SKIN DISEASE RECOGNITION AND CLASSIFICATION THROUGH MOBILE CAMERA USING MACHINE LEARNING AND IMAGE PROCESSING TECHNIQUES

SKIN DISEASE IMAGE CLASSIFICATION

MASTER'S THESIS

DECEMBER 2023

AKASH DAS

CONTENT

INTRODUCTION

Skin, the largest organ can affect due to UV, dust, fungi or virus attack. Some skin diseases can mature very fast and can become life threat. Due to inter diseases similarity, very experienced doctor required to identify correct disease. Also in remote areas experienced medical professional is in scares.

With advancement of AI, Its become possible and easier to classify images correctly using CAD. So this study aims to create an algorithm to classify few specific skin disease images accurately.

With automated disease classification, it will help in experienced resource scarcity situation. Also with easy internet and mobile device availability, it can support remote areas and save more skin disease patients.

LITERATURE REVIEW

With advancement of AI, many researchers have used different approaches to solve skin diseases image classification issue. Some authors have explored binary classification, others used multiclassification approaches. Some prominent process as follows.

1. **MACHINE LEARNING APPROACH** : Use of mathematical modelling and computer application to do feature engineering to extract image features and use of machine learning classification methods to classify.
2. **DEEP NEURAL NETWORK** : I. Transfer learning-based techniques. II. Custom CNN. III. Hybrid models.
3. **IMBALANCED DATASET AND LIMITED DATA** : style-based GAN, Application of proportionate weights to different samples in the loss function, custom mini batch aur data augmentation.
4. **GENERALIZATION ABILITY ACROSS DOMAIN AND GEOGRAPHY** : Researchers started enriching data by adding multiple ethnic groups, sample image collection using handheld device. Mixing different image types such as dermoscopic and histological images.
5. **MULTIMODAL FUSION** : Learning from different model or data source and combine their findings using a set of criteria. Simple majority voting rule, merging image and patient's clinical data
6. **FASTER MODELS SUITABLE FOR EDGE DEVICES** : Lighter mobile deep learning models, Parameter or weight pruning techniques or Knowledge distillation using student teacher network.
7. **EFFECTIVENESS AGAINST ADVERSARIAL ATTACKS AND OTHER NOISES**: saving models from black box and white box attacks using FGSM and PGD methods

PROBLEM STATEMENT

- According to available research works, most of the research were more focused on binary classification to identify one particular class or explored complex network to achieve higher accuracy.
- However, those systems might hinder the effort to democratize the remote skin disease identification system for rural area where experienced dermatologist and highly skilled diagnostician is not available.
- ✓ Therefore, it is essential to create an automatic skin cancer classification system which is more precise, more affordable, and easier to recognize. Additionally, using such automated diagnostic technologies can significantly reduce skin cancer mortality, which is advantageous for both patients and healthcare systems. An online automated systems decrease the wait times for medical dermoscopic testing and help to increase the survival rate of patients from larger geography.
- So, this research aims to find answer for the following research questions.
    - What machine learning (ML) model can be suggested that can run on low-end hardware or an edge device?
    - Will the suggested model's performance be comparable to that of the current state of the are image classification model?
    - Will the model's performance be enhanced by the feature fusion of image features and the patient's demographic data?

AIM & OBJECTIVES

SUGGESTED MODEL

To propose a predictive model for classification of skin disease in images.

PERFORMANCE COMPARISONS

To evaluate the efficiency of the suggested method against other available methods.

EFFECTIVENESS OF DEMOGRAPHIC DATA

To demonstrate the effect of demographic data usage on model performance.

METHODOLOGY

DATA

ISIC 2019 dataset with 8 skin disease classes. Do EDA to understand the data and fill missing values in patient's demographic data.

split the data in 70%, 20% and 10% train, test and validation set.

Data preprocesses to normalize light, intensity and other noises or reduce class imbalance.

