Skin Cancer Prediction Using Transfer Learning and Ensemble Learning

[1] Shailendra Bahirwar, Artificial Intelligence and Data Science, Vishwakarma Institute Of Information Technology
[2] Ms. Shalini Wankhede, Department of Information Technology, Vishwakarma Institute Of Information Technology

Abstract

Skin cancer is a prevalent and potentially deadly disease, but early detection greatly improves treatment outcomes. This study presents a deep learning approach for skin cancer prediction using various Transfer Learning and Ensemble Learning trained model which consists of MobileNet, Inception and Xception Learning on pre-trained models for prediction of Skin Cancer based on the image scans of the infected part of the body. The need for early detection of Skin Cancer has become a crucial need in the medical industry and a very demanding application in need, this article examines the use of a skin cancer software for public to alert them through early prediction.

1. Background

Skin cancer is the most commonly diagnosed cancer globally, with a rapidly increasing incidence rate, particularly in regions with high ultraviolet (UV) exposure such as Australia, North America, and parts of Europe. According to the World Health Organization, approximately 2 to 3 million non-melanoma and 132,000 melanoma skin cancers occur worldwide annually. Although skin cancer is largely preventable and highly treatable when detected early, it continues to result in significant morbidity and mortality due to delayed diagnosis and limited access to specialized care in certain regions.

Skin cancer arises from the uncontrolled proliferation of abnormal skin cells, typically triggered by DNA damage induced by UV radiation. These cancers are broadly classified into **non-melanoma skin cancers (NMSC)**—which include **basal cell carcinoma (BCC)** and **squamous cell carcinoma (SCC)**—and **melanoma**, the most aggressive and deadly form. While NMSC is more common and generally less life-threatening, melanoma poses a significantly higher risk due to its tendency to metastasize early if not identified and treated promptly.

Basal cell carcinoma (BCC), accounting for nearly 80% of NMSC cases, originates in the basal cells located at the bottom of the epidermis. It typically grows slowly and rarely spreads to other parts of the body but can cause substantial local tissue damage if left untreated. **Squamous cell carcinoma (SCC)** arises from squamous cells in the upper epidermis and, while more aggressive than BCC, has a relatively favorable prognosis when detected early.

Melanoma, which develops from melanocytes (pigment-producing cells), is less common but responsible for the majority of skin cancer-related deaths. It can occur de novo or from pre-existing moles and is often identified by changes in lesion characteristics described by the ABCDE rule: Asymmetry, Border irregularity, Color variation, Diameter (>6 mm), and Evolution over time.

Risk factors for skin cancer include prolonged exposure to UV radiation, fair skin phenotype, a history of sunburns, genetic predisposition, the presence of atypical or numerous nevi, immunosuppression, and increasing age. Notably, individuals with darker skin tones may present with skin cancer at more advanced stages due to diagnostic oversight or lack of awareness, highlighting disparities in early detection efforts.

The clinical diagnosis of skin cancer primarily involves visual inspection and dermoscopic evaluation. However, the variability in lesion appearance across individuals and overlapping visual features between benign and malignant lesions make accurate diagnosis challenging. Studies have reported significant interobserver variability among dermatologists, and the misdiagnosis of melanoma remains a concern even in specialized clinical settings.

Biopsy and histopathological examination remain the gold standards for definitive diagnosis. Nonetheless, these procedures are invasive, time-consuming, and may not always be feasible in resource-limited settings or for screening large populations. Consequently, there is a growing demand for reliable, non-invasive, and automated diagnostic tools to aid clinicians in early detection and classification of skin lesions.

