International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 03 | Mar 2022 www.irjet.net p-ISSN: 2395-0072

Multi-Class Skin Disease Classification using Pre-Processing and Multi-Model Ensemble Techniques

Bhavya Bipin Gada[1], Shivangkumar Gandhi[1], Pruthvi Rathod[1], Amruta Sankhe[2]

[1]Dept. of Information Technology, Atharva College of Engineering, Maharashtra, India
[2]Assistant Professor, Dept. of Information Technology, Atharva College of Engineering, Maharashtra, India

---

Abstract - *Many cancer cases early are misdiagnosed, resulting in severe consequences, including the patient's death. In this project, both the above problems are addressed using deep neural networks and transfer learning architecture. This paper proposes a novel system automatically detecting and classifying Skin Diseases into seven classes. Publicly available HAM10000 databases provided by Harvard were used to train and test the model. First, the proposed system uses the U-Net Model for Semantic Image Segmentation followed by a state-of-the-art Transfer Learning InceptionV3 model. The U-Net predefined model produced a validation accuracy of 94.88%. The masked image predicted from this model is then merged with the original image, converting it into a disease-focused image. The proposed system then extracts 52 features from the state-of-the-art predefined InceptionV3 transfer learning model and combines them with other metadata (CSV file features). Resampling techniques are considered for balancing the dataset, producing 4,200 images contributed from 7 disease classes equally. In the last stage, the system uses an ensemble model combination of metadata and features extracted from InceptionV3 and XgBoost Classifier to predict seven classes based on 74 features. Our model achieves an accuracy of 0.95, precision 0.95, recall 0.95, F1 score 0.95, and ROC- AUC 0.99, which is better than the previous state-of-the-art approaches. The individual class f1 scores were: akiec - 0.96, bcc - 0.96, bkl - 0.93, df - 1.00, mel - 0.92, nv - 0.90, vasc - 0.99.*

Key Words: **Multi-Model Ensemble**, **ResNet50, InceptionV3 and Xception**, **Data Point Extraction**, **Ensemble Data Points, InceptionV3+XGBoost Classifier**

1. INTRODUCTION

Skin diseases are more common than other diseases. In general, skin diseases are chronic, infectious, and sometimes may burgeon to be carcinoma. Therefore, skin diseases must be diagnosed during the early stages to slow down their development and spread. However, skin disease detection is a complicated process, and sometimes, even dermatologists (skin specialists) may find it difficult to diagnose it. The advancements in lasers and photonics-based medical technology have made it possible to diagnose skin diseases far more quickly and accurately. But the cost of diagnosis remains expensive. Therefore, we propose an image processing-based approach to diagnose skin diseases. This method takes the digital image of the disease affecting the skin area then uses image analysis to identify the type of disease. The proposed approach is swift, simple, and inexpensive as it only requires a camera and a processor to diagnose, both of which are present in a mobile phone.

2. LITERATURE REVIEW

In [1], fully Convolutional networks like InceptionV3 and MobileNet were used as standard classification transfer learning techniques on the MNIST HAM10000 dataset to detect skin diseases which achieved an accuracy of 72% with InceptionV3 and 58% with MobileNet. Dermnet research dataset was accumulated in [2], which produced 72.2 % average data accuracy using AlexNet Convolutional Network. The segmentation and classification approach was used in [3] on ISBI 2017 dataset was able to generate an IOU of 74.67% on the UNET segmentation model and 91.3% on the FCRN Classification algorithm with a recall of 0.22.

[4] studied various classification models and data preprocessing techniques that included Noise Removal techniques and GrayScale conversion, discovering accuracies ranging from 80%-89% amongst different available datasets. [5] addresses a new framework by fine-tuning layers of ResNet152 and InceptionResNet-V2 models with a triplet loss function. This framework, first, learns the embedding from input images into Euclidean space by using deep CNN ResNet152 and InceptionResNet-V2 model. It classifies the input images using L-2 distances to learn the discriminative features of skin disease images using the triplet loss function. Human face skin disease images used in this framework are acquired from the Hospital in Wuhan, China.

