Analysis of Skin Lesion Classification using Transfer Learning Models

Final Project by:
as4098 | Anirudh Srinivasa Raghavan
vm623 | Viswajith Menon
spa44 | Sreeniketh Aathreya

Introduction

- The project aims to use a benchmark dataset of skin lesion images to conduct a statistical analysis and build models that obtains accurate results of image classification.

- The project works on incorporating Deep Learning and tests how accurately transfer learning can help detect lesions of various different types.

- For this project, pre-trained models are used as classifiers and these pre trained networks perform the classification directly.

Problem Statement

The project aims to :

- Use deep learning techniques to classify lesion samples from a benchmark datasets in an accurate manner.
- Provide strong insight on the statistical comparison of various pre trained neural networks which are used for transfer learning and image processing.
- Perform an in-depth analysis of how each pre trained model performs. The models' performance is evaluated based on evaluation metrics like precision, recall, F1 score etc.

Literature Survey

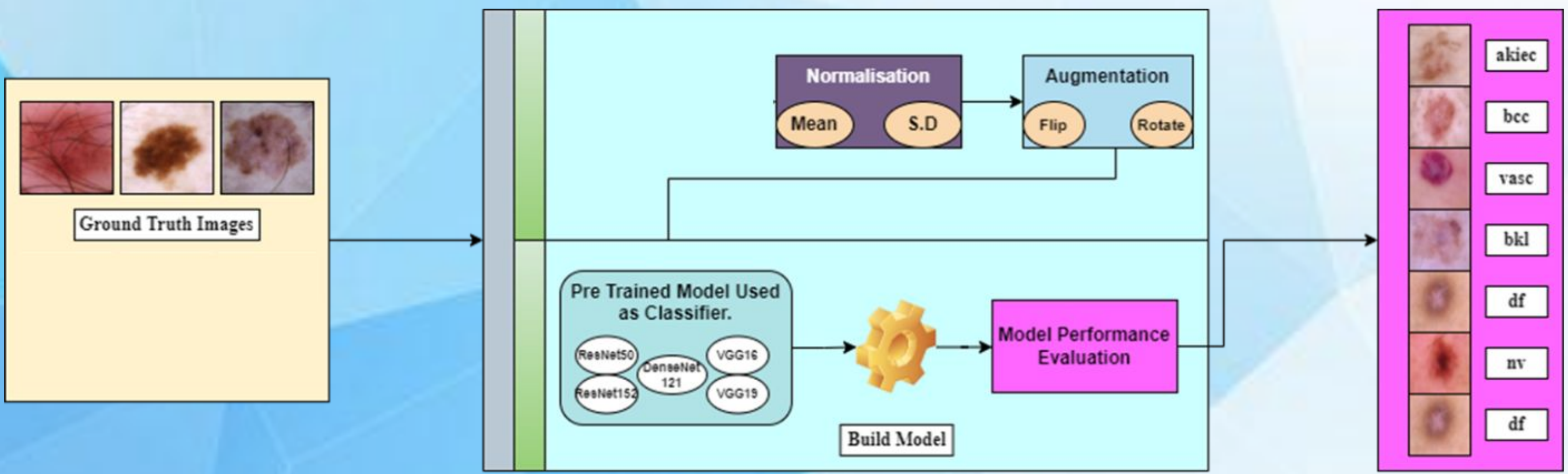
PAPER	YEAR	MECHANISM	PROS	CONS
Accessible Melanoma Detection Using Smartphones and Mobile Image Analysis5	2018	Accessible detection of malignant melanoma using mobile image analysis.	Mobile and portable ML model	MobileNet accuracy is not very precise and is arbitrary.
Melanoma Lesion Detection and Segmentation Using YOLOv4-DarkNet and Active Contour	2020	Enables YOLO v4 algorithm with Active Contour to detect lesion shape	Accurate detection of lesion in shape	No classification of malignant or benign lesions
Convolutional descriptors aggregation via cross-net for skin lesion recognition	2020	Framework for automatic skin lesion recognition using cross-net based aggregation of multiple convolutional networks	More powerful and more distinguishing feature representation capabilities	Inefficient Segmentation
Medical image classification using synergic deep learning	2019	Multi-CNN collaborative training dermoscopy image lesion recognition Model.	Improves the robustness of lesion identification and verified the effectiveness of the proposed method on related data sets	Does not give a accurate reading of if lesion is of any danger

