IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Skin Lesion Detection using VGG-16 and ResNet-50 based Hybrid CNN Model

[1]Preeti Gupta, [2]Sachin Meshram

[1]M. Tech. Scholar, [2]Assistant Professor and HOD

[1] Department of Electronics and Telecommunications,

Chouksey Engineering Collage, Bilaspur, Chhattisgarh-497101, India

Abstract: One of the most prevalent cancer forms, skin cancer has become more widespread in recent decades. It is critical to correctly diagnose skin lesions to distinguish between benign and malignant lesions in order to provide proper patient care. While there are various computerized approaches for classifying skin lesions, convolutional neural networks (CNNs) have been demonstrated to outperform traditional methods. We offer an hybrid CNN model for skin lesion categorization in this paper. We fuse an optimized pre-trained VGG-16 model with a pre-trained ResNet-50 CNN model to construct an hybrid model for skin lesion classification. In this work, we have used ISIC 2016-17 dermoscopic images to test the performance of the proposed hybrid CNN model. The proposed method achieves very good classification performance when tested on the ISIC Dataset, with an accuracy of 0.8565.

Index Terms* – *Convolutional Neural Networks* (*CNN*), *Deep Learning (DL), ResNet-50, Skin Lesion, VGG-16.

I. Introduction

I.1 Background

The skin is the human body's biggest organ. Skin cancers form when skin cells become disorganized and out of control, with the potential to disseminate to different regions of the body. The most prevalent disease on the globe is skin cancer. Skin cancer, for example, is the most frequent cancer with one in every five people in developed countries developing the disease over their lifetime. One of the worst kinds of skin cancer is malignant melanoma (the most lethal form) kills 10,000 people per year. It can be treated with a simple excision if caught early enough, but a delayed diagnosis is linked to a greater risk of death (an estimated 50 percent mortality rate). We provide a computational approach for skin lesion categorization that is entirely automated [1-3]. Deep features optimized from a variety of eminent abstraction levels and CNN architectures are used in the proposed technique. The most lethal kind of skin cancer is melanoma and its prevalence has skyrocketed in the previous 30 years. The most effective way to treat melanoma is to discover it early. Melanoma has a 98 percent five-year relative survival rate when it is confined, but it drops to approximately 14 percent when it is progressed. As a result, early and accurate detection of melanoma is crucial. To identify pigmented skin lesions, dermoscopy imaging is employed in order to diagnose melanoma or suspected skin lesions. As an initial step, it is exploited to identify suspicious skin lesions using a non-invasive method. Fig. 1 shows some sample images of various skin cancers.

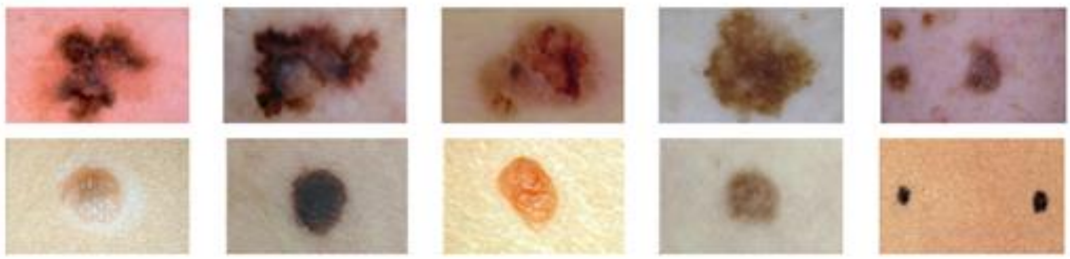


Fig. 1: Some samples of Skin Lesion and normal dermoscopic images.

I.2 Motivation

Cancer is an extremist life threat to human life. It can sometimes result in a human's death. In the human body, various types of cancer can exist, and one of the most rapidly growing malignancies is skin cancer that can lead to death. It is triggered by a variety of circumstances, including smoking, alcohol consumption, allergies, infections, viruses, physical activity, changes in the environment, and exposure to ultraviolet (UV) light, among others. UV rays from the sun have the potential to destroy the DNA inside skin cells. Skin cancer can also be caused by odd swellings of the human body. As skin cancer is linked to a higher risk of death - the estimated mortality rate is 50%. If discovered early enough, it can be treated with a simple excision, but a later diagnosis is linked to a higher risk of death. To reduce the death rate due to skin cancer where in early detection can save the life of the person. Thus there is a dire need of non-invasive efficient skin lesion detection system.