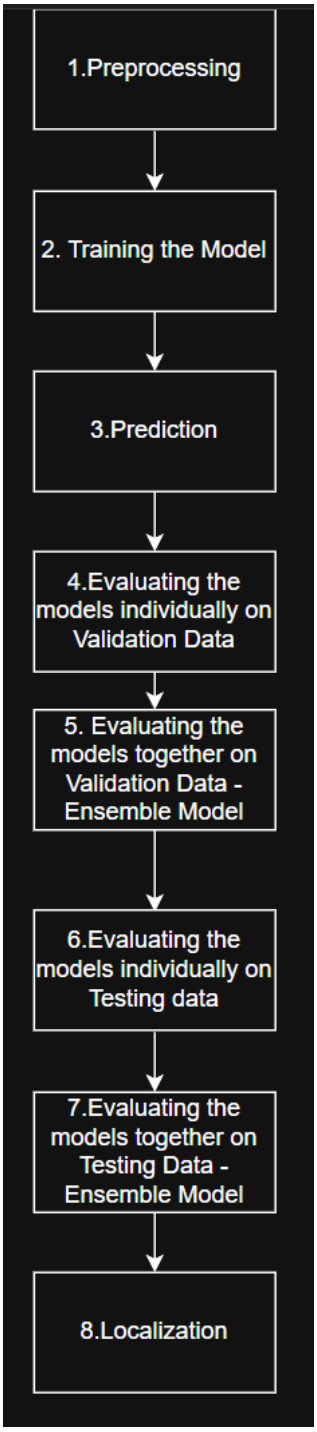
In recent years, the integration of digital dermoscopy with artificial intelligence (AI) and machine learning techniques—particularly **deep learning using Convolutional Neural Networks (CNNs)**—has shown considerable promise in addressing these diagnostic challenges. By learning complex features directly from image data, CNNs can distinguish between various types of skin lesions with accuracy levels comparable to those of expert dermatologists. This has laid the groundwork for developing computer-aided diagnostic systems aimed at improving early detection, reducing diagnostic errors, and ultimately enhancing patient outcomes.

This study builds upon these developments by exploring an ensemble approach that combines multiple CNN architectures to improve skin cancer classification performance. A deeper understanding of the medical context and diagnostic complexities associated with skin cancer is essential for developing AI systems that are clinically relevant, reliable, and scalable.

2. Architecture

The basic working and the flow of the Machine Learning Model can be seen with the help of the Architecture shown in the figure. The overall model has been divided into 10 parts to make the model more understandable to the general user and so that it is easy to be worked on in the future in order to help improve the results. The detailed working of every part of the model has been discussed below along with the flowchart of the working of the model. Some of the advantages of the Learning Model are:

- Tensors are used for faster processing
- Transfer Learning gives us the advantage of using the well-known pre-trained models which can help improve the performance on the dataset
- Due to transfer learning, the used models have been pre-trained on a general data set and needs to be trained only for the specific characteristics and since the models are deep neural networks, so it can train easily even on a smaller number of epochs which can reduce the amount of computation, processing and use of resources .
- Ensemble model is used to help improve the overall accuracy of the model by bringing together several models as one and create a new deeper architecture.
- In this work, the image file of the patient is upload into a public use soft-ware, which is a GUI-based interface, developed with the help of Tkinter and Python, and it consists of the model and the software processes the image through the model and predict the results which can help people and doctors to start with the medication way earlier instead of waiting for the laboratory tests and reports for the confirmation.



3. Methodology

PREPROCCESING

All the required libraries are first loaded into the model and then the dataset is also loaded into the model using the Google Drive as the Dataset has been uploaded to the Google Drive it can be accessed directly through it.
After the Google Drive is connected and authorized for use, the CSV files of the Training, Validation and Testing Data are read one-by-one and the data which they contain is displayed. Further, it is required for the dataset images to have their directory location attached to the filenames because then the images can be accessed directly with it without changing the directory which might interrupt any further operation.
After the directory locations are attached to the filenames, the images are loaded into the model one-by-one and they are converted into tensors. After obtaining the tensors, they are stored in a separate location so that these tensors can be used directly, and it eliminates the need for loading the dataset and processing it again and again and instead these

tensor files can just be loaded into the model and used for further training and processing. This also saves a lot of time because we do not need to create the tensors again and again as the tensors will be same each time they are being created because it is the same images which might be used during the development of the research work.

TRAINING MODEL

For this, we have used Transfer Learning and Ensemble Modelling. For the Trans-fer Learning, we have used the following pre-trained models:

- Mobile Net Architecture
- Inception V3 Architecture
- Xception Architecture

For the training, the pre-trained models are first loaded in as a function and then a since we have a checkpoint of these models, so we can start training these models from that checkpoint again. We start training it further by fitting it on the data which we have using the tensors.

PREDICITON

This prediction is done on a sample image just to check that all the models are working and are predicting. Working of this section denotes that the model can be taken to the further stage of evaluation of the validation data and the testing data. For this part, we take the sample image and convert that image into a tensor. After that, the Individual Models and their respective weights are loaded so that the pre-diction can be made using the predictive function which we have defined. This is the function which will also be used for the Interface Application for the prediction on the image which will be uploaded by the user.

