The fact that their model got a specificity rate of 87.63% and a mean sensitivity of 88.2% across three different types of skin lesions showed how flexible the ResNet series is when it comes to classifying skin lesions. A full skin lesion classification method was created by Xie et al. [15]. This method divided skin lesions into two groups: normal and cancerous. An ensemble Neural Network (NN) model was used to classify the cancer in three steps: lesion extraction, feature extraction, and classification. This system did much better than past classifiers, with a 91.11% success rate. The authors Aswin et al. [16] came up with a new way to find skin cancer by combining genetic algorithm (GA) and artificial neural network (ANN) methods. They got an overall success rate of 88% with their method, which included preparation to get rid of hair and using the GLCM method to pull out features. After that, classification was done using a mix of genetic algorithms and artificial neural networks.

Backpropagation artificial neural networks were used to build an automatic system by Jaleel J.A. [17] that can find skin cancer. This system used a 2D-wavelet transform method to identify features and put the images into groups for cancerous and noncancerous cells.Choudhari and Biday [18] used maximum entropy thresholding to create an artificial neural network (ANN)-based diagnostic system that separated cancer based on how they looked. The authors got an accuracy level of 86.66% by using a gray-level co-occurrence matrix (GLCM) to select features. The features were then put into groups based on whether the skin cancer was aggressive or normal.

In their paper [19], Brinker and Hekler showed a cutting-edge predictor that was built on CNN and used the same classifications that doctors do. Here, it was shown that CNN can be used to classify skin lesions in both clinical and non-clinical settings. Refiantiet.al., wrote a study [20] about their work on making software that uses CNN algorithms, especially LeNet5, to sort skin cancer into different groups. An accuracy rate of 93% during training and 100% during testing showed that their approach, which included fine-tuning and adding more data, worked pretty well. Transfer learning was used by Hosny et al. [21], who also took out the model's last layer and replaced it with Softmax. They used certain cancer from the Ph2 dataset for training and testing, which shows that transfer learning works in this field.Deep learning systems, especially CNNs and their versions, could help find and classify skin cancer more accurately, as shown by the study that was done for this reason. A lot of new data preparation techniques, feature extraction methods, and frameworks like ResNet and Nasnet Mobile are making machine learning even more important in changing the way dermatology tests are done.

III. Methodology

In our research methodology of dermatological diagnostics, specifically targeting the detection and classification of skin cancer using deep learning models and the methodology is shown in Fig.1. The 'Skin Cancer: Malignant vs. Benign' dataset [22] from Kaggle was picked as the most useful tool for our study because it has many images of skin lesions. This set of data gives us a diverse and accurate image of both types of skin cancer. The reason this dataset was chosen is because it has a good mix of normal and cancerous cases and is still useful for dermatologists today at diagnosing skin problems. This collection has high-resolution images that have been put into two separate groups: cancerous and noncancerous. This difference is necessary in order to build a model that can accurately tell the difference between these two very important groups. Getting the dataset ready for training and testing the Convolutional Neural Network (CNN) model is a complicated process that needs to be done correctly. This method is needed to correctly group skin cancer into the right groups. OpenCV (cv2) is a Python package for image processing that has been praised since it was first released [23]. It is used in the first step of this process. OpenCV was chosen as the tool for our planning process because it works