I.3 Contributions

In this work, we used two pre-trained CNN architectures namely: VGG-16 and ResNet-50 to propose a hybrid CNN framework for skin lesion classification. The major contributions of this work are as follows:

- Extensively analyze the performance of VGG-16 CNN model with Adam optimizer in multi-class skin lesion classification.
- Extensively analyze the performance of ResNet-50 CNN model with Adam optimizer in multi-class skin lesion classification.
- Proposed a hybrid CNN framework using VGG-16 and ResNet-50 for automatic non-invasive skin lesion grade detection.
- Used ISIC 2016-17 dermoscopic image dataset to train and test the proposed hybrid CNN model.

The rest of the paper has been structured as follows: After introduction section, a comprehensive literature review has been presented. Following that the brief description of VGG-16 and ResNet-50 has been delineated. Subsequently, the proposed scheme has been explained in detail in the next section. Later, the experimental analysis followed by the conclusion and future works has been presented.

II. Literature Review

Many researchers have proposed numerous types of Skin lesion detection systems. Those are mostly based on various CNN techniques such as AlexNet, ResNet, GoogLeNet [1-3]. However, there are just a few automated skin lesion classification systems on the market, the most of which are PC-based and require additional peripherals and/or regulated settings to gather data.

Ganster et al. [4] developed an integrated system, Automated Melanoma Recognition. A fusion of the outputs of three algorithms was used to achieve automated picture segmentation; as a consequence, 96 percent of the pictures were successfully segmented. When classifying into three classes, a 24-NN classifier has shown a melanoma detection rate of 73 percent, and 87 percent sensitivity and 92 percent specificity for the “not benign” class in a two-class situation.

Ercal et al. [5] developed a technique that combined using a commercial Neural Network classifier, dermatologists determined lesion boundaries that used computed characteristics to indicate irregularity and asymmetry in lesion forms, in addition to colors that are similar. This technique was successful in classifying 80% of the time. Neither of the aforementioned methods was created with the common public in mind or was intended to be used on mobile devices.

MoleSense [6] is a product of Opticom Data Research (Canada), which is a soft-ware program that collects photos of skin lesions, analyses them using ABCD principles, and derives feature values without categorizing them. However, the use of particular pre-trained network topologies or certain layers for extracting deep characteristics limits these studies. Furthermore, there was just one pre-trained network employed. Simultaneously, [7] utilized a single pre-trained AlexNet, [8] used a single pre-trained VGG16, and [10] used a single pre-trained Inception-v3 [9] network.

We used a pre-trained and improved version of the VGG-16 CNN and ResNet-50 architecture in this study to classify skin lesion pictures obtained using a camera into three categories: Nevus, Seborrheic Keratosis, and Malignant. We have used these two CNN architectures as both of them have demonstrated efficient performances for different application like image processing, object detection, medical image processing, image fusion, medical image fusion, and remote sensing [11-18].

III. VGG-16 CNN Model

In the race to enable computers to "see" the world, VGG16 has proven to be a crucial turning point. Since several decades ago, the field of computer vision (CV) has made significant progress in this area. One of the key inventions that helped pave the path for other advances in this sector is VGG16. It is a convolutional neural network (CNN) model that Andrew Zisserman and Karen Simonyan from the University of Oxford came up with. The model's concept was put out in 2013, but the actual model was presented in 2014 as part of the ILSVRC ImageNet Challenge. 2014's ILSVR (ImageNet) competition was won using the convolution neural network (CNN) architecture VGG16. It is regarded as one of the best vision model architectures created to date. The most distinctive feature of VGG16 is that it prioritized having convolution layers of 3x3 filters with a stride 1 and always utilized the same padding and maxpool layer of 2x2 filters with a stride 2. Throughout the whole design, convolution and max pool layers are arranged in the same manner. It concludes with two fully connected layers (FC) and a softmax for output. The 16 in VGG16 denotes the fact that there are 16 layers with weights. There are around 138 million parameters in this network, making it a sizable network. Further, Fig. 2 demonstrates internal details of VGG-16 CNN model, whereas Fig. 3 depicts the basic flow diagram of the typical VGG-16 CNN model.