EVALUATING THE MODELS INDIVIDUALLY ON VALIDATION DATA

For the evaluation of the models on the Validation Data, we have used the Receiving Operating Characteristic Curve. A Receiver Operating Characteristic Curve, or ROC curve, is a highly useful method to obtain the graphical plot that illustrates the excellent diagnostic ability of a Binary Classifier system as its discrimination thresh-old is varied.

For this, we will be computing the test set predictions and then the model will be evaluated based on the Receiving Operating Characteristic Curve whose values will be obtained using the Confusion Matrix. Then, we will plot the values obtained for this to obtain the graph and check for the best possible values which can be used for prediction.

Further, after obtaining all these values, we will be calculating the Precision, Recall, Specificity, Sensitivity, Negative Predictive Value and Accuracy for further evaluation and after this, we can check the parameters, and if possible, we can tune the parameters to obtain better results.

At the end, we can compare the test results for the pre-trained architecture, as we have noted that the loss had reduced to the most efficient values possible and if we will train the model further with more epochs on the architectures, then we might make the model over fit on the data which can decrease the results and might lead to a higher number of wrong predictions, furthermore reducing the accuracy of the model and its efficiency.

EVALUATING THE MODELS TOGETHER ON VALIDATION DATA – ENSEMBLING THE MODELS

Defining a specific input shape for all the input images. The models will be brought together and appended so that the image passes through all the models one-by-one. The ensemble model will be defined which will contain all the models and the then the weights of the model will be combined, and a file will be generated so that the weights are together into one and then they can be used together for the complete architecture.

For evaluating the models, we will be using the test set predictions and the ROC curve will be used whose computational values will be obtained using the Confusion Matrix. The other evaluation metrics are Precision, Re-call, Sensitivity, Specificity, Negative Predictive Value and Accuracy. The results obtained for the Ensemble Model are better than each architecture individually.

EVALUATING THE MODELS INDIVIDUALLY ON TESTING DATA

For the evaluation of the models on the Testing Data, we have used the Receiving Operating Characteristic Curve. A Receiver Operating Characteristic Curve, or ROC curve, is a highly useful method to obtain the graphIcal plot that illustrates the excellent diagnostic ability of a Binary Classifier system as its discrimination thresh-old is varied.

For this, we will be computing the test set predictions and then the model will be evaluated based on the Receiving Operating Characteristic Curve whose values will be obtained using the Confusion Matrix. Then, we will plot the values obtained for this to obtain the graph and check for the best possible values which can be used for prediction.

Further, after obtaining all these values, we will be calculating the Precision, Recall, Specificity, Sensitivity. Negative Predictive Value and Accuracy for further evaluation and after this, we can check the parameters, and if possible, we can tune the parameters to obtain better results.

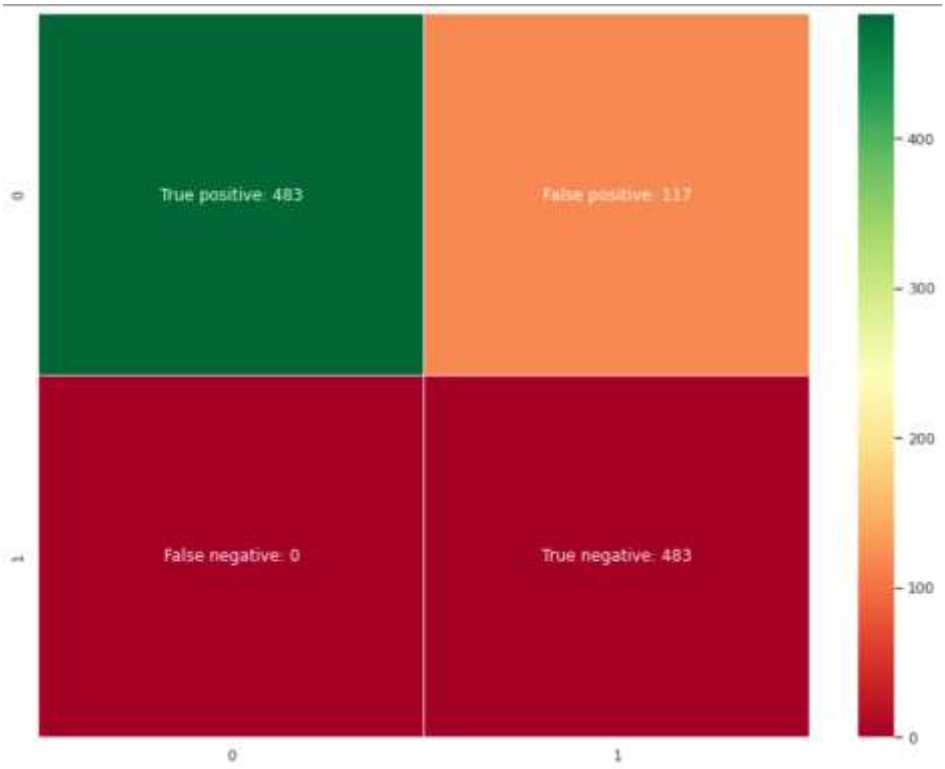
At the end, we can compare the test results for the pre-trained architecture, as we have noted that the loss had reduced to the most efficient values possible and if we will train the model further with more epochs on the architectures, then we might make the model over fit on the data which can decrease the results and might lead to a higher number of wrong predictions, furthermore reducing the accuracy of the model and its efficiency.