Dataset Description

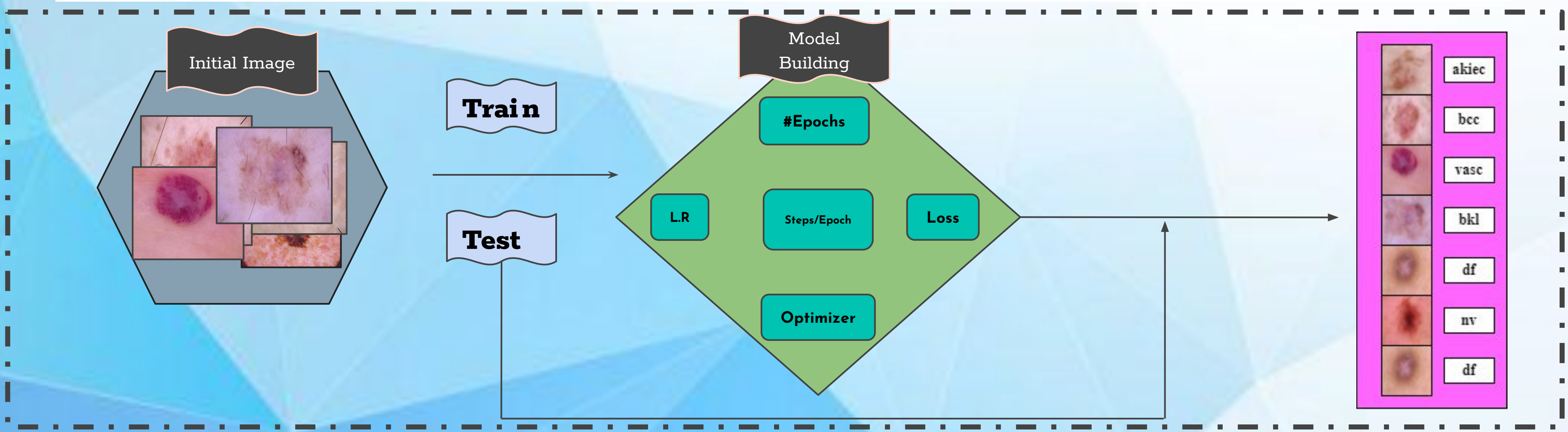
- The dataset consists of 10015 dermatoscopic images from different populations, acquired and stored by different modalities.
- The images are a representative collection of all important diagnostic categories in the realm of pigmented lesions. These are classified into 7 types of skin lesion namely
  - *Melanocytic nevus (nv)*
  - *Actinic keratosis (akiec)*
  - *Basal cell carcinoma (bcc)*
  - *Dermatofibroma (df)*
  - *Vascular lesion (vasc)*
  - *Malignant melanoma (mel)*
  - *Benign keratosis (bkl).*

- The dataset is a highly imbalanced with Melanocytic nevus having the most images.
- In order to balance the dataset, several data augmentation techniques like flipping, rotating and tilting have been implemented.

System Architecture



Working: Model Fitting



Train Test Split

The Dataset is split. Images are attributed to train the model and the rest are used for testing.

Pre-Trained Model as Classifier

Five Pre-Trained models are used for this project. They are ResNet50, VGG16, ResNet152, DenseNet121, and VGG19

Hyperparameter Tuning

The following hyperparameters are used:

Learning Rate(L.R) : 1e-3
Loss Function: Cross Entropy Loss
Optimizer: Adam
No of Epochs: 10
Steps per epoch: 1024