Structural Details of VGG-16 CNN Model

- 2 x convolution layer of 64 channel of 3x3 kernal and same padding.
- 1 x maxpool layer of 2x2 pool size and stride 2x2.
- 2 x convolution layer of 128 channel of 3x3 kernal and same padding.
- 1 x maxpool layer of 2x2 pool size and stride 2x2.
- 3 x convolution layer of 256 channel of 3x3 kernal and same padding.
- 1 x maxpool layer of 2x2 pool size and stride 2x2.
- 3 x convolution layer of 512 channel of 3x3 kernal and same padding.
- 1 x maxpool layer of 2x2 pool size and stride 2x2.
- 3 x convolution layer of 512 channel of 3x3 kernal and same padding.
- 1 x maxpool layer of 2x2 pool size and stride 2x2.

Fig. 2: Internal structural details of the VGG-16 CNN model.

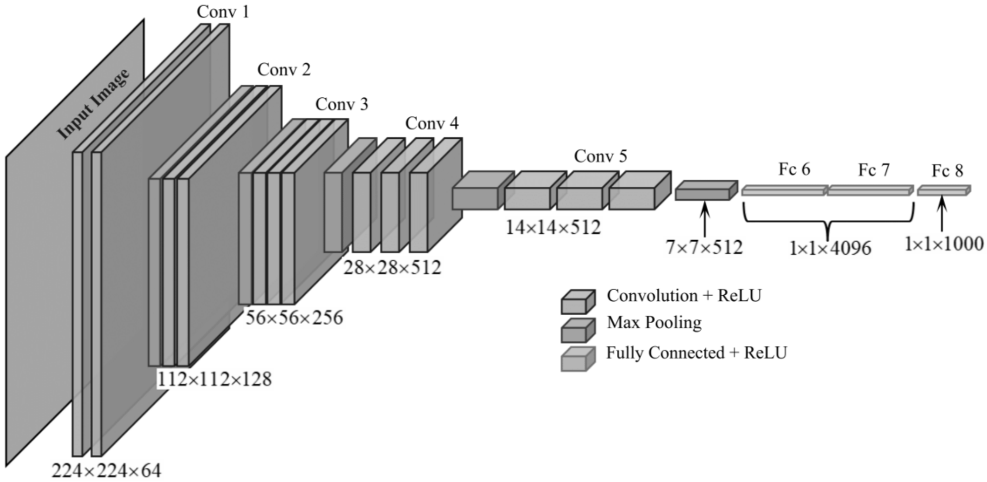


Fig. 3: Schematic flow diagram of VGG-16 CNN model.

IV. RESNET-50 CNN MODEL

ResNet, which stands for Residual Networks, is a well-known neural network that serves as the foundation for many computer vision applications. The 2015 ImageNet competition was won by this model. ResNet was a significant advancement in that it effectively enabled us to train incredibly deep neural networks with more than 150 layers. Due to the issue of vanishing gradients, very deep neural network training was challenging before ResNet. There are 5 stages in the ResNet-50 model, each having a convolution and an identity block. Each identity block and each convolution block each have three convolution layers. Over 23 million parameters may be trained with the ResNet-50. Further, Fig. 4 demonstrates internal details of ResNet-50 CNN model, whereas Fig. 5 depicts the basic flow diagram of the typical ResNet-50 CNN model.

Structural Details of ResNet-50 Model

- A convolution with a kernel size of 7 x 7 and 64 different kernels all with a stride of size 2 giving us 1 layer.
- Next we see max pooling with also a stride size of 2.
- In the next convolution there is a 1 x 1, 64 kernel following this 3 x 3, 64 kernel and at last a 1 x 1, 256 kernel, these three layers are repeated in total 3 time so giving us 9 layers in this step.
- Next we see kernel of 1 x 1, 128 after that a kernel of 3 x 3, 128 and at last a kernel of 1 x 1, 512 this step was repeated 4 time so giving us 12 layers in this step.
- After that there is a kernal of 1 x 1, 256 and two more kernels with 3 x 3,256 and 1 x 1, 1024 and this is repeated 6 times giving us a total of 18 layers.
- And then again a 1 x 1, 512 kernels with two more of 3 x 3, 512 and 1 x 1, 2048 and this was repeated 3 times giving us a total of 9 layers.
- After that we do an average pool and end it with a fully connected layer containing n nodes and at the end a softmax function so this gives us 1 layer.

