EXPLORING SKIN LESION CLASSIFICATION THROUGH EXPLAINABLE AI - LEVERAGING Grad-CAM FOR INTERPRETABILITY

B.Kishore Kumar[1], G.Spurthy Vahini[2], B.S.S.Durga Tanmayee[3], M.Devi Sri Chandana [4],M.Ashwini [5],P.Abheesta[6]

[1] Assistant professor, BVRIT HYDERABAD College of Engineering for women,kishorekumar.b@bvrithyderabad.edu.in
[2] B.Tech Student, BVRIT HYDERABAD College of Engineering for women,20wh1a6605@bvrithyderabad.edu.in
[3] B.Tech Student, BVRIT HYDERABAD College of Engineering for women,20wh1a6603@bvrithyderabad.edu.in
[4] B.Tech Student, BVRIT HYDERABAD College of Engineering for women,20wh1a6624@bvrithyderabad.edu.in
[5] B.Tech Student, BVRIT HYDERABAD College of Engineering for women,21wh5a6605@bvrithyderabad.edu.in
[6] B.Tech Student, BVRIT HYDERABAD College of Engineering for women,20wh1a6604@bvrithyderabad.edu.in

Abstract

Advancements in the field of AI and ML have had a great impact on the field of healthcare and automating medical assistance. Yet the underlying skepticism prevails in the usage of such technology in such a field as it has hithered to been handled by professionals of the field. Hence the advent of Explainable Artificial Intelligence(XAI) assumes significance in this context. The proposed work aims at tackling such a problem in the context of skin lesion classification using XAI techniques such as Grad-CAM to provide an understanding of the prediction made by the model which has been trained on the HAM-10000 dataset. Using data augmentation techniques the Mobile Net model has been trained leading to better results than the original dataset in early detection of the seven conditions of skin lesions. Using Grad-CAM the image is visualized with an overlay of the heat map resulting in the prediction of the model hence aiding in understanding the result better and generating a factor of trust towards the machine's predictions.

Keywords: Machine learning, Explainable Artificial intelligence, skin lesion classification, data augmentation, Grad-CAM

1 Introduction

Skin lesions represent areas of your skin that deviate from its normal appearance. They can vary widely in size, shape, color, and texture and may arise from diverse causes, including infections, inflammation, trauma, autoimmune reactions, or malignant growths. Lesions can manifest anywhere on the body and often serve as visible clues to underlying health issues. They can arise from various causes, including injuries or damage such as sunburn.

While common, they can also serve as indicators of underlying conditions, such as infections or autoimmune disorders. Understanding the different types of skin lesions is crucial for accurate diagnosis and appropriate management.

Dermatology, as a field, faces challenges in optimizing patient analysis methods, which could potentially reduce costs. Skin diseases are prevalent globally, affecting a large percentage of the population, and rank high in terms of health burden worldwide.

Dermatology stands as a pivotal discipline within medicine, its significance underscored by the sheer prevalence of skin diseases. Surpassing the collective occurrence of hypertension, obesity, and cancer, skin conditions emerge as among the most widespread human ailments.

Affecting individuals across all ages, genders, and cultural backgrounds, these conditions collectively impact a substantial portion of the population, ranging from 30% to 70% in the United States alone.

Therefore, skin diseases are an issue on a global scale, positioning on 18th in a global rank of health burden worldwide. Additionally, medical imaging emerges as a valuable resource within dermatology, given the extensive array of illnesses that the field must address.

Medical imaging, particularly in dermatology, holds promise for improving diagnostics, with the potential to replace verbal descriptions and aid in early detection. Skin cancer, including melanoma, basal, and squamous cell carcinomas, is most common in populations with predominantly white skin or in countries with high sun exposure.

In many countries, the prevalence of skin cancer presents notable challenges for healthcare systems, leading to considerable economic impact. Early detection substantially improves survival rates, yet late-stage diagnoses incur greater treatment expenses and poorer prognoses.

In the field of medical imaging and diagnosis, particularly in the domain of dermatology and oncology, the integration of deep learning methodologies has revolutionized the detection and classification of skin lesions, including those indicative of various types of cancer. Among the arsenal of deep learning models utilized for this purpose, MobileNet and Inception v3 stand out for their efficacy in analyzing medical images with high accuracy and efficiency.

The process of detecting cancerous lesions through skin lesion analysis begins with the acquisition of medical images, typically through imaging modalities like dermoscopy or digital photography. These images capture intricate details of skin lesions, ranging from color and texture to shape and size. Once obtained, these images serve as inputs to the deep learning models.