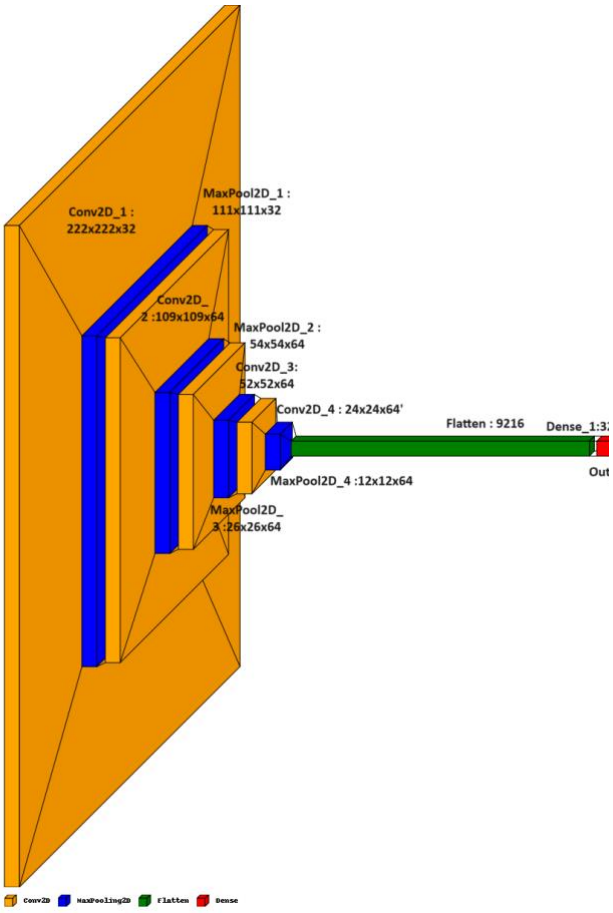

Fig.2. CNN Model Architecture

well with images data and can handle and change it quickly. Because of this, OpenCV is a very useful tool. Make sure all the cancer are in the same file. This is one of the most important things to do during the preparation step. Few things, like differences in color channels, images depth, and other format-related features, can really slow down the performance of a neural network. OpenCV gives us the tools we need to make these parts consistent across the whole collection. This makes sure that every image is in a style that is uniform and good for neural network processing. As soon as the cancer are all set up the same way, they are shrunk down to a standard size of 224 pixels by 224 pixels. This reshaping is very important because CNNs need all their inputs to be the same size. It is very important to fix this problem so that there are no errors that could lead to incorrect feature extraction by the CNN, which would then make it harder for the model to learn. This is why the process of resizing is carefully tweaked to keep each image as close as possible to its original size while still staying within the size that was given. The next step in our process is to figure out what the images are. Each image is marked as either cancerous or not cancerous before it is put into a category. This gets the guided learning method ready to go. With this two-way sorting, the model can learn new things and find connections between parts of the image and the groups that go with them. To begin, these names are changed into a number system, where 0 means "harmless" and 1 means "cancerous." This means that the neural network can use them to process and evaluate them. We are almost done getting the data ready. Now our dataset is divided into two groups: a training set and a testing set, about 80% of all the data is used as the training set. This is what the model learns. This lets it learn from a lot of different data points. Our model needs the testing set, which is about 20% of the rest of the data, to see how well the model works on data it has never seen before. We need this split to train the model well and make sure that the way we judge its success is fair. We also use a technique called "stratified splitting" to make sure the model is free of any biases and can correctly identify both groups[24]. This method makes sure that in both the training set and the testing set, the relative spread of cancerous and noncancerous cases stays the same. Lastly, we have a lot of care and make sure that every step of our planning process is done right. Part of this is calling out and breaking down information, as well as changing and scaling cancer. The experiment needs to reach this level of accuracy for the model to grow.