Fig. 4: Internal structural details of the ResNet-50architecture.

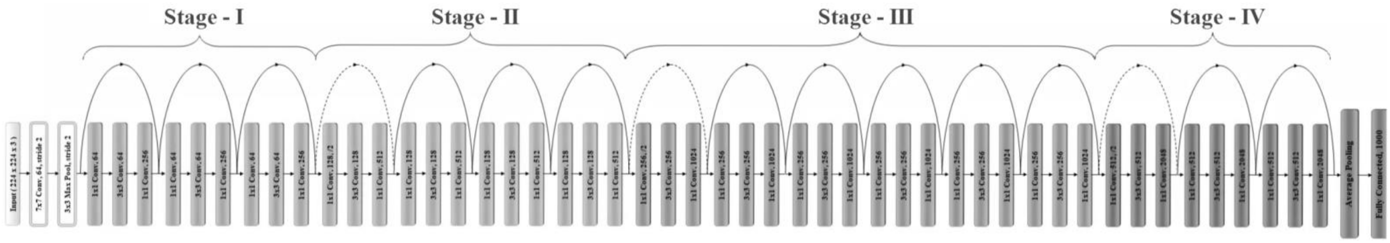


Fig. 5: Schematic flow diagram of the ResNet-50architecture.

V. Proposed Model

Deep learning has many tools of invaluable measure that has helped in the movement towards detecting skin lesions using computer-aided diagnosis. One of the requirements for the application of deep learning methods is their need for enormous training data. The main problem with the medical domain is the lack of quality and quantity for the open data. This is a characteristic of medical field. There are many clinics and hospitals which holds large amount of data but they cannot make it public because of many reasons mainly because of privacy issues. Despite of this problem there are many authors who pushed their limits and technologies and were successful in overcome these barriers. Also, with the help of creation like ISIC Archive, it is possible to get data and hence utilize the power of CNN for diagnosis in the medical field. In the field of skin lesion there are majorly two research trends that can be visualized. The first strategy utilizes the CNN architectures with the help of transfer learning and hence overcoming the availability of limitation of data to some extent. Another approach uses the larger and unlabeled data that is available through unsupervised learning techniques, and then supervised learning is applied for classification. The work presented in this article utilized the former approach, that is, transfer learning due to unavailability of larger dataset.

In this work, we have proposed a hybrid CNN model that will fuse features of pre-trained VGG-16 and ResNet-50 to detect possibility of skin lesion using a dermoscopic image. Initially a dermoscopic image has been provided as an input to both the selected CNN models. Next, we have selected the output of the final max pooling layer of VGG-16 architecture and supplied the selected features as input to a new 1×1024 Fully Connected (FC) layer with ReLU activation function. At the same time we have selected the output of the final average pooling layer of ResNet-50 architecture and supplied the selected features as input to a new 1×1024 FC layer with ReLU activation function. Next, we have constructed a new 1×2048 FC layer that will fuse the features generated from 1×1024 FC layer of VGG-16 and 1×1024 FC layer of ResNet-50 architecture. Finally, we have used another 1×512 FC layer to extract more salient features. Following that, we have used our final 1×3 FC layer with Softmax activation function to carry out the final classification task. Moreover, the proposed hybrid CNN model has been retrained using 80% of the images of the dataset. Whereas, the remaining 20\% images of the dataset has been used to validate the performance of the proposed hybrid model. Next, Fig. 6 displays the detailed structural diagram of the proposed hybrid CNN model. Subsequently, Fig. 7 depicts the algorithmic steps of the proposed classification scheme.