The [6] system classifies skin lesions into benign or malignant lesions based on a novel regularizer technique. A binary classifier discriminates between benign or malignant lesions. This achieved an average accuracy of 97.49%. The performance of CNN in terms of AUC-ROC with an embedded novel regularizer was tested on multiple use cases. [6] The area under the curve (AUC) achieved was 0.77 for nevus against melanoma and 0.93 for seborrheic keratosis v bcc. Also, seborrheic keratosis v melanoma had obtained 0.85, whereas solar lentigo v

melanoma achieved 0.86. Ideas from each of the above-mentioned papers led us to generate a methodology which is mentioned below.

3. METHODOLOGY

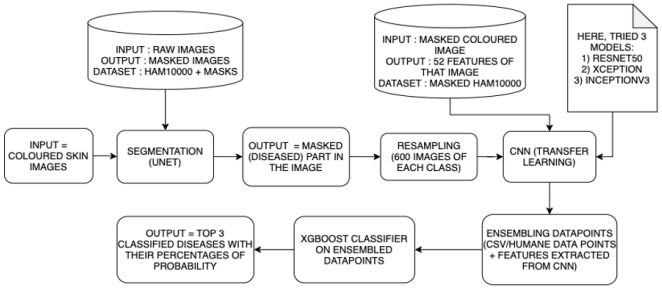

Fig-1: System Design

The problem is formulated into four main stages:

i. Segmentation Model: U-net-based architecture is implemented for precise segmentation of the image data from the HAM 10000 images. The model is evaluated over Intersection Over Union (IOU)

ii. Classification Model: The Segmented skin images are then passed to the classification model, the Inception V3 transfer learning algorithm. It is measured in terms of accuracy.

iii. Feature Extraction and ensembling Data Points: 52 image features are extracted from the InceptionV3 model for any given image. This is achieved by pulling out the last seven layers of the neural network. The extracted image features are then merged and mapped according to the data points of the metadata in the form of CSV in training. While testing, the same parameters are achieved from the data the user enters while uploading the image.

iv: XgBoost Classification: The obtained ensembled data points are used to train the XgBoost model. Predicting top 3 diseases, with an accuracy of 95%.

4. DATA-SET AND RESOURCES

The HAM10000 dataset is a collection of dermatoscopic images of skin lesions, found in the Harvard Dataverse [7]. The dataset consists of 10,015 images that can be used for academic deep learning purposes. It is a set of dermatoscopic images acquired and stored by different modalities from different populations. Cases include a representative collection of all critical diagnostic categories in the realm of pigmented lesions: Actinic keratoses and intraepithelial carcinoma / Bowen's disease (akiec), basal cell carcinoma (bcc), benign keratosis-like lesions (solar lentigines / seborrheic keratoses and lichen-planus like keratosis, bkl), dermatofibroma (df), melanoma (mel), melanocytic nevi (nv) and vascular lesions (angiomas, angiokeratomas, pyogenic granulomas and hemorrhage, vasc).

More than 50% of lesions are confirmed through histopathology (histo), the rest of the cases is either follow-up examination (follow_up), expert consensus (consensus), or confirmation by (confocal). In addition, the dataset includes lesions with multiple images, which the lesion_id-column within the HAM10000_metadata file can track.

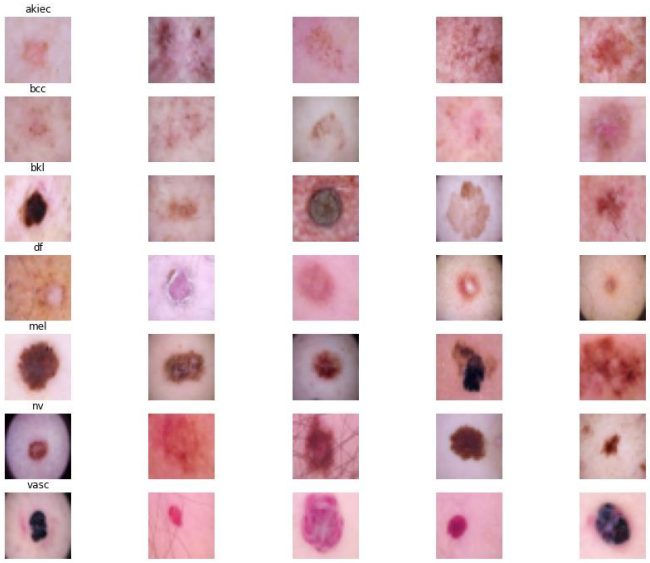

Fig-2: HAM10000 skin lesions

5. IMPLEMENTATION

Overall implementation is a superset of Pre-processing, Model training, comparison between models, model evaluation and finally, GUI.