A. The Design and Architecture of Convolutional Neural Networks (CNN):

The Convolutional Neural Network (CNN) [25] has been carefully planned and built in terms of its layout and design to do its difficult job of classifying skin cancer. Our CNN model's main layer is a convolutional layer (Conv2D) that has 32 filters, each of which is 3x3 pixels in size. With this design, it is possible to find the right balance between the need to find all features and the need for fast computing. When dermatology images are first being looked at, this layer's main job is to find and record basic visual elements like lines and simple patterns. These things are necessary. As the model gets better, more and more convolutional layers are used, and each one has 64 filters. So that there are more options, this is done. It is possible for these levels to get deeper into the complicated images data and pull-out features that are even more specific. It is necessary to make things more complicated to correctly describe skin cancer, which often have minor and changing visible traits. What is used as an activation function in these convolutional layers is the Rectified Linear Unit (ReLU). ReLU was chosen because it is good at training deep neural networks, especially because it is linear and doesn't get too full. Because of this, it was chosen as the main reason. This is a great way to deal with the disappearing gradient problem, which comes up a lot when deep neural networks are being trained. In this case, this trait is especially helpful. We add a MaxPooling layer, which is also known as MaxPooling2D, after each convolutional layer. These layers have the function of reducing the spatial dimensions of the feature maps. This greatly lowers the amount of work that needs to be done on the computer and the number of factors that need to be included in the model. But they do more than just shrink cancer; they also help pull out the most important parts of the images and keep them safe. It is important to have this level of abstraction so that the model can better apply what it has learned from the training data to images it has never seen before. The flattening process is a big change that means our CNN system is changing in a big way. This process changes the two-dimensional feature maps into a one-dimensional feature vector so that there is a link between the convolutional layers and the thick layers that come after them. Changing the data in this way is an important step in getting them ready for the last stage of classification. The model then goes on to thick layers, which oversee processing the information it has learned even further. As the last and most noticeable part of the design, it has a thick layer with a sigmoid activation function. Because it's this layer's job to tell us how likely it is that a tumour is cancerous, our goal for binary classification was met. It was decided that the Adam algorithm would be the

best way to make our model better. It is possible to make more accurate changes during the training process with Adam, which is known for its flexible learning rate. This is an advantage that stands out even more when working with files that have a lot of dimensions, like the one we have. Further, binary crossentropy loss function helps us with our binary classification job. It gives a number value to the difference between the model's expectations of probabilities and the real binary labels, which helps the model make more accurate predictions. With a total of 388,225 parameters, our CNN design shows how powerful the model is for finding trends and extracting specific features. This level of detail is important because medical images are very complicated and vary from one another, especially ones that are used to find skin cancer. The architecture of CNN was shown in the Fig.2. will help us to make a CNN model that can regularly and correctly identify skin cancers. Therefore, a medical images analysis tool that is both very accurate and reliable in the field of dermatology. This complicated building design is a tribute to our commitment.

B. The VGG16 Model Implementation:

For our study, using transfer learning to add the VGG16 [26] model is an important part of making our skin disease labelling system work better. The VGG16 model, which is known for being very complete and working well in the ImageNet competition, is what our method is based on as its main design. The design of it, which includes both deep and sequential convolutional layers, lets it pull out a lot of different features from complex visual data. Because of this skill, VGG16 works really well in medical image analysis, where being able to see small, intricate patterns is very important for making the right diagnosis.It was the goal of the VGG16 design to capture traits that were already there at a lot of different levels of complexity. The first stages of the process are used to recognize simple patterns and backgrounds. As our process goes deeper, more complicated features are found. It is very important to use this tiered method for feature extraction when working with dermatology shots. They do this because each layer shows more complicated data that is needed for a correct study.

Using the idea of transfer learning, we change the VGG16's pre-trained weights so that they can be used for our task of classifying skin lesions. By using this approach, we don't have to do a lot of training right away. Instead, we can rely on the model's strong feature extraction skills that it has learned from the ImageNet dataset. Medical imaging uses this method a lot because they need to get useful information from very large datasets that haven't been labelled very well and have images that are very complicated. One way to do this is with pre-trained models like VGG16. Because we want VGG16 to work the way we want it to, we get rid of its first top layers, which were set up for a classification task with 1,000 classes. Instead, we've added a GlobalAveragePooling2D layer that takes each feature map and turns it into a single value. This makes it possible to successfully describe the most important features that the neural layers found. We need to focus on the most important information for our binary classification goal, which is to tell the difference between skin cancer that are normal and those that are cancerous. The summary report is an important part of narrowing down the network's outputs which was depicted in the Fig.3. The last thick layer with one neuron and a sigmoid activation function is added so that the system can be changed even more. This layer interprets the reduced traits and then gives a statistical result that shows how likely it is that the tumour is dangerous. This statistical result is important for making complex decisions in classification tasks because it gives us a sense of how likely it is that the model's statements are to be true. The reason we split our information into training sets and testing sets at the same time is to make sure the model is strong and reliable. This split is important for figuring out how well the model works with new data that hasn't been seen before because it gives a fair image of how well the model can adapt. To get around the problems that come with the small dataset sizes that are common in dermatology imaging, we use several different data enrichment techniques.