CNN MODEL ARCHITECTURE

Download pretrained EffecientNetV2-B0, B1, B2, B3 and small models.

Add additional layers to produce 8 class classification result.

Determine hyperparameters such as Image shape, batch size, learning rate and epoch.

Experiment for best model, using only image data.

Create a parallel network to process demographic data for feature fusion.

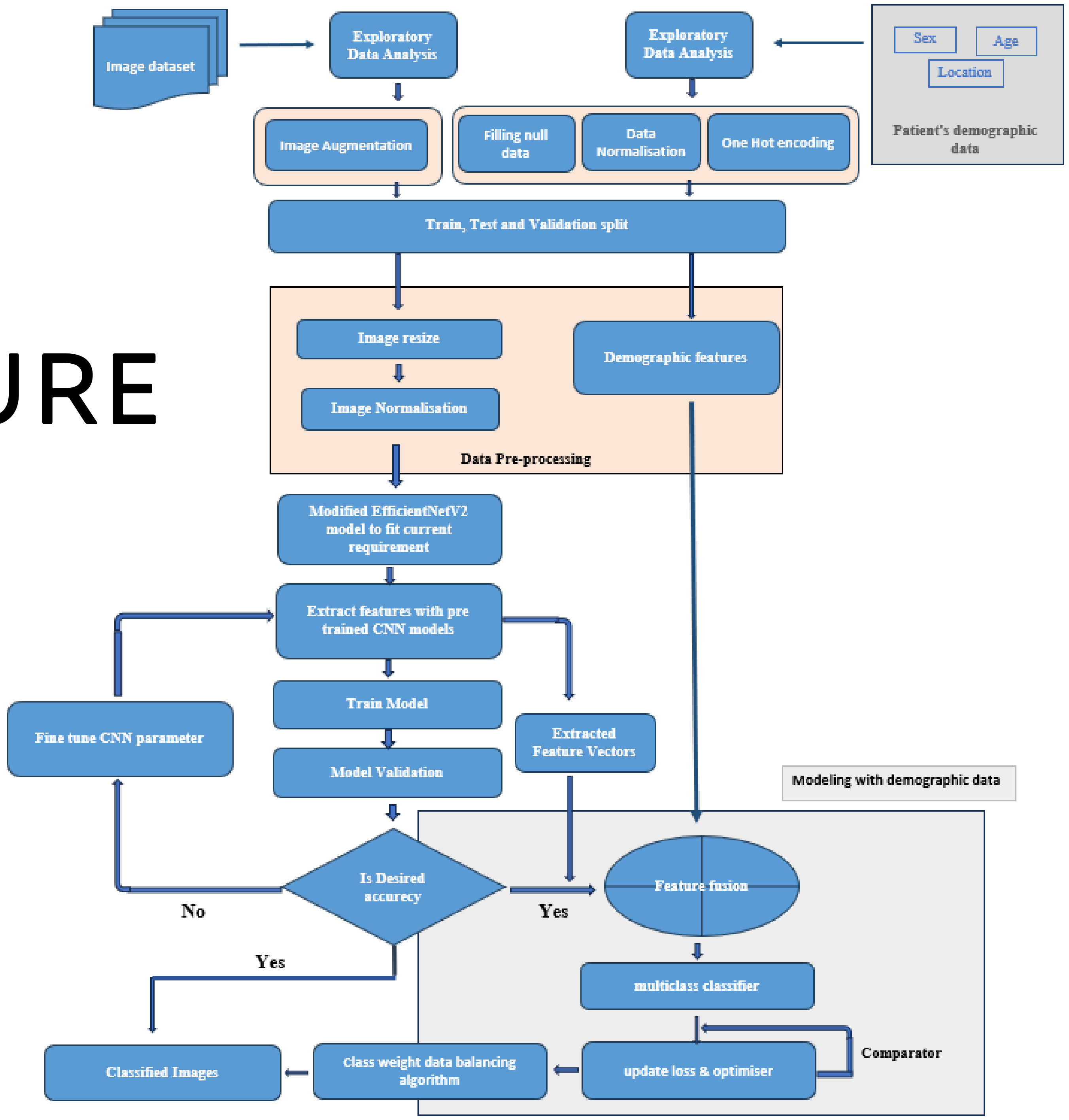
Experiment for best result using image and patient's demographic data.