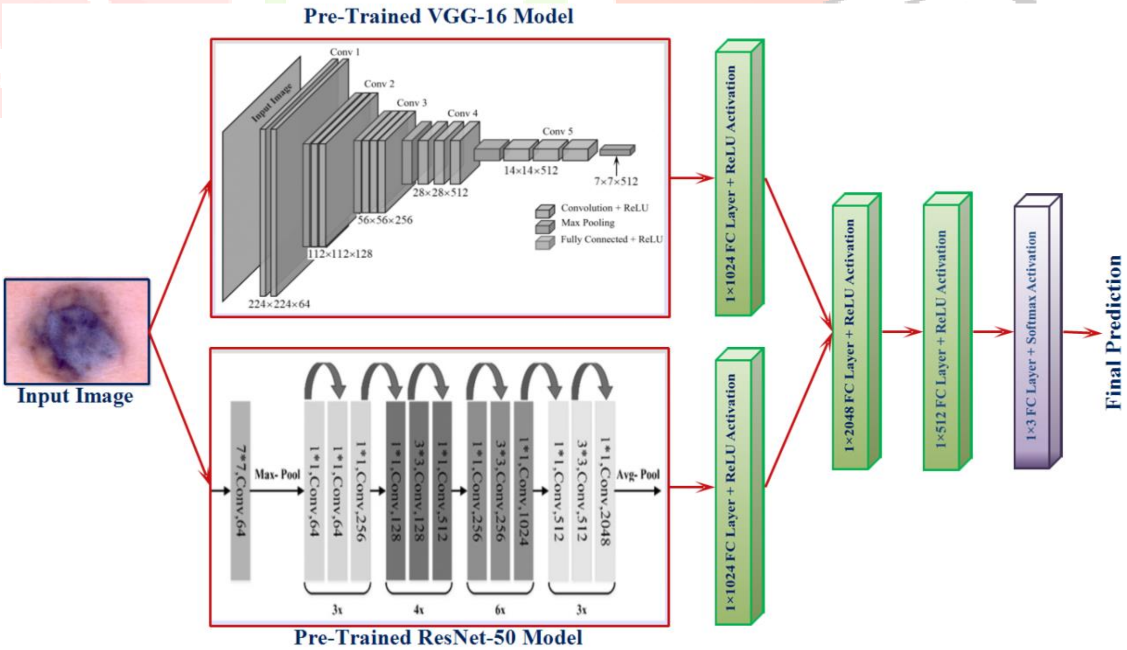


Fig. 6: Detailed structural diagram of the proposed hybrid CNN model.

Algorithm 1: Algorithm: Skin Lesion Detection.

Input: Input a Skin Lesion query image I_q.
Output: Identified Class $\mathcal{C}_q$ of the image I_q.
Parameter: I_q must be a Skin Lesion Demographic color image.

```
Create and instance of pre-trained VGG-16 CNN architecture.
Create and instance of pre-trained ResNet-50 CNN architecture.
Construct the proposed hybrid CNN architecture as suggested in Fig. 6.
Select the Skin Lesion Data-set Z_n, where n = |Z|.
Split the Data-set Z_n in two parts as Z_nT and Z_nV.
Assign 80% image of Z_n to Z_nT for training of the CNN model.
Assign class label information to every image of Z_nT for training of hybrid CNN architecture.
Assign 20% image of Z_n to Z_nV for testing and validation of CNN.
Perform the suggested Phase-I training on the CNN model.
for ∀I_i ∈ Z_nT do
    Used Adam optimizer to perform supervised trainining of the hybricd CNN architecture.
Select the input query image I_q ∈ Z_nV.
Resize the I_q to 224 × 224 × 3 pixels.
Feed the I_q to the trained hybrid CNN architecture for class label prediction.
Display the predicted class value C_q as final output.
```

Fig. 7: Algorithmic steps of the proposed skin lesion classification scheme.

VI. RESULT AND DISCUSSION

VI.1 Dataset Description

The Dataset used in this work consists of skin mole images that are Nevus, Seborrheic Keratosis, and Malignant for skin lesion detection by ISIC 2016-17 challenges which is publicly available to use from their official website, which consists of the skin lesion images of Nevus, Seborrheic Keratosis, and Malignant skin lesion. The dataset contains 4946 images in total. Among all these images, 1372 images belong to Nevus class, 1880 images belong to Melanoma class, and 1694 images belong to Seborrheic Keratosis class. Moreover, the training part of dataset consists of 80% of the images and rest 20% of the images have been utilized for testing.

VI.2 Results and Outcomes

We divided our dataset taken from Kaggle into 80:20 partitions for training set and test set respectively. Upon training our model we got an accuracy of more than 90% in most of the cases. Further we found that running the algorithm for more than 100 iterations the model overfits the given data and for 5 iterations it underfits the data, so we settled for 10 to 90 iterations which give optimal results. We also tested the accuracy of determining whether a given picture of a skin lesion is malignant or benign by personally examining certain test photographs. The outcome of the sample test images have been presented in Fig. 8.

Fig. 8: Outcome of the sample test images.