MobileNet and Inception v3, renowned for their versatility and performance, are employed to process these images. These models have been trained on extensive datasets comprising various classes of skin lesions, including benign and malignant types. Through a process known as convolutional neural networks (CNNs), these models meticulously analyze the features present within the images, extracting relevant information crucial for lesion classification.

While the overall accuracy of the project has been determined, the focus now shifts towards assessing the accuracies of individual classes.To achieve this, the plan entails implementing class-specific accuracy evaluation techniques.

Additionally, a Gradient-weighted Class Activation Map (Grad-CAM) module will be integrated for each class. Grad-CAM modules provide insights into the regions of an image that contribute most to the model's prediction for a particular class. By incorporating Grad-CAM modules, the project aims to enhance interpretability and understanding of model predictions at a granular level.

2 Literature Survey

Dhivyaa et[1] on two datasets, proposed Random Forest and Decision Tree algorithms for skin class category and model that can produce feature maps of high resolution that can be used to assist in the preservation of the image's spatial information.

Polat, Kemal, and Kaan Onur Koc5 [2] claim to have obtained very encouraging results in the identification of skin lesions and have proposed a system that uses no filtering and feature extraction.

Kumar et al.[3]introduced a system employing RGB color-space, GLCM, and Local Binary Pattern (LBP) techniques for initial processing and segmentation of images. Their methodology includes fuzzy-c clustering for further refinement, leading to promising outcomes in skin lesion identification. This system utilizes a combination of methods to enhance images and extract meaningful features, ultimately achieving encouraging results in lesion detection.

Adegun and Viriri[4]introduced a method incorporating an encoder-decoder Conditional Random Field (CRF) module to refine contours and localize lesion boundaries. They employed a linear combination of Gaussian kernels for paired edge potentials in this process. Additionally, they utilized a Fully Convolutional Network (FCN) with optimized hyper-parameters, which yielded satisfactory outcomes.

Srinivasu et al.[5] introduced data augmentation strategies aimed at balancing diverse forms of lesions within the same range of images. Their proposed model, based on LSTM and MobileNet V2 approaches, demonstrated effectiveness in both classifying and detecting skin diseases while requiring minimal computational resources. This suggests a promising solution for efficient and accurate diagnosis of skin conditions

Zhang et al.[6]showcased the efficacy of CNN optimization using an enhanced version of the whale optimization technique. This method is applied to refine the network's weights and biases, aiming to minimize the disparity between the network's output and the desired output. Their approach highlights a novel method for enhancing CNN performance through advanced optimization techniques.

Hameed et al.[7]introduced a Multi-Class Multilevel (MCML) classification technique, drawing inspiration from the "divide and conquer" strategy. Their proposed algorithm integrates both machine learning and deep learning methods, offering a comprehensive approach to classification tasks. This innovative method demonstrates the fusion of traditional and advanced techniques to enhance classification accuracy and robustness.

Mohsin Shaikh[8] proposed different deep learning models which include EfficientNets, SENet,and ResNeXt WSL . Their approach rely on multiple model input resolutions and adopt multiple cropping strategies and utilize images at various resolutions, enhancing model robustness. These models are selected through a rigorous search strategy, ensuring optimal performance and also achieve accurate results of classification.

In their study, Mahbod et al.[9] validated that image size significantly impacts skin lesion categorization performance when employing CNN transfer learning. They found that image cropping surpasses image resizing in terms of performance. Additionally, they demonstrated that the most effective classification performance is achieved through a straightforward ensembling strategy, which combines results from images clipped at six scales and three fine-tuned CNNs. This highlights the importance of considering image size and employing ensemble techniques for improved skin lesion classification accuracy

Mohammed A. Al-masni[10] proposed an consolidated diagnostic framework that combines the skin lesion boundary segmentation stage and the multiple skin lesion classification stage. They used a full resolution convolutional network to segment skin lesion boundaries from whole dermoscopic images, also they used Inception-v3, ResNet-50, Inception-ResNet-v2, and DenseNet-201 to classify segmented skin lesions. The classification performance of each classifier model in terms of accuracy was 88.05%, 89.28%, 87.74%, and 88.70%, respectively, and the final results were obtained by merging the accuracy of these classifiers.

Catarina Barata and Jorge S. Marques [11]classified skin lesions hierarchically into melanocytic and nonmelanocytic lesions based on their origins, further they were classified into benign or malignant lesions depending on the malignancy degree. The classification method used may be similar to that used for binary tree problems, and DenseNet-161 and ResNet-Inception were used as the architectures of the input image encoder. For images that underwent feature extraction through the image encoder, a long short-term memory network was used to perform the process of the image decoder to derive a hierarchical diagnosis. The final accuracy was 78.0%, and the classification accuracy between melanocytic and nonmelanocytic lesions was 92.2%.