Algorithms Used

Normalisation

- The algorithms used for Normalization are:

$$\bar{E} = \frac{1}{(n)} \sum_{i=1}^{n} V_i$$

$$std(E) = \sqrt{\frac{1}{(n-1)} \sum_{i=1}^{n} \left(V_i - \bar{E}\right)^2}$$

$$\bar{E} = \frac{1}{(n)} \sum_{i=1}^{n} V_i$$

Model Evaluation

- The algorithms used in model evaluation metrics are:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

$$F1Score = \frac{2 * Precision * Recall}{Precision + Recall}$$

Pre-trained Models Used -

1. **ResNet50** - ResNet50 is a deep residual neural network architecture that has 50 layers. The ResNet50 architecture utilizes skip connections to address the vanishing gradient problem and allows for the training of deeper neural networks.
2. **VGG11**: VGG11 is a deep convolutional neural network architecture with 11 layers. It was introduced in 2014. The VGG11 architecture uses small 3x3 filters with a stride of 1 and a padding of 1 to maintain the spatial resolution of the input.
3. **ResNet101:** ResNet101 is a variant of the ResNet architecture that has 101 layers. It was introduced in 2015 and achieved state-of-the-art results on the ImageNet dataset. ResNet101 uses skip connections to address the vanishing gradient problem and allows for the training of even deeper neural networks.
4. **DenseNet121**: DenseNet121 is a deep neural network architecture that was introduced in 2016. It uses densely connected blocks, where each layer receives input from all previous layers in the block, to encourage feature reuse and reduce the number of parameters in the model. The DenseNet121 architecture has 121 layers and achieved state-of-the-art results on the ImageNet dataset.
5. **VGG19**: VGG19 is a variant of the VGG architecture with 19 layers. Similar to VGG 11, VGG19 architecture uses small 3x3 filters with a stride of 1 and a padding of 1 to maintain the spatial resolution of the input.

Classification Results

1. **Confusion Matrix –**

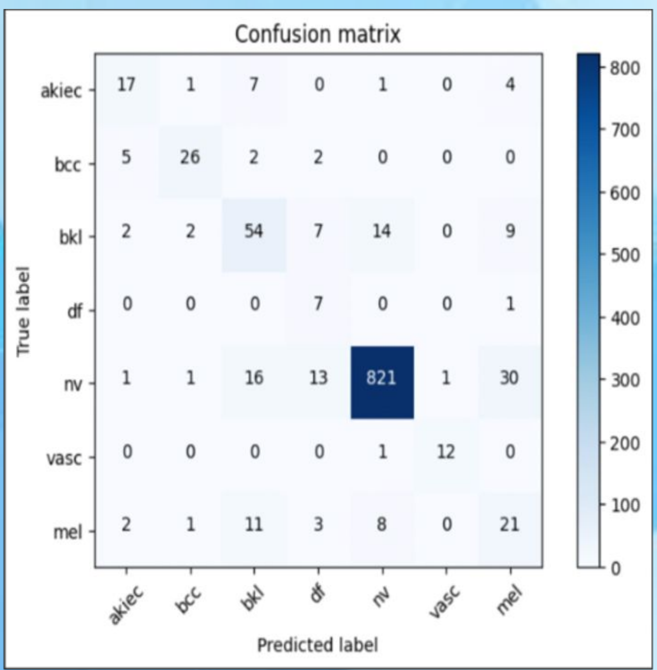
A confusion matrix is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known. It is a table with 4 different combinations of predicted and actual values.

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

2. **Classification Report –**

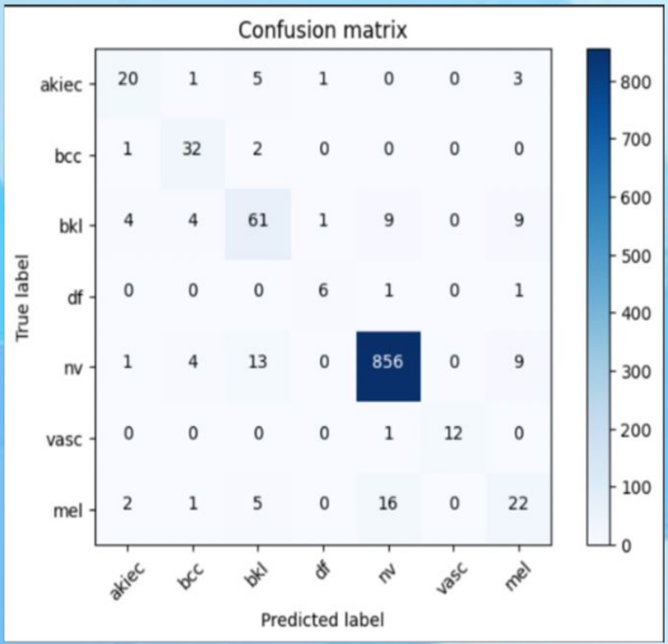
It is one of the performance evaluation metrics of a classification-based machine learning model. It displays your model's precision, recall, F1 score and support. It provides a better understanding of the overall performance of our trained model.