The results obtained on the Testing data are very similar to the results obtained on the Validation data.

EVALUATING THE MODELS TOGETHER ON TESTING DATA – ENSEMBLING THE MODELS

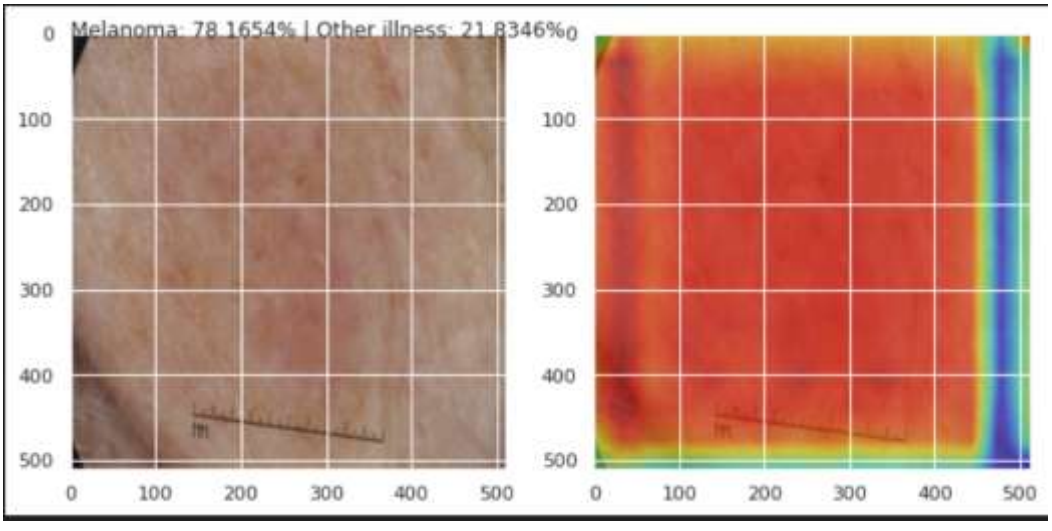
Defining a specific input shape for all the input images. The models will be brought together and appended so that the image passes through all the models one-by-one. The ensemble model will be defined which will contain all the models and the then the weights of the model will be combined, and a file will be generated so that the weights are together into one and then they can be used together for the complete architecture.

For evaluating the models, we will be using the test set predictions and the ROC curve will be used whose computational values will be obtained using the Confusion Matrix. The other evaluation metrics are Precision, Re-call, Sensitivity, Specificity, Negative Predictive Value and Accuracy. The results obtained for the Ensemble Model are better than each architecture individually as given in figure.



LOCALIZATION

For this, we define the Class Activation Map and then we create another function using which we can display the results obtained on a particular file, as we can use the address of the that image directly and then plot to even check the results obtained for the model which we have created. The results are displayed along with the plot which is created containing the Image and another heatmap image displayed along.