Andre Estava[12]proposed a system that performs as well as dermatologists when it comes to identifying malignant lesions. They used a GoogLeNet Inception-V3 NN and trained it using 129,450 skin lesion images to be able to classify skin lesions as benign/malignant. The output of the system was compared with the performance of 21 different dermatologists and the outcome was favourable.

3 Problem Statement

Classifying skin lesions aids doctors and patients in many ways. One of the main ways being protection from cancerous diseases like melanoma. Early detection of such a disease leads to it being cured quickly, yet it is easily overlooked like any other lesion by the patients leading to the increase of the disease.

The problem of classifying skin lesions till date has been experimented upon by various, had results that were compelling enough, yet the usage of such a machine in a human driven environment plays out to be futile due to various reasons. One of the main reasons being the lack of trust in the system. To tackle such problem Explainable Artificial Intelligence(XAI) techniques like Grad-CAM can be used. They aid in providing the explainability factor which explains what the machine's decision process is like. It would build trust among the users.

Not using such automated devices is greatly hindering the speed of diagnosis these days. Hence by using XAI,people can be encouraged to use the machines for automation of diagnosis.

4 Proposed Methodology

The proposed methodology is a congruence of explainability using Grad-CAM and Deep learning models such as MobileNet. To achieve the desired output, the following steps are proposed.

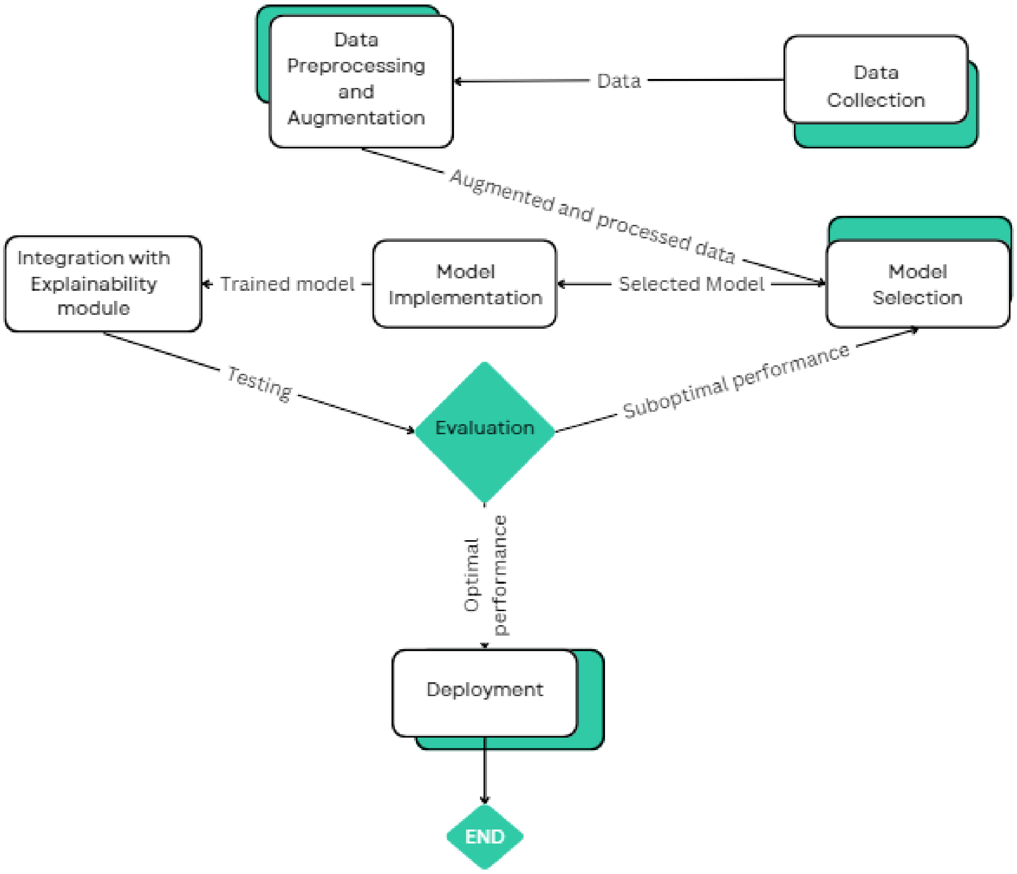

Figure 1: WorkFlow

i.Data Collection Diverse images of various skin lesions have been used. The dataset is annotated according to the labels by medical professionals hence ensuring its reliability.