Further, the proposed hybrid model has been trained for multiple times (5 times each) for 200 epochs, with Adam as optimizer, learning rate as 1 and batch size as 32. This has been done so as to make comparative study more authentic. As concept of transfer learning is used for this work, the last classification layer is kept common for VGG-16 and ResNet-50 models. After training of the proposed model, the trained model has been deployed for the testing and validation. Fig. 9(a) shows the detailed performance of the hybrid model that uses only VGG-16 network as a backbone. Here, we have achieved 0.8040 of accuracy. Further, Fig. 9(b) shows the detailed performance of the hybrid model that uses only ResNet-50 network as a backbone. Here, we have achieved 0.8343 of accuracy. Further, Fig. 9(c) shows the detailed performance of the hybrid model that uses both VGG-16 and ResNet-50 network as a backbone. Here, we have achieved 0.8565 of accuracy. Next, Fig. 10 demonstrates the ROC curve of the proposed hybrid model, whereas Fig. 11 shows the confusion matrix of the proposed model.

	precision	recall	f1-score	support
0	0.86	0.86	0.86	390
1	0.77	0.88	0.82	274
2	0.92	0.83	0.87	326
micro avg	0.85	0.85	0.85	990
macro avg	0.85	0.85	0.85	990
weighted avg	0.86	0.85	0.85	990
samples avg	0.85	0.85	0.85	990

(a) VGG-16 based hybrid model

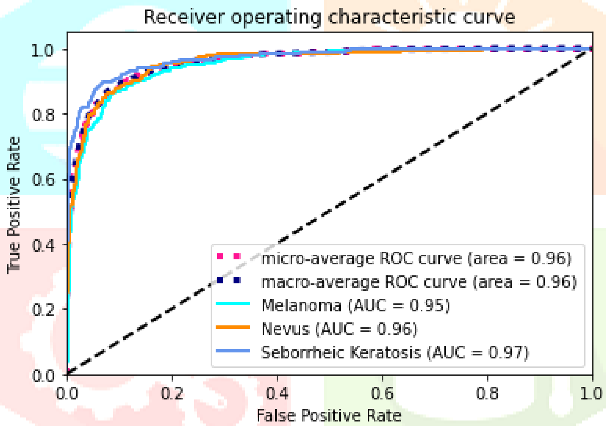
	precision	recall	f1-score	support
0	0.90	0.82	0.86	384
1	0.82	0.87	0.84	270
2	0.87	0.92	0.89	336
micro avg	0.87	0.87	0.87	990
macro avg	0.86	0.87	0.87	990
weighted avg	0.87	0.87	0.87	990
samples avg	0.87	0.87	0.87	990

(b) ResNet-50 based hybrid model

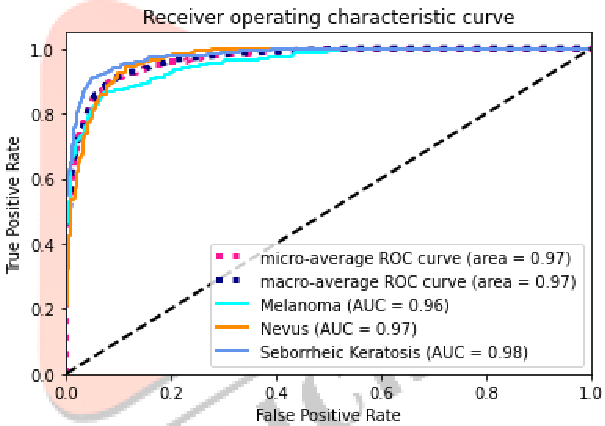
	precision	recall	f1-score	support
0	0.94	0.82	0.87	376
1	0.83	0.93	0.88	283
2	0.88	0.92	0.90	331
micro avg	0.88	0.88	0.88	990
macro avg	0.88	0.89	0.88	990
weighted avg	0.89	0.88	0.88	990
samples avg	0.88	0.88	0.88	990

(c) Prposed hybrid model (VGG-16 + ResNet-50)

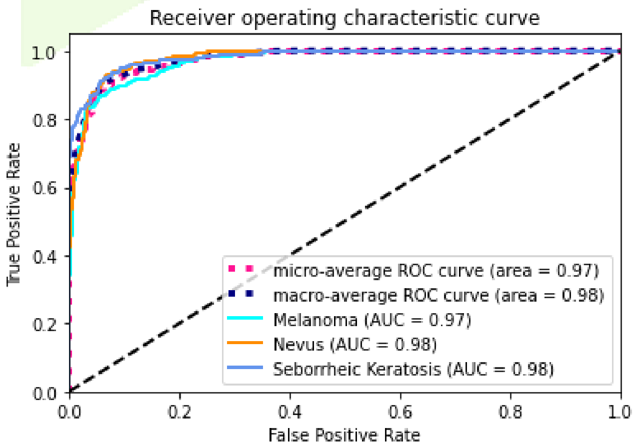
Fig. 9: Detaied perfroamcens of the proposed hybrid model.