Resnet 50



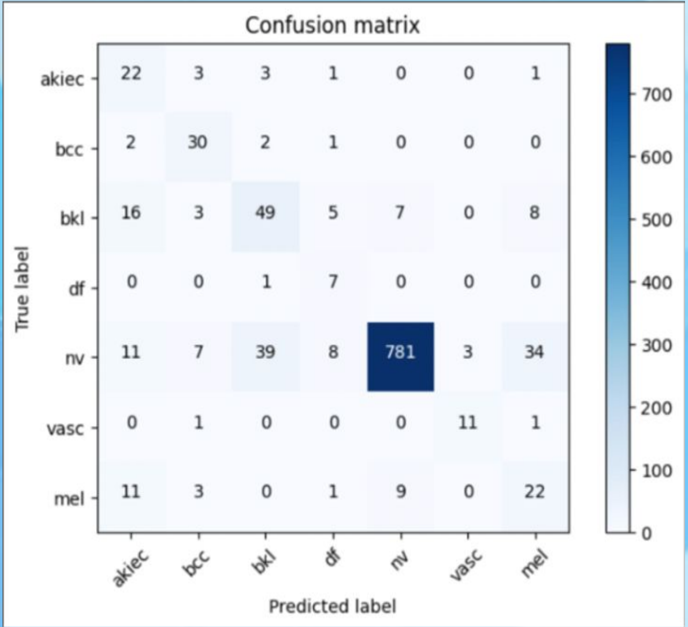
	precision	recall	f1-score
akiec	0.63	0.57	0.60
bcc	0.84	0.74	0.79
bkl	0.60	0.61	0.61
df	0.22	0.88	0.35
nv	0.97	0.93	0.95
vasc	0.92	0.92	0.92
mel	0.32	0.46	0.38
accuracy			0.87
macro avg	0.64	0.73	0.66
weighted avg	0.90	0.87	0.88

DenseNet121



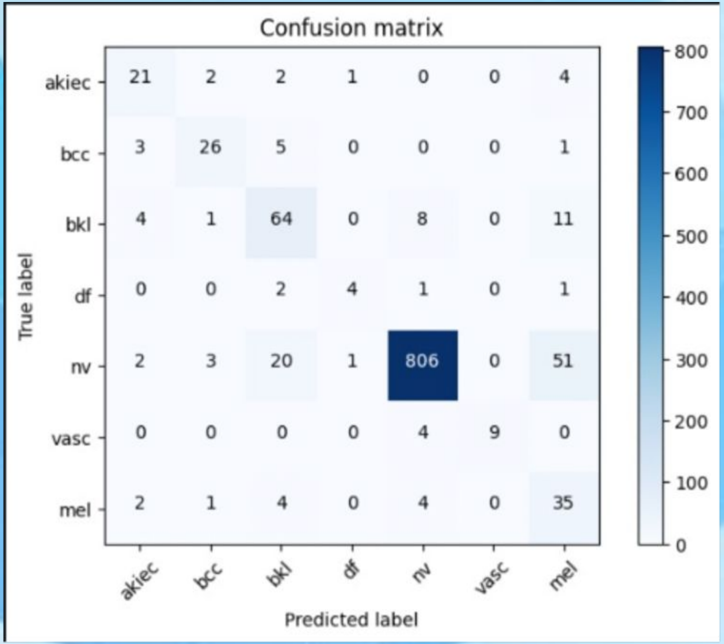
	precision	recall	f1-score
akiec	0.71	0.67	0.69
bcc	0.76	0.91	0.83
bkl	0.71	0.69	0.70
df	0.75	0.75	0.75
nv	0.97	0.97	0.97
vasc	1.00	0.92	0.96
mel	0.50	0.48	0.49
accuracy			0.91
macro avg	0.77	0.77	0.77
weighted avg	0.91	0.91	0.91