1. Sample Dataset:

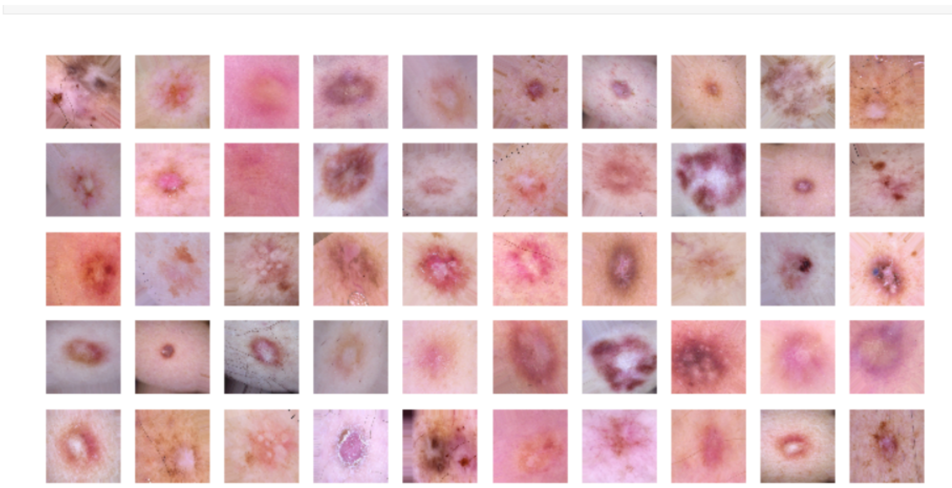

Figure 2: Sample Datset

2. Source and Size of the Dataset

   Dataset is sourced from the ISIC 2018 challenge which had 10015 dermatoscopic images carefully curated to showcase seven different skin lesion possibilities.

3. Dataset Description

   The images have been sourced from the Department of Dermatology at the Medical University of Vienna, Austria, and the skin cancer practice of Cliff Rosendahl in Queensland, Australia over a time of 20 years. They meticulously labeled and finally checked manually so as to provide accurate results. They are categorized into seven different classes, namely

   ⋆ akiec - actinic keratoses and intraepithelial carcinoma/Bowen disease
   ⋆ bcc - basal cell carcinoma
   ⋆ bkl - benign lesions of the keratosis type
   ⋆ df - dermatofibroma
   ⋆ mel - melanoma
   ⋆ nv - melanocytic nevi
   ⋆ vasc - vascular lesions

   Such images highly aid in the classification yet they are imbalanced in terms of the classes with nv class having the highest number of images and df having the least.

ii.Data Augmentation Though the dataset was robust in itself, there was a significant challenge in the balancing of classes. It was observed that there was a severe imbalance in the distribution of images given and hence we have used augmentation to increase the dataset size hence resulting in a dataset of size of around 40000.

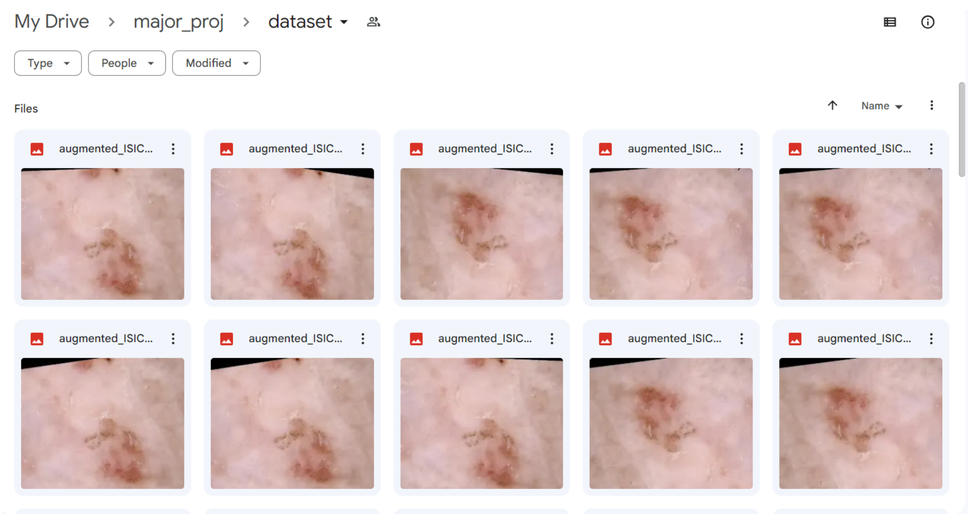

Figure 3: Augmented Dataset

iii.Adapting Transfer Learning Pretrained models such as MobileNet, Resnet,DenseNet, InceptionResnet can be used to train the current data by recognizing which one works the best for image classification, particularly for the dataset considered.