(a) ROC of VGG-16 based hybrid model

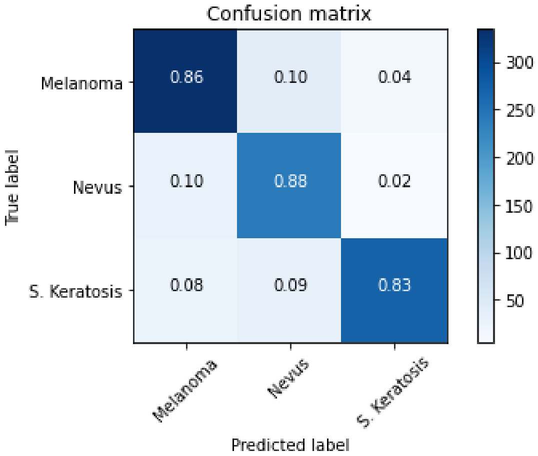


(b) ROC of ResNet-50 based hybrid model

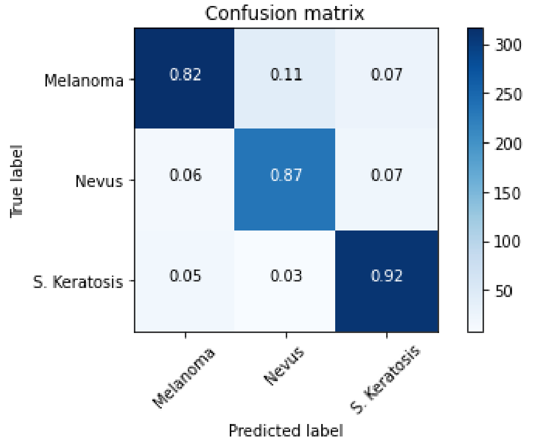


(c) ROC of Prposed hybrid model (VGG-16 + ResNet-50)

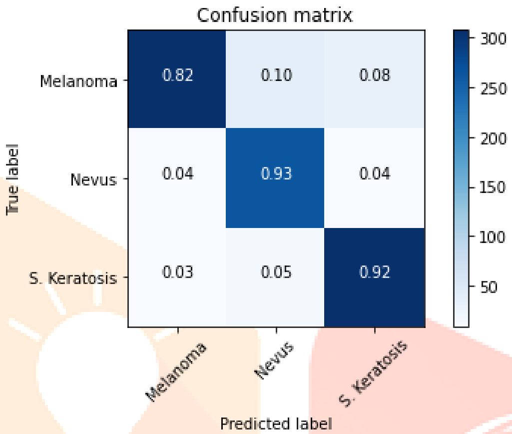
Fig. 10: Detaied perfroamcens of the proposed hybrid model in terms of ROC curve.



(a) CM of VGG-16 based hybrid model



(b) CM of ResNet-50 based hybrid model



(c) CM of Prposed hybrid model (VGG-16 + ResNet-50)

Fig. 10: Detaied perfroamcens of the proposed hybrid model in terms of Confusion Matrix (CM).

As we see from above results that the proposed hybrid model gave the best average test accuracies out of all the three models. Although all the related work mostly used VGG-16 or ResNet models for their work, but for this dataset the proposed hybrid model outperformed both of them. This proves the fact, that even for the same line work, it is not necessary that the same model will always work better. At the same time, the recorded results confirm that the hybrid model is a better choice for skin lesion classification task.