ResNet101



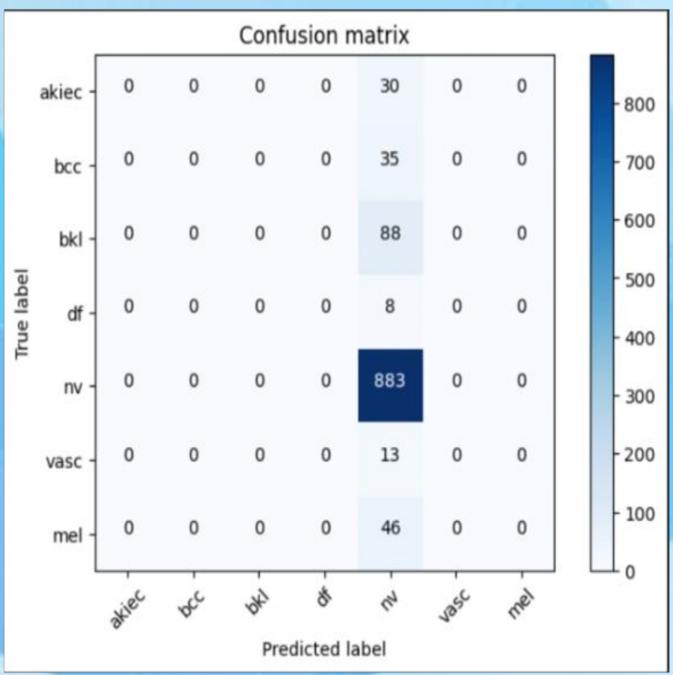
	precision	recall	f1-score
akiec	0.35	0.73	0.48
bcc	0.64	0.86	0.73
bkl	0.52	0.56	0.54
df	0.30	0.88	0.45
nv	0.98	0.88	0.93
vasc	0.79	0.85	0.81
mel	0.33	0.48	0.39
accuracy			0.84
macro avg	0.56	0.75	0.62
weighted avg	0.88	0.84	0.85

VGG11



	precision	recall	f1-score
akiec	0.66	0.70	0.68
bcc	0.79	0.74	0.76
bkl	0.66	0.73	0.69
df	0.67	0.50	0.57
nv	0.98	0.91	0.94
vasc	1.00	0.69	0.82
mel	0.34	0.76	0.47
accuracy			0.87
macro avg	0.73	0.72	0.71
weighted avg	0.91	0.87	0.89

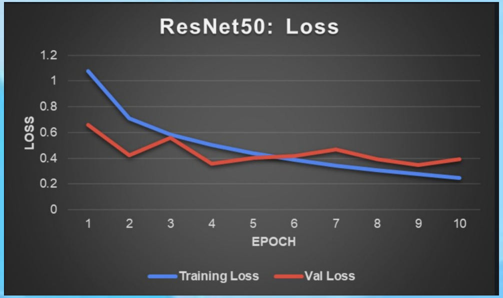
VGG19_BN



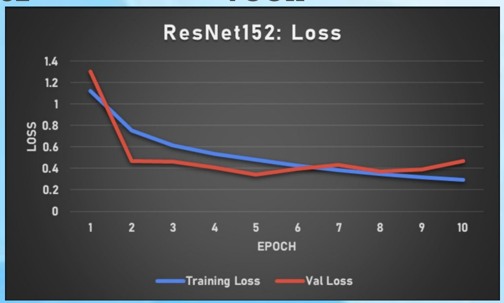
	precision	recall	f1-score
akiec	0.00	0.00	0.00
bcc	0.00	0.00	0.00
bkl	0.00	0.00	0.00
df	0.00	0.00	0.00
nv	0.80	1.00	0.89
vasc	0.00	0.00	0.00
mel	0.00	0.00	0.00
accuracy			0.80
macro avg	0.11	0.14	0.13
weighted avg	0.64	0.80	0.71