A. DATA PRE-PROCESSING

Pre-processing of the training and validation data is performed in three parts, data segmentation, data resampling and data resizing.

i. Data Segmentation for UNET

Data segmentation targets accurate division or partition of the skin lesion (diseased area) from the actual dermoscopic image, i.e., division into multiple segments like diseased and healthy areas of skin. This comes under pre-processing and forms a black mask around the healthy part of the skin. This results in hiding inconsistencies like noise and intense colour or illumination effects which reduces artefacts and provides better training. The system uses UNet architecture to segment the input images to obtain a masked image.

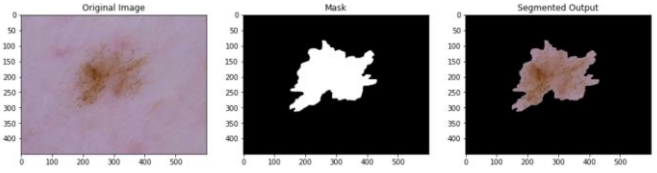

Fig-3: Segmentation of image using U-Net

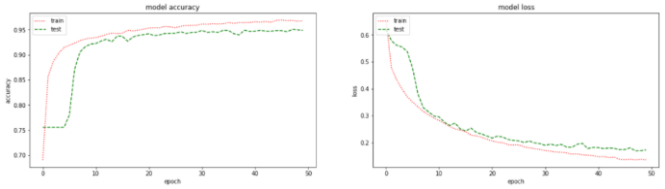

Fig-4: U-Net Segmentation Model Accuracy and Loss

In Fig-3, the first image from the left, titled "Original image", would be the image uploaded by the patient. The middle image titled "Mask" is the predicted mask that covers the healthy part of the skin for better training. Finally, the last image titled "Segmented Output" is the output of the segmentation process used for training. The UNet Segmentation model generates a masked image, concentrating only on the affected part and covering the surrounding skin.

This is a binary classification model and is able to generate masks for test images. The model has an accuracy of 93.74%, with a training accuracy of 97.37%. The f1 Score ROC-AUC curve values are 0.94 each, and the Validation accuracy was 94.88% as shown in Fig-4.

ii. Resampling

The dataset is highly imbalanced, with most data of Melanocytic nevi, Melanoma, and Benign keratosis-like lesions. The system uses resampling techniques to balance the data, reducing the dataset to 600 images per class with techniques of UpSampling and DownSampling. On purpose, the training images were not cleaned, thus still containing some noise. This comes mainly in the form of intense colours, illumination effects, and sometimes wrong labels. All images were rescaled to have a maximum side length of (256,192) pixels for the UNet Model.

Table-1: Dataset Size Comparison after Resampling

Class Data	Initial	After Resampling
Melanocytic nevi	6705	600
Melanoma	1113	600
Benign keratosis-like lesions	1099	600
Basal cell carcinoma	514	600
Actinic keratoses	327	600
Vascular lesions	142	600
Dermatofibroma	115	600

iii. Image Resizing

Resizing images is an essential pre-processing step in training machine learning models. Models train faster on smaller images, but shrinking can cause the deformation of features and patterns inside the image. Additionally, various deep learning architectures need images of the exact dimensions even though the raw collected images may be different in size. For example, the original dermoscopic image in the HAM10000 dataset has a size of 600x450 pixels in the RGB format. The images are then resized to 128x128 pixels to all layers.

B. PRETRAINED MODEL

While working with pre-trained models like InceptionV3, ResNet50 and Xception, as our transfer learning models, we had images in our dataset divided into the ratio of Training:Testing:Validation set, which was precisely 2520:1050:630 summing up to a total of 4200 images used in the experiment. The data arrays for training and testing were created using train_test_split function of the Sklearn model.