EVALUATION MATRIX

Overall accuracy, f1 score and confusion metrices for different model comparison.

Precision, recall, f1 score, roc-auc curve and auc value for each class to understand model's efficiency to predict each class separately.

MODEL ARCHITECTURE



RESULTS & DISCUSSION

HYPERPARAMETER CHANGES AND DATA BALANCING TECHNIQUES

EVALUATION OF DIFFERENT MODELS

COMPARISON RESULT WITH MULTIMODAL APPROACH

RESULTS & DISCUSSION

HYPERPARAMETER CHANGES AND DATA BALANCING TECHNIQUES

IMAGE SIZE AND BATCH SIZE EFFECTS

With EffecientNetV2-small model with 32 batch size and 20 epochs, 3 experiments conducted using 3 different image shape. Image shape 224x224 outperforms 150x150 with 3-6% of improvement in f1 score for each class. Image shape 314x314 gave a memory issue due to limitation of system configuration.

In 3 experiments with batch size of 32, 64 and 128, batch size of 64 outperforms 32 batch size where batch size of 128 gave an memory limitation error.

CHANGE IN NUMBER OF EPOCHS

All considered EffecientNetV2 models with 64 batch size and 224x224 image shape, 2 experiments with 20 and 40 epochs conducted. After 20 epoch for couple of epochs accuracy increases with very low fraction and become static before 30 epochs. Only small version of model gave 3% of improvement in accuracy in 40 epoch training. Rest of models maintained same accuracy through out the run with slight fluctuation. So, 20 epochs finalized from this experiments.

DATA BALANCING TECHNIQUES

Image imputation making 2000 max count for a class if less image available versus class weight algorithm, 2 experiment considered. Image imputation making the training overfitting by achieving high train score, however other one maintained lesser difference. In test data image imputation achieved < 1% higher score. Considering this class weight techniques has been considered further. It also support script implementation for 3rd objective to check efficiency with demographic data.

RESULTS & DISCUSSION

EVALUATION OF DIFFERENT MODELS

EXPERIMENTAL MODELS

Results from different lightweight version of EffecientNetV2 models has been considered.

EffecientNetV2 models has much less parameters, small in size. train faster and able to achieve similar or better accuracy than its predecessor EffecientNet.

On overall, effecientNetV2-B2 have highest accuracy score of 83.4%.

Interclass comparison shows B2 and small version of model able predict different class with higher score.

EFFECIENTNETV2-B0, B1, B2, B3 , SMALL

Evaluation of interclass and intermodal efficiency

		Weighted Average Score				Model Size	
	Model	Accurecy	precision	recall	f1-score	Total params	Size (MB)
1	EffecientNetV2-B0	0.822	0.825	0.822	0.821	63,20,216	24.11
2	EffecientNetV2-B1	0.813	0.816	0.813	0.814	73,32,028	27.97
3	EffecientNetV2-B2	0.834	0.840	0.834	0.835	92,03,558	35.11
4	EffecientNetV2-b3	0.811	0.821	0.811	0.814	1,33,98,086	51.11
5	EffecientNetV2-S	0.832	0.839	0.832	0.834	2,07,32,264	79.09

		f1 score				
	class	B0	B1	B2	B3	S
0	Actinic_keratosis	0.663	0.696	0.696	0.689	0.658
1	Basal_cell_carcinoma	0.814	0.844	0.859	0.827	0.838
2	Benign_keratosis	0.716	0.718	0.745	0.684	0.707
3	Dermatofibroma	0.653	0.667	0.720	0.679	0.735
4	Melanocytic_nevus	0.899	0.899	0.898	0.883	0.901
5	Melanoma	0.740	0.752	0.743	0.733	0.773
6	Squamous_cell_carcinoma	0.632	0.613	0.651	0.645	0.672
7	Vascular_lesion	0.926	0.962	0.926	0.893	0.889