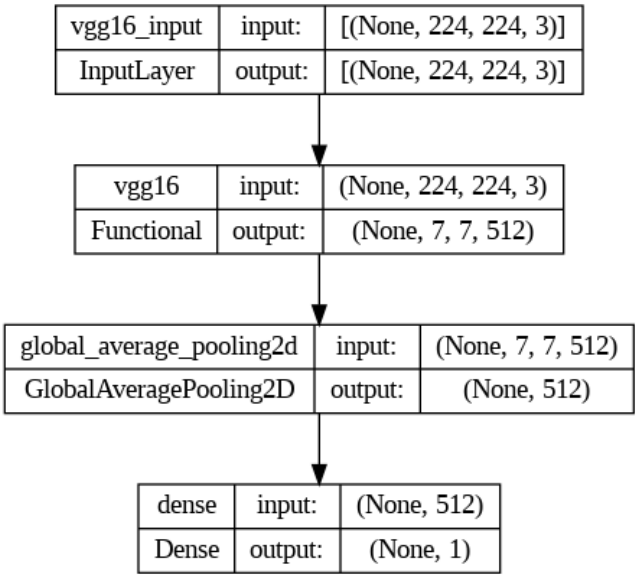

Fig.3. VGG16 Model Parametric Architecture

The goal of these methods is to make the training sample bigger than it really is. Operations like rotation, shifts, shear, zoom, and turning are some of these methods. The model can generalize better across a wider range of appearances and situations, which give it more examples of different types of lesions.The updated VGG16 model, which has all 14,715,201 features, is an example of a design that can fully handle and analyse complex images data. The model's ability to reliably identify features is shown by the large number of factors it has. This ability is necessary for tasks that need to be done quietly, like putting skin sores into groups. In conclusion, our method is an advanced way to classify skin lesions because it uses a CNN that was built just for our needs and a VGG16 model that was changed to work together. This two-model method combines the advanced feature extraction skills of a network that has already been taught with the flexibility of a custom design, which lets for even more customization. The model can generalize much better when a lot of extra data is added. This, along with VGG16's strong feature extraction tools, makes a big difference. This is especially important in medical imaging, where different images displays can make things very hard. As of now, deep learning has been used a lot to help dermatologists make assessments. Our analytical approach of using parameters for two models also stands out as a big step forward in this direction which was shown in the Table. I.

TABLE I. MODEL PARAMETRIC INVESTIGATION

Feature	Custom CNN Model	VGG16 Model
Total Parameters	3,88,225	1,47,15,201
Trainable Parameters	3,88,225	1,47,15,201
Non-trainable Params	0	0
Layers	- Conv2D (32 filters) - MaxPooling2D - Conv2D_1 (64 filters) - MaxPooling2D_1 - Conv2D_2 (64 filters) - MaxPooling2D_2 - Conv2D_3 (64 filters) - MaxPooling2D_3 - Flatten - Dense (32 units) - Dense_1 (1 unit)	- VGG16 (Functional, 512 filters) - Global Average Pooling2D - Dense (1 unit)
Output Shape	- Various, ending in (None, 1)	- Various, ending in (None, 1)