Batch Size was determined based on the number of data trains(training) that gave the best result when spread across the neural network in each iteration (train steps). In this training process, the batch size was defined as 256, which means that each step was to be spread to 256 data train to the neural network. The input images were in (128,128,3) shape and were run for 50 epochs for all models. Outputs were generated by a combination of different layers like convolutional layers, activation layers, pooling layers, batch normalization layers, and concatenation layers, summing up to 59 layers that generated output based on several parameters compared in Table-2.

C. CLASSIFICATION MODELS COMPARISON

The three models in consideration, namely InceptionV3, ResNet50 and Xception, were then compared based on multiple areas of attention like the number of parameters the models generated and a comparison of trainable and non-trainable parameters, in Table-3, the accuracy and loss graphs of each of the models were investigated along with the ease of learning taken into account, in Fig-5,6,7, and finally, the classification results were examined individually for all evaluation parameters, in table-3.

The study showed little difference between the models when compared to the entire spectrum. A contrasting result where ResNet50 was the highest amongst all others

in terms of trainable parameters, particularly difficulties in learning when time and GPU utilization were taken into account. It also had a non-monotonous testing accuracy and loss graph contradicting its monotonous training accuracy and loss graph. When such parameters were considered for comparison of Xception model, it fell back at mostly all starting from classification average accuracy of only 0.788.

Although InceptionV3 had a classification accuracy of 0.003 less than that of Resnet50, its learning was smoother and less computationally expensive than Resnet50, with a minimal difference in the ratio of non-trainable parameters. Hence, it was taken into account that InceptionV3 is the best transfer learning algorithm for classifying the seven skin diseases into consideration among all three models in the picture.

Table-2: Comparison of Classification Models based on a number of parameters.

Pre-Trained Model	Parameters		
	Total	Trainable	Non-Trainable
ResNet50	40,441,611	40,387,075	54,536
InceptionV3	26,073,771	26,037,923	35,848
Xception	30,375,347	30,319,403	55,944

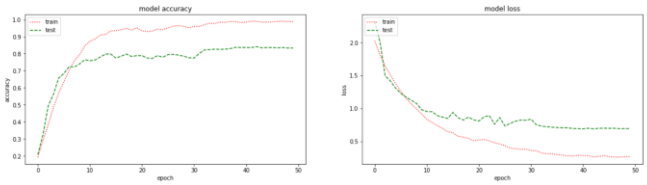

Fig-5: Inception V3 Accuracy and Loss

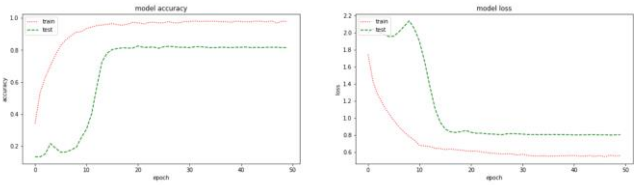

Fig-6: ResNet50 Accuracy and Loss

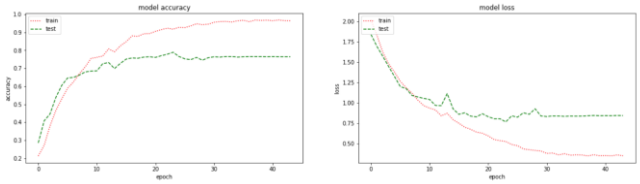

Fig-7: Xception Accuracy and Loss

Table-3: Classification Results of All Models
A = Accuracy, B = Precision, C = Recall, D = F1 Score, E = ROC-AUC, F = Test Accuracy, G = Test Accuracy

Pre-Trained Model	A	B	C	D	E	F	G
ResNet50	0.81	0.81	0.81	0.81	0.81	0.815	1.000
InceptionV3	0.81	0.81	0.81	0.81	0.81	0.812	1.000
Xception	0.79	0.79	0.79	0.79	0.79	0.788	1.000

D. Feature extraction and merging CSV data points

Once it was established that InceptionV3 was the model to choose for further research, we moved to the next part of our approach, which was feature extraction. InceptionV3 had 59 layers in our experiment. Out of which, the last seven layers were excluded from extracting 52 vital features the model analyzed from the image. These 52 features were ensembled with 22 data points from the MetaData(CSV file). These included skin type, age, skin localization and sex as well. This resulted in a dataset of 74 data points in which a few data points like gender, localization, cell type was hot encoded to further pass them for boosting.