The proposed system implements MobileNet which has been selected after numerous trial and error procedures.

MobileNet is a lightweight convolutional neural network architecture made to be fast, efficient and easily deployable for mobile devices. It has 87 layers which work for a different reason to reach a cohesive output.

The layers include

⋆ An input layer through which the images are taken as the input for the model.

⋆ Convolution layers to extract the features of the image using different activation functions.

⋆ Depthwise separable convolution layers which are used to reduce the number of parameters required.

⋆ GlobalAveragePooling layer to reduce the spatial dimensions.

⋆ Dropout and Classification layers to finally provide the required output.

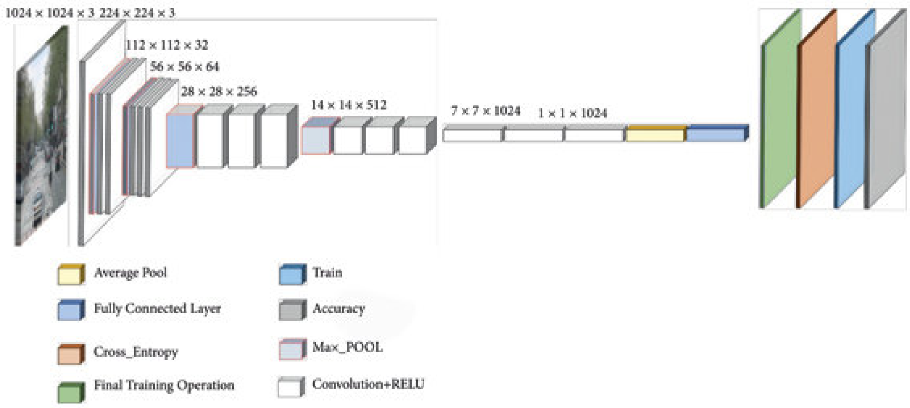

Figure 4: Overview of MobileNet

The major virtues of the architecture which led to it as the ultimate choice was how it interacted with the data as amongst the experimentation performed it showed the highest accuracy in terms of metrics and was easily trainable.

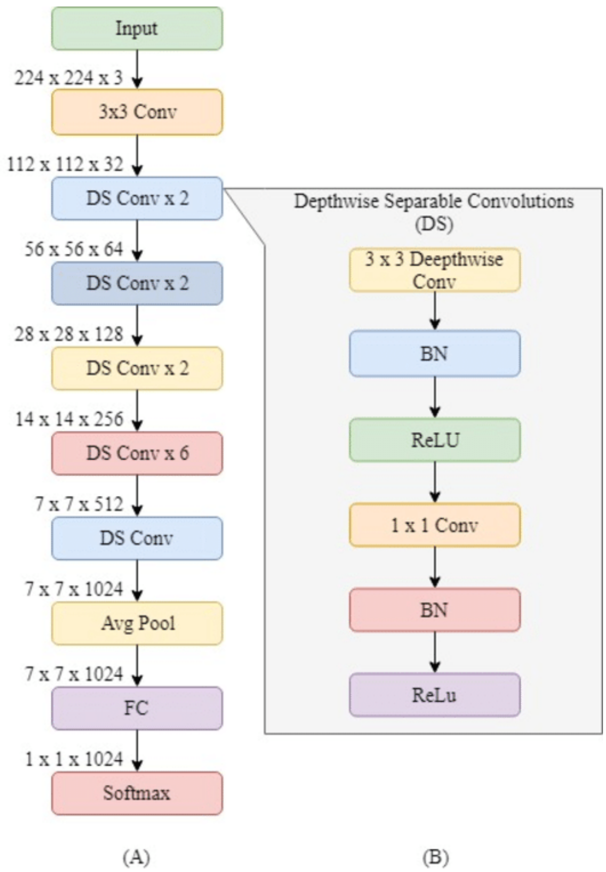

Figure 5: Architecture of MobileNet

iv.Transfer Learning integration Selected MobileNet model was initialized with the pretrained weights. Then on it was trained on the dataset to provide better results.

Training was carried on by using various methods which included

- ⋆ Fine tuning the parameters at different levels.
- ⋆ Training at different class weights for every n epochs to provide attention to the classes which seemed lacking to compensate for the class wise accuracy. It was done after examining the performance of the model towards different classes, the classweights were changed accordingly to provide more attention to the classes which had little to less accuracy to achieve a balance in the accuracies.
- ⋆ Adding a layers of Dropout and regularization to the existing architecture.