IV. EXPERIMENTAL RESULTS

The experimental framework of this study was meticulously structured to assess and compare the performance of two distinct convolutional neural network architectures for medical image classification. One important way to find out how well custom-designed models work is to compare them to well-known models that have already been taught how to do certain image processing jobs. The main tool used in this study is the one-of-a-kind Convolutional Neural Network. Many layers of convolution and pooling make up this model, which was made from scratch. It then has thick layers that are fully linked. The plan was to learn from medical cancer and record its features, with the main goal of telling the difference between dangerous and healthy places. On the other hand, the VGG16 model was used as a more advanced way. A lot of people have already trained this network. ImageNet is where the VGG16 model got its start, and it is famous for how well it can sort pictures into groups. It helps us see how helpful and good pre-trained models are in a certain area, like medical picture analysis, by using it in this work. A computer program called Python [27], which is famous in data science and machine learning, was used for the tests. It was built and trained with Keras, a high-level neural networks API that runs on top of TensorFlow. It was built on top of the OpenCV (cv2) library, which is powerful and customizable. The tool for making complex models in Keras [28] is easier to use. The most important part of this work is the information. There are medical cancer spots on it that have been marked as either normal or dangerous. This set of data is very important because these models are used to make medical diagnoses, and how well people do could depend on how well the models work. The training set has 1,440 cancers that aren't very dangerous and 1,197 tumors that are. The models can get a lot of training because there are a lot of cancers in each group. This helps them learn all the unique traits and trends that are connected to each group. The collection has 360 cases of cancer that isn't too bad and 300 cases of cancer that is very bad. This split makes sure that there is a good way to test how well the models work and how well they can change based on new information. The file's pictures are now all 224x224 pixels in size. In the field of deep learning for cancer, this many variables is normal. While still being able to pull out specific features, it makes good use of as little computer power as possible. So that the data stays the same, it is very important that all the cancers are the same size. This way, the models can be taught and tested on data that is set up in the same way. A number of important tests are used to see how well the custom-built CNN and the pre-trained VGG16 models can tell the difference between medical cancer that is not dangerous and cancer that is. There are a lot of these, such as the uncertainty matrix, accuracy, precision, memory, and f1-score. There were a certain number of cases that were correctly labeled as benign, malignant, false positives, and false negatives. The confusion matrix shows the number of true positives and false negatives. A true positive is a case that was correctly labeled as malignant when it was actually benign. Also shown is the number of false negatives and false positives. The model's general soundness can be judged by how accurate it is. The memory shows how well it can find all cases of that class. If the f1-score is a harmonic mean of accuracy and memory, it means that it is a good mix of the two.

The custom CNN model shows an accuracy of 82.12%, with precision values of 0.77 for benign cases and 0.87 for malignant cases. Its recall is 0.86 for benign cases but lower at 0.79 for malignant cases, leading to f1-scores of 0.81 for benign and 0.83 for malignant cases. From the Fig.4, the training and testing accuracy and loss of CNN are determined where the training loss for the CNN model decreases steadily from the first epoch to the sixth, indicating that the model is learning and improving its predictions over time. However, there is a significant spike in validation loss at the ninth epoch, suggesting that the model may be overfitting to the training data and not generalizing well to the validation data. This is corrected at the tenth epoch, where the validation loss drops, suggesting some recovery from overfitting.The accuracy graph for the CNN model shows fluctuations, but there is an overall upward trend in both training and validation accuracy. Notably, the validation accuracy surpasses the training accuracy at several points, which is unusual and typically suggests that the model might be underfitting. However, it can also indicate that the validation set might not be perfectly representative of the problem space, or there might be randomness in the training process due to factors like mini-batch gradient descent or data shuffling.

In contrast, the VGG16 model exhibits a slightly higher accuracy of 83.33%, with precision values of 0.75 for benign and 0.94 for malignant cases. Its recall is higher for benign cases at 0.94 but lower for malignant cases at 0.74, resulting in f1-scores of 0.84 for benign and 0.83 for malignant cases. From the Fig.5, the training and testing accuracy and loss of VGG16 are determined where the VGG16 model's training and validation loss shows more volatility compared to the custom CNN. Particularly, the validation loss exhibits sharp increases at epochs 2, 5, and 7, which suggests that the model's performance on the validation set is inconsistent. This could be due to the complexity of the model, the learning rate settings, or the data augmentation effects that introduce more variability in the training data.The VGG16 model’s accuracy, much like its loss, is also more volatile. The validation accuracy peaks at epoch 3, showing that the model performed best at this point on the validation set. The sharp drops in epochs 5 and 7 for validation accuracy mirror the spikes seen in the validation loss graph, reinforcing the model's inconsistency on the validation set.

the precision of malignant predictions and the recall of benign cases. However, both models have areas needing