VII. Conclusions

This article discussed the importance of the automated classification method, which can support skin lesion diagnosis. We have discussed different approaches which were followed to increase the accuracy of the predictions, which can be proved valuable in the medical field. However, there are still some problems and difficulties in achieving this goal; even more, when we study clinical images usually present in immense diversity because of different variables such as cameras and other environmental effects. Consequently, the work gave a hybrid CNN model on which two different CNN models, namely – VGG-16 and ResNet-50, were employed and finally made sufficiently well for classifying three skin lesions that have given acceptable results. Additionally, the presented work has given techniques for the model selection decision process by providing a systematic set of procedures that is general and can be used in any other application also. The work presented has shown a way of utilizing the best model to construct a hybrid CNN model for the work. Then, we used different experiments on the model and evaluated their performance in terms of accuracy, precision, recall, and f-measure.

References

[1] Jayapriya K, Jacob IJ. Hybrid fully convolutional networks-based skin lesion segmentation and melanoma detection using deep feature. International Journal of Imaging Systems and Technology. 2020 Jun; 30(2):348-57.
[2] Murphree DH, Ngufor C. Transfer learning for melanoma detection: Participation in ISIC 2017 skin lesion classification challenge. arXiv preprint arXiv:1703.05235. 2017 Mar 15.
[3] Qassim H, Verma A, Feinzimer D. Compressed residual-VGG16 CNN model for big data places image recognition. In2018 IEEE 8th annual computing and communication workshop and conference (CCWC) 2018 Jan 8; 169-175.
[4] Ganster H, Pinz P, Rohrer R, Wildling E, Binder M, Kittler H. Automated melanoma recognition. IEEE transactions on medical imaging. 2001 Mar; 20(3):233-9.
[5] Salah B, Alshraideh M, Beidas R, Hayajneh F. Skin cancer recognition by using a neuro-fuzzy system. Cancer informatics. 2011 Jan;10:CIN-S5950.
[6] Ramlakhan K, Shang Y. A mobile automated skin lesion classification system. In2011 IEEE 23rd International Conference on Tools with Artificial Intelligence 2011 Nov 7; 138-141.
[7] Kawahara J, BenTaieb A, Hamarneh G. Deep features to classify skin lesions. In2016 IEEE 13th international symposium on biomedical imaging (ISBI) 2016 Apr 13; 1397-1400.

[8] Lopez AR, Giro-i-Nieto X, Burdick J, Marques O. Skin lesion classification from dermoscopic images using deep learning techniques. In2017 13th IASTED international conference on biomedical engineering (BioMed) 2017 Feb 20; 49-54.
[9] Mirunalini P, Chandrabose A, Gokul V, Jaisakthi SM. Deep learning for skin lesion classification. arXiv preprint arXiv:1703.04364. 2017 Mar 13.
[10] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. InProceedings of the IEEE conference on computer vision and pattern recognition 2016; 2818-2826.
[11] Pradhan J, Kumar S, Pal AK, Banka H. Texture and color visual features based CBIR using 2D DT-CWT and histograms. InInternational Conference on Mathematics and Computing 2018 Jan 9 (pp. 84-96). Springer, Singapore.
[12] Majhi M, Pal AK, Pradhan J, Islam SK, Khan MK. Computational intelligence based secure three-party CBIR scheme for medical data for cloud-assisted healthcare applications. Multimedia Tools and Applications. 2021 Feb 16:1-33.
[13] Keles A, Keles MB, Keles A. COV19-CNNet and COV19-ResNet: diagnostic inference Engines for early detection of COVID-19. Cognitive Computation. 2021 Jan 6:1-1.
[14] Kumar S, Pradhan J, Pal AK. A CBIR technique based on the combination of shape and color features. InAdvanced Computational and Communication Paradigms 2018; 737-744.
[15] Soni P, Pradhan J, Pal AK, Islam SH. Cybersecurity Attack-resilience Authentication Mechanism for Intelligent Healthcare System. IEEE Transactions on Industrial Informatics. 2022 Jun 1.
[16] Pradhan J, Bhaya C, Pal AK, Dhuriya A. Content-Based Image Retrieval using DNA Transcription and Translation. IEEE Transactions on NanoBioscience. 2022 Apr 29.
[17] Raj A, Pradhan J, Pal AK, Banka H. Multi-scale image fusion scheme based on non-sub sampled contourlet transform and four neighborhood Shannon entropy scheme. In2018 4th International Conference on Recent Advances in Information Technology (RAIT) 2018 Mar 15; 1-6).
[18] Pradhan J, Pal AK, Banka H. Medical Image Retrieval System Using Deep Learning Techniques. InDeep Learning for Biomedical Data Analysis 2021;101-128.