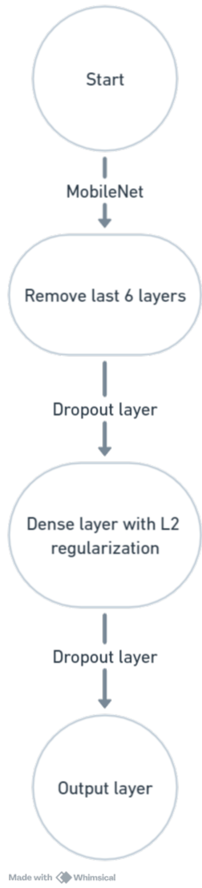

Figure 6: Modified Architecture

v.Optimization The model is then optimized using optimization techniques which include the experimenting with various hyperparameters, dropout rates,callbacks and so forth to increase the speed of convergence.

1. Callbacks:Two very important call backs were used as to achieve better results listed below.

   - ModelCheckpoint callback to save the model weights when the accuracy is better than the previous one over the course of training.
   - ReduceLROnPlateau callback to prevent over fitting by reducing the learning rate as required.

   Both of the callbacks were monitored on top-2-accuracy so as to avoid overfitting.

2. Hyperparameters and dropout rates: Hyperparameters such as batch size were monitored parallely to the dropout rates.

vi.Grad-CAM integration Explainability factor is added to the model using Gradient-weighted Class Activation Maps(Grad-CAM). It is a technique used to understand deep learning models for image classification tasks. It does so by highlighting the area in the image which led to the ultimate decision of the model.

The gradients of the score corresponding to the target class (i.e., the class of interest) with respect to the feature maps of the final convolutional layer are computed using backpropagation.Then the gradients are globally averaged indicating the it's importance to the target class.

Then it calculates the weighted combination of all the features to generate the heat map. The heat map is applied as an overlay over the existing image to exactly pin point where the model's focus is on.

This would aid the viewer to understand the model better while evaluating its performance.These visualizations can aid in increasing trustworthiness, explainability factor which would concrete the credibility to the people using it. Hence yielding to better usage and increased improvement.

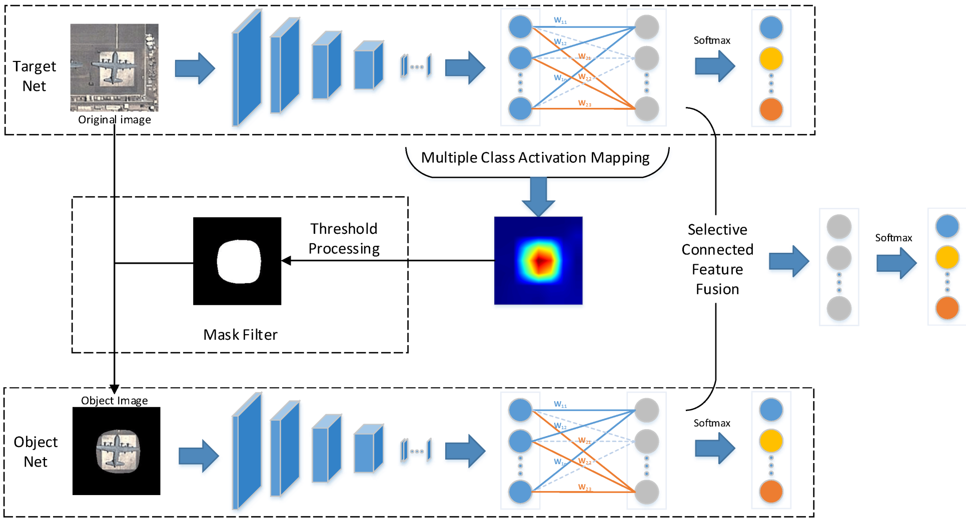

Figure 7: Class Activation Maps

vii.Metrics Metrics used to evaluate the model's performance are categorical accuracy, accuracy by class. Since one of the main mottos towards building such a model was to make the model perform equally well towards all the classes instead over fitting to a class that is most prone to be easily understood by machines, accuracy towards all the classes was considered to evaluate how the model performed. It is also evaluated on multiple test cases that would cover a diverse range of diseases.

$$\text{Categorical Accuracy} = \frac{1}{n}\sum_{i=1}^{n} 1(\text{argmax}(\text{y}_{\text{pred}}[i]) = \text{argmax}(\text{y}_{\text{true}}[i])) \tag{1}$$

where:

$$\begin{aligned} n &: \text{total number of samples} \\ \text{y}_{\text{pred}}[i] &: \text{predicted probabilities for sample } i \\ \text{y}_{\text{true}}[i] &: \text{true probabilities (one-hot encoded) for sample } i \\ 1(\cdot) &: \text{indicator function, returns 1 if condition is true, else 0} \end{aligned}$$

5 Training Details

The training happened in series of 30 epochs with each trained model analyzed to make the model's performance better. The first 30 epochs were trained with the initial class weights and it's performance was observed. Then the class weights were altered according to which class had the least accuracy when tested. Weight of such classes were increased and the others were depreciated. This carried on till three sets of the epochs until the model reached convergence at the last set of class weights.Thus the model trained for 120 epochs leveraging GPU present on kaggle platform.