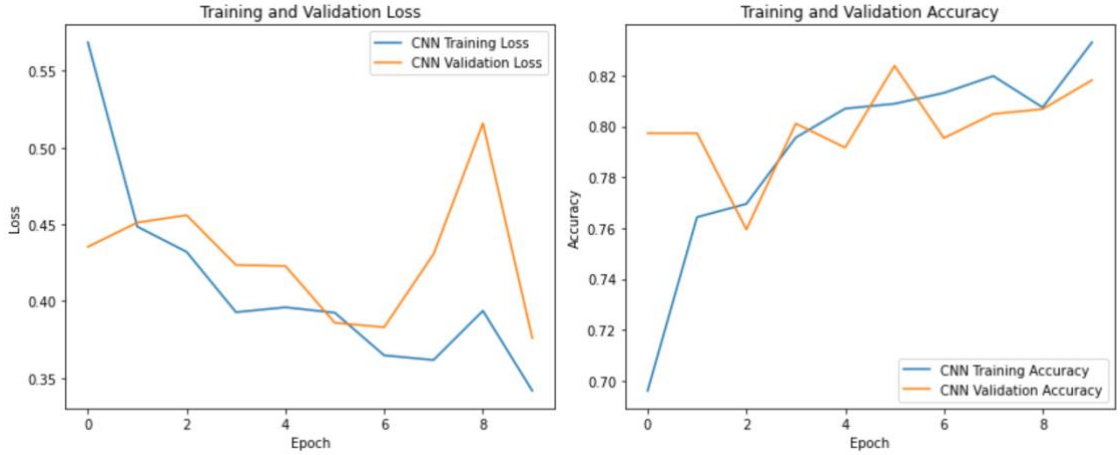

Fig. 1. Fig.4. CNN Training and Testing Loss and Accuracy

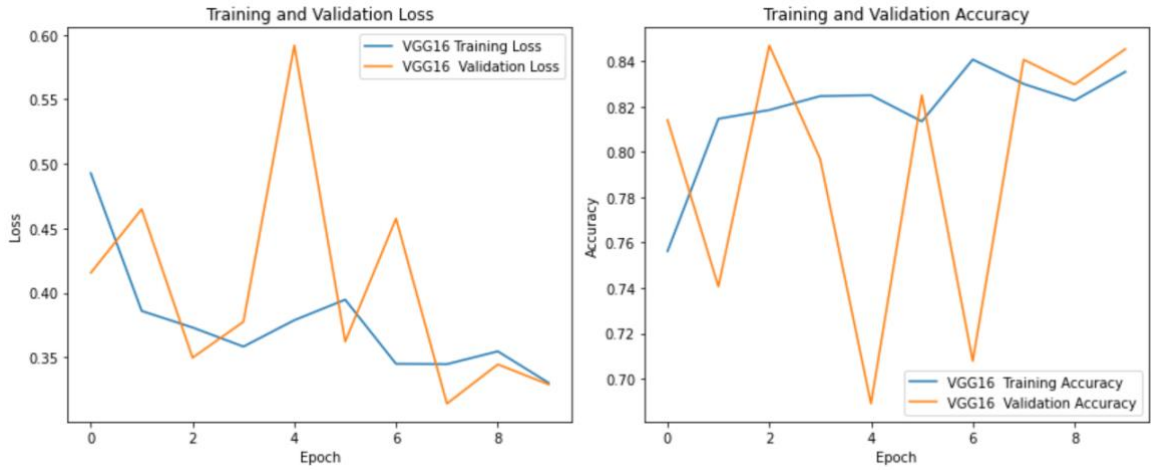

Fig. 2 Fig.5. VGG16Training and Testing Loss and Accuracy

TABLE II. COMPARISION OF MODEL AND THEIR PERFORMANCE METRICS

Metric	Custom CNN Model	VGG16 Model
Accuracy	82.12%	83.33%
Precision (Benign)	0.77	0.75
Precision (Malignant)	0.87	0.94
Recall (Benign)	0.86	0.94
Recall (Malignant)	0.79	0.74
F1-Score (Benign)	0.81	0.84
F1-Score (Malignant)	0.83	0.83
True Positives (Benign)	283	259
True Negatives (Malignant)	267	283
False Positives (Benign)	17	41
False Negatives (Malignant)	93	77

The comparison of metrics is mentioned in the Table. II where both models show similar levels of accuracy, the VGG16 model demonstrates a slight advantage, especially in improvement. The CNN model could benefit from a reduction in false negatives for malignant cases, while the VGG16 model needs to improve its detection of all malignant cases. Further tuning of hyperparameters and data augmentation techniques might enhance the generalizability and stability of both models, underscoring their strengths and limitations in medical image classification.