E. Training Model (Inception V3 + XgBoost)

XGBoost, which itself is a scalable ensemble technique based on gradient boosting [10], was chosen to boost the ensembled data-point set and to classify the images into 7 categories at the question, namely Vascular lesions, Basal cell carcinoma, Benign keratosis-like lesions, Actinic keratoses, Melanoma, Dermatofibroma, Melanocytic nevi.All the data points were combined and taken as X, and the labels(name of diseases for classification) were taken as Y. This Multi-Model Ensembled approach led to a surprising accuracy of 0.95 f1-score which is seen in Table-4.

6. RESULTS

Finally, the model was evaluated on parameters like Precision, Recall, F1-Score and Support which was then compared with the same evaluation parameters achieved from the InceptionV3 model, calculated without ensemble and boosting. Before boosting, the Accuracy calculated by f1-score was 0.81 which went to a surprising score of 0.95 when boosted with the ensembled data points.

Table-4: Classification Report of:
X=InceptionV3 model
Y=InceptionV3 + XgBoost model

/	Precision		Recall		f1-score		support	
	X	Y	X	Y	X	Y	X	Y
akiec	0.81	0.97	0.85	0.95	0.83	0.96	156	121
bcc	0.83	0.93	0.84	0.98	0.84	0.96	156	117
bkl	0.64	0.94	0.75	0.91	0.69	0.93	146	112
df	0.91	1.00	0.97	1.00	0.94	1.00	151	120
mel	0.77	0.90	0.66	0.94	0.71	0.92	139	127
nv	0.75	0.93	0.62	0.87	0.68	0.90	152	122
vasc	0.97	0.98	0.99	1.00	0.98	0.99	150	121
Accuracy	-	-	-	-	0.81	**0.95**	1050	840
Macro avg	0.81	0.95	0.81	0.95	0.81	0.95	1050	840
Weighted avg	0.81	0.95	0.81	0.95	0.81	0.95	1050	840

7. GUI

The model was executed in Graphical User Interface (GUI) form using [8] Streamlit library, which is a python library that allows developers to make data-based web applications quickly and easily. Along with Streamlit, the project hosted the Web Application using [9] remote.it which is a SaaS, allowing remote users to user your web applications via an autogenerate URL. It was decided to use [9] since it is safer and better than ngrok.

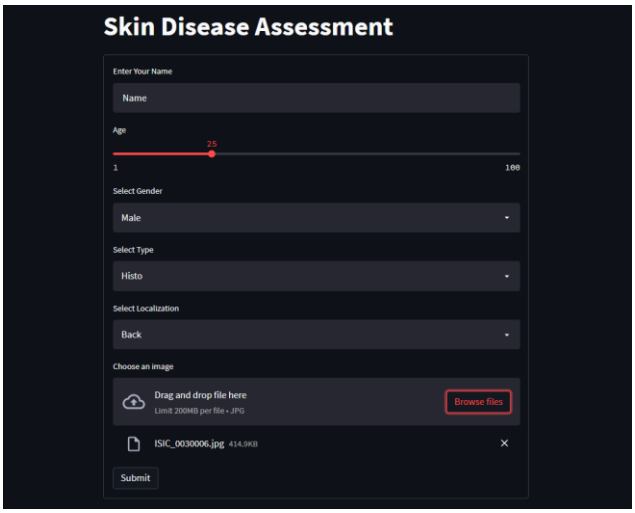

Fig-8: Taking Inputs from the user

This Fig-8 shows the GUI of the skin disease classification system and accepts user data along with the test image to classify. Whereas, Fig-9 shows the GUI of the skin disease classified system and generates a report with the Top 3 probable diseases along with segmented images