Table 1: Class weights for different epochs

Epochs	akeic	bcc	bkl	df	mel	nv	vasc
1-30	3	1	3	3	3	1	3
31-90	2	1	3	3	2	1	2

6 Results and Analysis

The model achieved a precision,recall,F1-score and accuracy of 92%,92%,92%,91.79% respectively. Below are the classwise accuracies.

Table 2: Results based on classes

Lesion Name	Accuracy Achieved
akeic	65%
bcc	93%
bkl	70%
df	83%
mel	56%
nv	96%
vasc	90%

Below is the confusion matrix plotted over validation dataset.

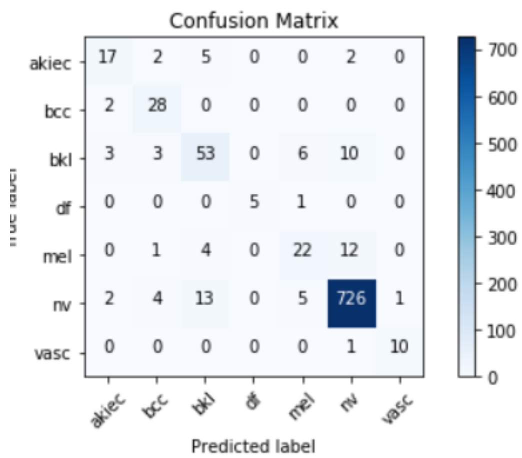

Figure 8: Confusion matrix

Understanding the above confusion matrix we can observe that more dangerous conditions such as Actinic Keratosis (AKIEC), Basal Cell Carcinoma (BCC), Melanoma (MEL), and Vascular Lesions (VASC) performed fairly well.

Once the model has been chosen it was integrated with Grad-CAM to provide the following results.

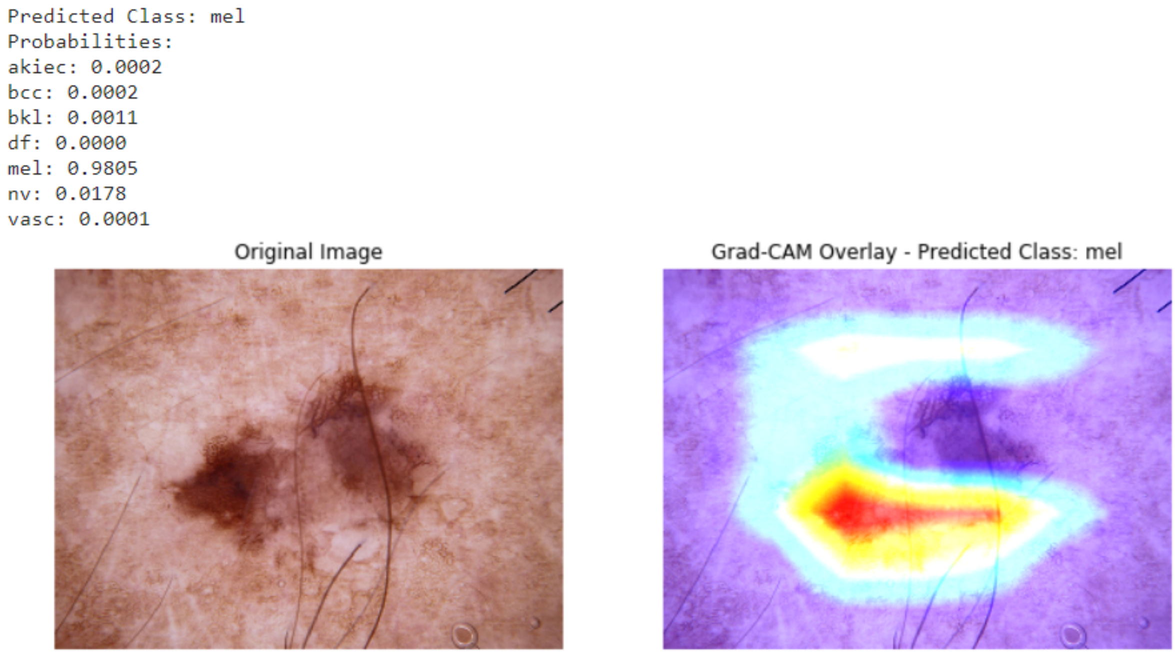

Figure 9: Grad-CAM integrated result

The above picture is of a melanoma lesion which was fed to the model, the model then classified it into it's respective class along with the heatmap overlay of Grad-CAM, in order for the user to understand where the model's attention was when predicted.