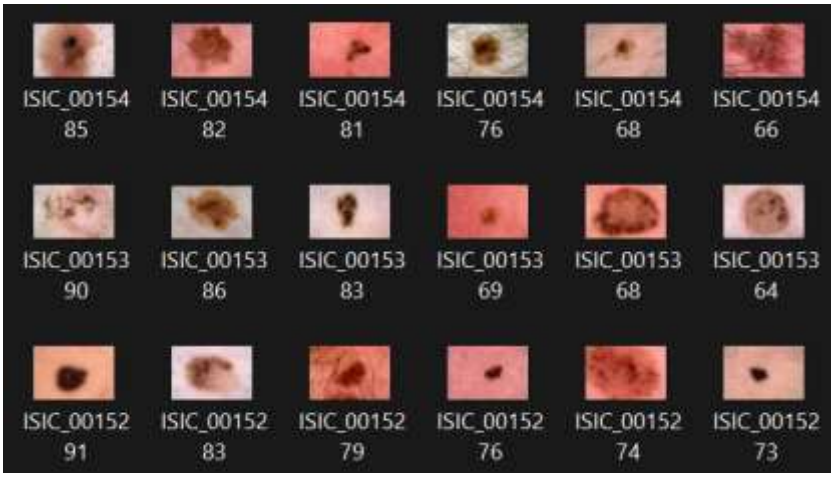
4. Results and Discussions

DATA SELECTION

The goal of the dataset selected is to help develop an image analysis that will help ensemble automated diagnosis of melanoma from the dermascopic images. Image analysis of the skin lesions consists of 3 parts :

- Lesion Segmentation
- Detection and Localization of Visual Dermascopic Features/Patterns
- Disease Classification

The dataset provides training data (150 images), and blind held-out test dataset (~600 images) will be provided to generate automated results .



COMPARATIVE STUDY

The comparative analysis of various skin lesion model is presented in the below table and from what we infer that ensemble models provides a good accuracy when compared to the CNN models.

Evaluation Type	Architecture	Confusion Matrix				Sensitivity/ Recall	Specificity	Precision	Negative Predictive Value	Accuracy
		True Positive	False Positive	True Negative	False Negative					
Individual Models on Validation Data	Mobile Net	120	30	0	0	1	0	0.8	0	0.8
	Inception	120	30	0	120	1	0	0.8	0	0.8
	Xception	120	30	0	120	1	0	0.8	0	0.8
Ensemble Model on Validation Data	Ensemble	120	30	0	120	1	0	0.8	0	0.8
Individual Models on Test Data	Mobile Net	483	117	0	0	1	0	0.805	0	0.8
	Inception	483	117	0	483	1	0	0.805	0	0.8
	Xception	483	117	0	483	1	0	0.805	0	0.8
Ensemble Model on Test Data	Ensemble	483	117	0	483	1	0	0.805	0	0.8

5. Conclusions

From the above study it is clear that an alarming disease for any human kind is a skin cancer. An early diagnosis in the most important nowadays to reduce the Melanoma skin cancer cases, as it leads to increasing in death rate. The appropriate system for early detection is encourageable and the above discussed ensemble learning plays a vital role in detection the cancerous cells with highest accuracy when compared to the other models

6. References

1. Masood, A.A. Al-Jumaily, Review article computer aided diagnostic support system for skin cancer: a review of techniques and algorithms , Int J Biomed Imaging (2013)

2. A. Esteva1, B. Kuprel1, R.A. Novoa, J. Ko, S.M. Swetter, H.M. Blau, S. Thrun, Dermatologist-level classification of skin cancer with deep neural networks, Nature (2017), pp. 115-126

3. N. Codella1, J. Cai1, M. Abedini, R. Garnavi, A. Halpern, J.R. Smith1, Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images Springe LNCS (2015)

4. T.J. Brinker1, A. Hekler1, J.S. Utika, N. Grabe, D. Schadendorf, J. Klode, C. Berking, T. Steeb, A.H.E.C. von Kalle1 Skin cancer classification using convolutional neural networks: systematic review J Med Internet Res (2018), Article e11936

5. Hasan, N., Nadaf, A., Imran, M. *et al.* Skin cancer: understanding the journey of transformation from conventional to advanced treatment approaches. *Mol Cancer* **22**, 168 (2023).

6. Stanley R.J., Stoecker W.V., Moss R.H., Gupta K., Jella P. Color and Structural Features for Automatic Skin Lesion Discrimination in Dermoscopy Images; Proceedings of the 6th World Congress on Melanoma; Vancouver, BC, Canada. 6–10 September 2005.

7. Khan M.Q., Hussain A., Rehman S.U., Khan U., Maqsood M., Mehmood K., Khan M.A. Classification of Melanoma and Nevus in Digital Images for Diagnosis of Skin Cancer. IEEE Access. 2019;7:90132–90144. doi: 10.1109/ACCESS.2019.2926837.

8. P. Sabouri, H. GholamHosseini ,Lesion border detection using deep learning .Congress on evolutionary computation (CEC), IEEE, Vancouver, BC, Canada (24–29 July 2016)