RESULTS & DISCUSSION

COMPARISON RESULT WITH MULTIMODAL APPROACH

SUGGESTED MODEL – EFFECIENTNETV2 – B2 / SMALL

In image only model EffecientNetV2-B0 and small version of models work better for different specific cases. So architecture updated to include cleaned demographic data.

In this experiment setup, small vision of model works better than B2 version of model.

Addition of metadata feature, made the models to produce less accuracy than image only models.

When compared small version of model achieved 83.2% of accuracy score where it able to achieve 82.7% including metadata.

Possible reason has been discussed in conclusion slides.

		Only Image		With Metadata	
	Model	Accurecy	f1-score	Accurecy	f1-score
1	EffecientNetV2-B2	0.834	0.835	0.784	0.789
2	EffecientNetV2-S	0.832	0.834	0.827	0.828

CONCLUSIONS

SUGGESTED MODEL

PERFORMANCE COMPARISONS

EFFECTIVENESS OF DEMOGRAPHIC DATA

FUTURE RECOMMENDATION

CONCLUSIONS

SUGGESTED MODEL

With review of all considered models, EffecientNetV2-B2 produced best result on ISIC2019 dataset. It achieved 83.4% of overall accuracy and 83.5% of f1 score. Total algorithm consists of 9M parameters and 35MB of size, which makes it fit for the objective to find a lightweight model for skin diseases classification.

CONCLUSIONS

PERFORMANCE COMPARISONS

As per research scope, only considered lightweight CNN models used in multiclass classification for performance comparison. In earlier research, proposed very light weight models like MobileNet, SqueezeNet produced 79% accuracy. Proposed model in this research achieved 83.4% of accuracy. The score is similar to other similar models like EffecientNet, ResNet50 etc. which has produced 82-86% of accuracy.

CONCLUSIONS

EFFECTIVENESS OF DEMOGRAPHIC DATA

In ISIC2019 dataset, patient's demographic information has large amount of empty data randomly or for a source/hospital the data received. After filling the data in statistical way and creating a feature fusion with image features, current research setup not able to prove that demographic data can help in improving classification score. Missing data, overlapping information may have confused the model and a bright scope for future experiments.

CONCLUSIONS

FUTURE RECOMMENDATION

- Data from more region and more balanced class will help the models to get better accuracy.
- The interpretability study of deep learning model could help to know the model behavior and its limitation.
- Combine faster hair removal or masking algorithm.
- An availability of non-missing patient's demographic data with good variability opens up a bright scope to experiment with heterogeneous data.

MAY

Research interest form submission

JUL

Topic submission

AUG

Research proposal submission

OCT

Mid thesis (interim report) submission

DEC

Final thesis and video presentation submission

TIMELINE IN YEAR 2023

AKASH DAS

HEARTFUL GRATITUDE

WORDS CANNOT EXPRESS MY GRATITUDE TO MY PROFESSORS AND SUPERVISORS FOR THEIR PATIENCE, INVALUABLE GUIDANCE AND FEEDBACK. THIS ENDEAVOUR WOULD NOT HAVE BEEN POSSIBLE WITHOUT THE GENEROUS SUPPORT FROM THE PLATFORM TEAM TO MAKE THE LEARNING EXPERIENCE MOST ENJOYABLE. AT LAST SPECIAL THANKS TO PROF. MANOJ FOR HIS WEEKLY INTERACTION AND FOR PROVIDING ANSWERS TO EACH QUERY PATIENTLY.

THANK YOU

AKASH DAS
ID: 1088937 | LIVERPOOL JOHN MOORES UNIVERSITY