V. CONCLUSION

In summary, our research presents a comparative study of skin cancer classification utilizing a custom-built CNN and the pre-trained VGG16 model. The custom CNN achieved an accuracy of 82.12%, with notable precision in classifying malignant cases. The VGG16 model slightly outperformed the custom CNN, achieving an accuracy of 83.33% and demonstrating robustness in identifying benign cases. Despite their strengths, both models exhibited signs of overfitting and inconsistency across training epochs, highlighting areas for potential model refinement. Our findings suggest that both deep learning approaches possess significant potential for aiding in skin cancer diagnostics, with the VGG16 model's transfer learning providing a slight advantage in precision and recall. Future efforts will focus on optimizing these models through advanced regularization and data augmentation to

improve their diagnostic utility in clinical settings. This study contributes to the body of knowledge in machine learning applications for healthcare, indicating promising pathways for the development of automated diagnostic systems in skin cancer detection.

REFERENCES

[1] H. W. Rogers et al., "Incidence estimate of nonmelanoma skin cancer (keratinocyte carcinomas) in the US population," JAMA Dermatol., vol. 151, no. 10, pp. 1081–1086, Apr. 2015.

[2] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, "Cancer statistics, 2022," CA Cancer J. Clin., vol. 72, no. 1, pp. 7–33, 2022, doi: 10.3322/caac.21708.

[3] Agency for Healthcare Research and Quality, "Medical Expenditure Panel Survey," Rockville, MD, Jan. 2014. [Online]. Available: http://meps.ahrq.gov/mepsweb/data_stats/download_data_files.jsp.

[4] Skin Cancer Foundation, "Skin Cancer Facts & Statistics," [Online]. Available: https://www.skincancer.org/skin-cancer-information/skin-cancer-facts/. [Accessed: Oct. 10, 2023].

[5] P. K. Dhanyamraju and T. N. Patel, "Melanoma therapeutics: a literature review," J. Biomed. Res., vol. 36, no. 2, p. 77, 2022.

[6] A. R. A. Ali and T. M. Deserno, "A systematic review of automated melanoma detection in dermatoscopic images and its ground truth data," Proc. SPIE Int. Soc. Opt. Eng., vol. 8318, Feb. 2012, Art. no. 83180Q, doi: 10.1117/12.912389.

[7] G. Fabbrocini et al., "Teledermatology: From prevention to diagnosis of nonmelanoma and melanoma skin cancer," Int. J. Telemed. Appl., vol. 2011, Art. no. 125762, Sep. 2011, doi: 10.1155/2011/125762.

[8] K. M. Hosny, M. A. Kassem, and M. M. Foaud, "Classification of skin lesions using transfer learning and augmentation with Alex-net," PloS one, vol. 14, no. 5, Art. no. e0217293, May 2019.

[9] I. R. I. Haque and J. Neubert, "Deep learning approaches to biomedical image segmentation," Inform. Med. Unlocked, vol. 18, Art. no. 100297, 2020.

[10] G. Habib and S. Qureshi, "Compressed lightweight deep learning models for resource‐constrained Internet of things devices in the healthcare sector," Expert Syst., vol. 2023, Art. no. e13269, 2023.

[11] S. N. Haghighi et al., "A Deep Convolutional Neural Network for Melanoma Recognition in Dermoscopy Images," in Proc. 10th Int. Conf. Comput. Knowl. Eng. (ICCKE), Mashhad, Iran, 2020, pp. 453-456, doi: 10.1109/ICCKE50421.2020.9303684.