Analysis: Model Comparison (Loss/Epoch)

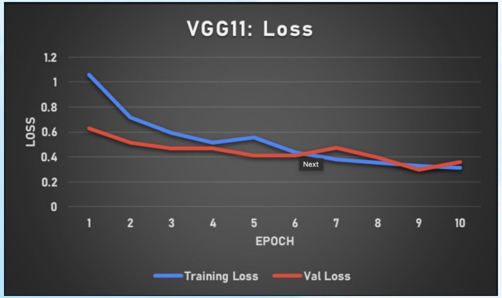
ResNet50 –



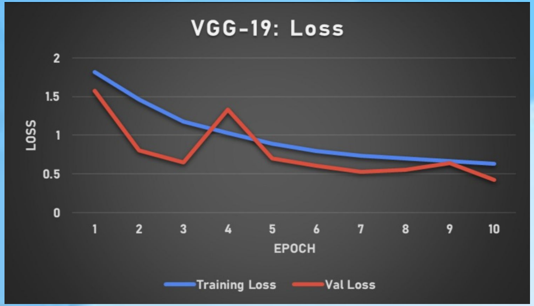
ResNet152 –



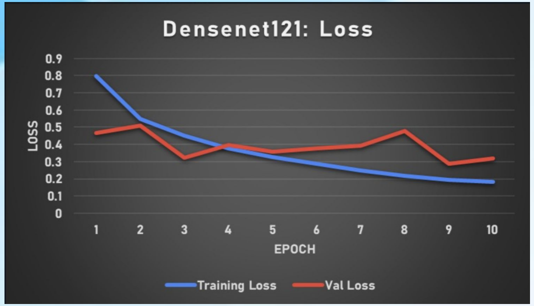
VGG11 –



VGG19_BN–

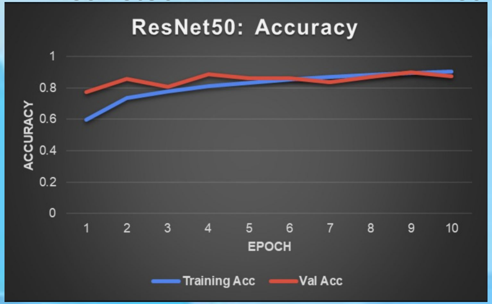


Densenet121 –

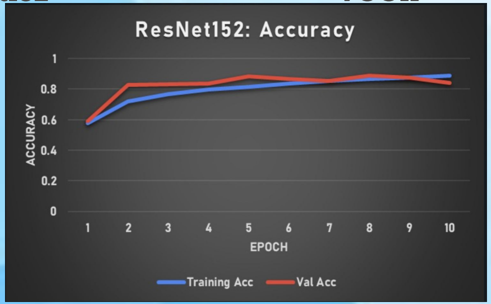


Analysis: Model Comparison (Accuracy/Epoch)