Conclusion and Future Work

In summary the proposed paper aims at creating a machine that can be easily used and trusted among people for faster diagnosis of harmful diseases that could be life threatening. Using concepts of transfer learning, it is aimed to create a model that can perform better when tested with the image of the lesion. A user interface that can integrate the user experience with diagnosis of the disease would help it reach a vast number of people. Using Grad-CAM, the trust among people is increased and so does the accountability. This helps in better human-machine interaction. Using the explanations garnered by Grad-CAM visualization, decision making is also improved as well informed decisions are made.

Looking ahead, the system has a continuous scope of improvement as more complex cases of skin lesions occur and lead to interesting deductions. The user interface can be refined so that ease of usage gets increased and lets the user interact more. Previous medical history of the patient can be compared and integrated so that it can aid in better predictions. The model itself can be modified so that it yields the best results.

References

[1]Dhivyaa, C.R., Sangeetha, K., Balamurugan, M., et al.: Skin lesion classification using decision trees and random forest algorithms. J Ambient Intell Human Com- put (2020) https://doi.org/10.1007/s12652-020-02675-8 . Received 14 August 2020; Accepted 03 November 2020; Published 17 November 2020

[2]Polat, K., Koc, K.O.: Detection of skin diseases from dermoscopy image using the combination of convolutional neural network and one-versus-all. Journal of Artificial Intelligence and Systems 2, 80–97 (2020) https://doi.org/10.33969/AIS. 2020.21006

[3]Kumar, M., Alshehri, M., AlGhamdi, R., et al.: A de-ann inspired skin cancer detection approach using fuzzy c-means clustering. Mobile Netw Appl 25, 1319– 1329 (2020) https://doi.org/10.1007/s11036-020-01550-2 . Published 08 June 2020; Issue Date August 2020

[4]Adegun, A.A., Viriri, S.: Fcn-based densenet framework for automated detection and classification of skin lesions in dermoscopy images. IEEE Access 8, 150377– 150396 (2020) https://doi.org/10.1109/ACCESS.2020.3016651

[5]Srinivasu, P.N., SivaSai, J.G., Ijaz, M.F., Bhoi, A.K., Kim, W., Kang, J.J.: Clas- sification of skin disease using deep learning neural networks with mobilenet v2 and lstm. Sensors 21(8), 2852 (2021) https://doi.org/10.3390/s21082852

[6]Zhang, L., Gao, H.J., Zhang, J., Badami, B.: Optimization of the convolutional neural networks for automatic detection of skin cancer. Open Med (Wars) 15, 27–37 (2020) https://doi.org/10.1515/med-2020-0006

[7]Hameed, N., Shabut, A.M., Ghosh, M.K., Hossain, M.A.: Multi-class multi-level classification algorithm for skin lesions classification using machine learning tech- niques. Expert Systems with Applications 141, 112961 (2020) https://doi.org/ 10.1016/j.eswa.2019.112961

[8]Gessert, N., Nielsen, M., Shaikh, M., Werner, R., Schlaefer, A.: Skin lesion classifi- cation using ensembles of multi-resolution efficientnets with meta data. MethodsX 7, 100864 (2020) https://doi.org/10.1016/j.mex.2020.100864

[9]Mahbod, A., Schaefer, G., Wang, C., Dorffner, G., Ecker, R., Ellinger, I.: Transfer learning using a multi-scale and multi-network ensemble for skin lesion classi- fication. Computer Methods and Programs in Biomedicine 193, 105475 (2020) https://doi.org/10.1016/j.cmpb.2020.105475

[10]Al-masni, M.A., Kim, D.-H., Kim, T.-S.: Multiple skin lesions diagnostics via integrated deep convolutional networks for segmentation and classification. Com- puter Methods and Programs in Biomedicine 190, 105351 (2020) https://doi.org/ 10.1016/j.cmpb.2020.105351

[11]Barata, C., Marques, J.S., Emre Celebi, M.: Deep attention model for the hier- archical diagnosis of skin lesions. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (2019)

[12]Esteva, A., Kuprel, B., Novoa, R., et al.: Dermatologist-level classification of skin cancer with deep neural networks. Nature 542, 115–118 (2017) https://doi. org/10.1038/nature21056 . Received 28 June 2016; Accepted 14 December 2016; Published 25 January 2017; Issue Date 02 February 2017