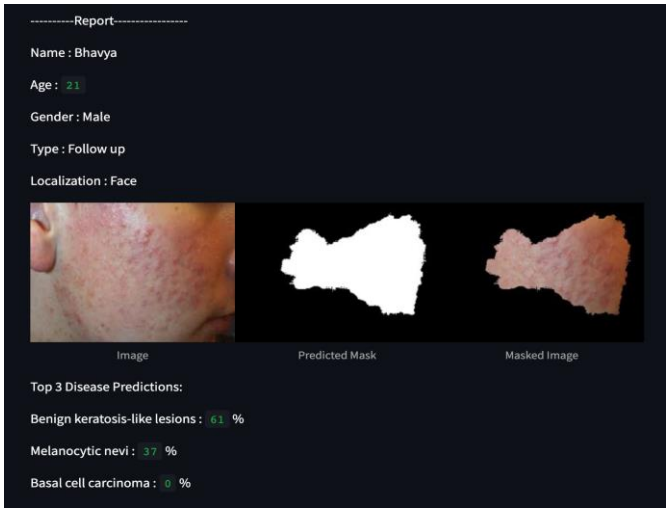

Fig-9: Generating the expected output

8. CONCLUSION

In Conclusion, the proposed solution to the problem at hand considers all the parts of the detection issues and selectively resolves them successfully. The paper presents the ideation of an ensembled multi-model approach to the HAM10000 dataset for skin disease detection. The successful implementation of the system, evaluation parameters, and execution via GUI is our active experimentation in defence of the proprietary pipeline.

9. REFERENCES

[1] K. E. Purnama et al., "Disease Classification based on Dermoscopic Skin Images Using Convolutional Neural Network in Teledermatology System," 2019 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM), Surabaya, Indonesia, 2019, pp. 1-5, doi: 10.1109/CENIM48368.2019.8973303.

[2] T. Shanthi, R.S. Sabeenian, R. Anand, Automatic diagnosis of skin diseases using convolution neural network, Microprocessors and Microsystems, Volume 76, 2020,103074, ISSN 0141-9331, https://doi.org/10.1016/j.micpro.2020.103074.

[3] V. B.N., P. J. Shah, V. Shekar, H. R. Vanamala and V. Krishna A., "Detection of Melanoma using Deep Learning Techniques," 2020 International Conference on Computation, Automation and Knowledge Management (ICCAKM), Dubai, United Arab Emirates, 2020, pp. 391-394, doi: 10.1109/ICCAKM46823.2020.9051495.

[4] L. -F. Li, X. Wang, W. -J. Hu, N. N. Xiong, Y. -X. Du and B. -S. Li, "Deep Learning in Skin Disease Image Recognition: A Review," in IEEE Access, vol. 8, pp. 208264-208280, 2020, doi: 10.1109/ACCESS.2020.3037258.

[5] B. Ahmad, M. Usama, C. Huang, K. Hwang, M. S. Hossain, and G. Muhammad, "Discriminative Feature Learning for Skin Disease Classification Using Deep Convolutional Neural Network," in IEEE Access, vol. 8, pp. 39025-39033, 2020, doi: 10.1109/ACCESS.2020.2975198.

[6] M. A. Al Bahar, "Skin Lesion Classification Using Convolutional Neural Network With Novel Regularizer," in IEEE Access, vol. 7, pp. 38306-38313, 2019, doi: 10.1109/ACCESS.2019.2906241.

[7] Tschandl, Philipp, 2018, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions", https://doi.org/10.7910/DVN/DBW86T, Harvard Dataverse, V3, UNF:6:/APKSsDGVDhwPBWzsStU5A== [fileUNF]

[8] Streamlit, tvst (2019) Streamlit [Source Code] https://github.com/streamlit/streamlit.git

[9] remot3.it, accessed on 10th February, 2022, < https://app.remote.it/#devices>.

[10] Bentéjac, Candice & Csörgő, Anna & Martínez-Muñoz, Gonzalo. (2019). A Comparative Analysis of XGBoost.

[11] https://www.kaggle.com/code/shivanggandhi/inceptionv3-xgboost-classification/notebook

[12] https://www.kaggle.com/code/shivanggandhi/unet-segmentation/notebook