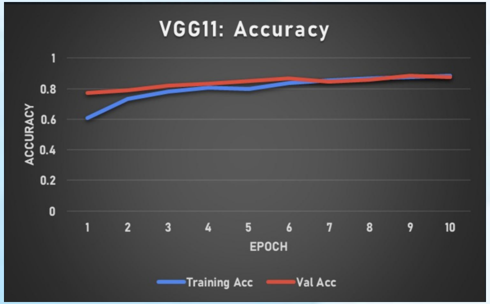
ResNet50 –



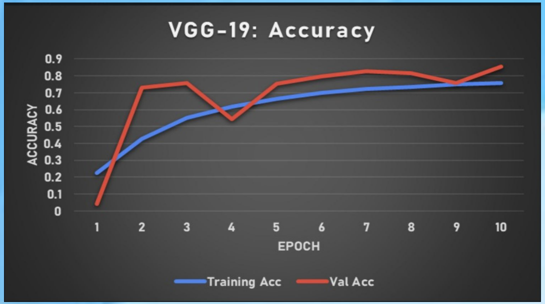
ResNet152 –



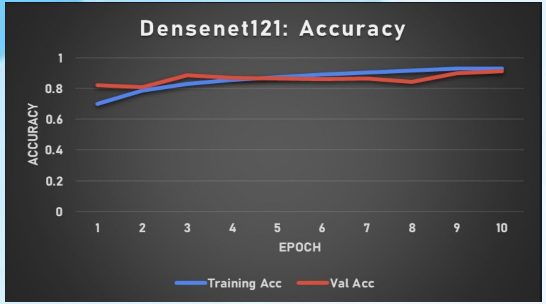
VGG11 –



VGG19_BN–



Densenet121 –



Analysis: Model Comparison Under Different Conditions

Validation Accuracy	Ground Truth
Resnet50	.87310
Resnet101	.84262
VGG11	0.8731
VGG19	.85423
Densenet121	.91327

Training Accuracy	Ground Truth
Resnet50	.90545
Resnet101	.88668
VGG11	0.90545
VGG19	.75929
Densenet121	.93045

Conclusion

- The conclusion of the project is that transfer learning is a useful technique to adopt when performing statistical analysis or comparison of pre-trained model performance. Once images are preprocessed and dataset is balanced, Transfer learning proves to be extremely useful.

- Best Performing Model – DenseNet121

- DenseNet121 gives us the best results for skin lesion classification as DenseNet121 generates dense feature maps, which may particularly be useful for detecting and classifying skin lesions. Dense feature maps contain more information than sparse feature maps, which may lead to better feature extraction and classification performance.

- DenseNet121 has a relatively small number of parameters compared to other deep learning architectures, which makes it more efficient and easier to train

References

- *O.R.P. P´erez V.d.J. A.Yabor A.M.Y. Palomo, M.d.J.D. P´erez and A. M. Fontaine. Melanoma maligno cut´aneo en pacientes de la provincia de lastunas. Revista Electr´onica Dr. Zoilo E. Marinello Vidaurret, 40, 2015.*
- *A. C. Society. Cancer facts and figures. 2016.*
- *H. Su J. Krause S. Satheesh S. Ma Z. Huang A. Karpathy A. Khosla M.Bernstein et al. O. Russakovsky, J. Deng. Imagenet large scale visual recognition challenge. International Journal of Computer Vision, 2016.*
- *Ammara Masood and Adel Al-Jumaily. Semi advised learning and classification algorithm for partially labeled skin cancer data analysis. In 2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), pages 1–4. IEEE, 2017.*
- *U˘gur Fidan, ˙Ismail Sarı, and Raziye K¨ubra Kumrular. Classification of skin lesions using ann. In 2016 Medical Technologies National Congress (TIPTEKNO), pages 1–4. IEEE, 2016.*
- *Mobeen ur Rehman, Sharzil Haris Khan, SM Danish Rizvi, Zeeshan Abbas, and Adil Zafar. Classification of skin lesion by interference of segmentation and convolotion neural network. In 2018 2nd International Conference on Engineering Innovation (ICEI), pages 81–85. IEEE, 2018.*

References

- *M Monisha, Alex Suresh, BR Tapas Bapu, and MR Rashmi. Classification of malignant melanoma and benign skin lesion by using back propagation neural network and abcd rule. Cluster Computing, 22(5):12897–12907, 2019.*
- *Marwan Ali Albahar. Skin lesion classification using convolutional neural network with novel regularizer. IEEE Access, 7:38306–38313, 2019.*
- *Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. nature, 542(7639):115–118, 2017.*
- *Fangfang Han, Huafeng Wang, Guopeng Zhang, Hao Han, Bowen Song, Lihong Li, William Moore, Hongbing Lu, Hong Zhao, and Zhengrong Liang. Texture feature analysis for computer-aided diagnosis on pulmonary nodules. Journal of digital imaging, 28(1):99–115, 2015.*

Thank you.