[12] M. Çakmak and M. E. Tenekeci, "Melanoma detection from dermoscopy images using Nasnet Mobile with Transfer Learning," in Proc. 29th Signal Process. Commun. Appl. Conf. (SIU), Istanbul, Turkey, Jun. 2021, pp. 1-4.

[13] T. J. Brinker et al., "Deep neural networks are superior to dermatologists in melanoma image classification," Eur. J. Cancer, vol. 119, pp. 11-17, 2019.

[14] S. S. Han et al., "Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm," J. Investig. Dermatol., vol. 138, no. 7, pp. 1529-1538, Jul. 2018.

[15] F. Xie et al., "Melanoma Classification on Dermoscopy Images Using a Neural Network Ensemble Model," IEEE Trans. Med. Imaging, vol. 36, no. 4, pp. 849-858, Apr. 2017, doi: 10.1109/TMI.2016.2633551.

[16] R. B. Aswin, J. A. Jaleel, and S. Salim, "Hybrid Genetic Algorithm: Artificial Neural Network Classifier for Skin Cancer Detection," in Proc. Int. Conf. Control, Instrum., Commun. Comput. Technol. (ICCICCT), Kanyakumari, India, Jul. 2014, pp. 1304-1309.

[17] J. A. Jaleel, S. Salim, and R. Aswin, "Artificial Neural Network Based Detection of Skin Cancer," Int. J. Adv. Res. Electr. Electron. Instrum. Eng., vol. 1, pp. 200-205, 2012.

[18] S. Choudhari and S. Biday, "Artificial Neural Network for Skin Cancer Detection," IJETTCS, vol. 3, pp. 147-153, 2014.

[19] T. J. Brinker et al., "Skin Cancer Classification Using Convolutional Neural Networks: Systematic Review," J. Med. Internet Res., vol. 20, no. 10, Art. no. e11936, Oct. 2018, doi: 10.2196/11936.

[20] R. Refianti, A. B. Mutiara, and R. P. Priyandini, "Classification of melanoma skin cancer using convolutional neural network," Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 3, 2019.

[21] K. M. Hosny, M. A. Kassem, and M. M. Foaud, "Skin Cancer Classification using Deep Learning and Transfer Learning," in Proc. 9th Cairo Int. Biomed. Eng. Conf. (CIBEC), Cairo, Egypt, 2018, pp. 90-93, doi: 10.1109/CIBEC.2018.8641762.

[22] Fanconic, "Skin Cancer: Malignant vs. Benign," Kaggle, [Online]. Available: https://www.kaggle.com/datasets/fanconic/skin-cancer-malignant-vs-benign/data.

[23] R. Garg, S. Maheshwari, and A. Shukla, "Decision support system for detection and classification of skin cancer using CNN," in Innovations in Computational Intelligence and Computer Vision: Proceedings of ICICV 2020, Springer Singapore, 2021, pp. 578-586.

[24] H. Tabrizchi, S. Parvizpour, and J. Razmara, "An improved VGG model for skin cancer detection," Neural Processing Letters, vol. 55, no. 4, pp. 3715-3732, 2023.

[25] G. Bradski, "The OpenCV Library," Dr. Dobb's Journal of Software Tools, 2000.

[26] F. Farias, T. Ludermir, and C. Bastos-Filho, "Similarity based stratified splitting: an approach to train better classifiers," arXiv preprint arXiv:2010.06099, 2020.

[27] G. Van Rossum and F. L. Drake Jr., Python Reference Manual, Centrum voor Wiskunde en Informatica Amsterdam, 1995.

[28] F. Chollet et al., "Keras," 2015. [Online]. Available: https://github.com/fchollet/keras.

[29] M. Abadi et al., "Tensorflow: A system for large-scale machine learning," in 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), 2016, pp. 265–283.

[30] Avaran, Alwin J., and K. S. Arun. "Automatic Helmet Detection: A Deep Learning Perspective." In Proceedings of the International Conference on Cognitive and Intelligent Computing: ICCIC 2021, Volume 2, pp. 335-343. Singapore: Springer Nature Singapore, 2023.