

# CSC580 Principals of Machine Learning Project

Caleb Jones Shibu, Kwang-Sung Jun

Department of Computer Science, University of Arizona, Tucson, Arizona, USA

## Explainable medical image classification

### 1. Introduction:

Classifying medical images effectively has been the essential goal in aiding clinical care and treatment. For example, some regions where radiologists are expensive and rare cause death to large number of people who suffer from pneumonia, in such circumstances artificial intelligence can be used to classify medical images. Such an approach could potentially save many lives.

Long ago the use of traditional machine learning methods such as Support Vector Machine (SVM) was used for medical image classification. However, these methods are time consuming, often provide low performance, and they require a lot of feature engineering. With the introduction of Convolutional Neural Networks we have seen its applications in variety of image classification tasks and it does so with high classification performance [1].

Studies [2] which used KNN and CNN for edge detection and they achieved a classification accuracy of 75% and 75.6% respectively to classify images into malignant or benign from International Skin Imaging dataset Collaboration (ISIC) dataset. Many studies used Deep convolutional neural network [3]–[5] where they achieved an average classification accuracy over 80 % in classifying multiclass skin cancer images. This goes to show that deep neural

networks can classify multiclass skin cancer images with high classification accuracy.

Likewise the same for Chest X-Ray classification where researchers have shown it is possible to classify chronic chest disease with a high classification accuracy of over 94% [6].

However Deep learning models are black boxes, they classify data with high accuracy without explaining why it did what it did. Additional research has been done to introduce explainability into Deep learning models, where extracting the spatial attention maps obtained with VGG-16 and DenseNet-161 [7] that it is possible to identify interpretable feature and regions from cancer images. Likewise, the same was done for introduce explainability in classification of Chest X-ray images [8].

In this study we aim to classify two types of multiclass data (1) Cancer and (2) Chest X-Ray while also introducing some level of explainability.

### 2. Dataset:

#### 2.1. Chest X-Ray dataset:

An open access COVID-19 chest X-Ray dataset [9] with three classes Normal, Viral Pneumonia and COVID was used to train the deep learning model.

#### 2.2. Skin Cancer dataset:

International Skin Imaging Dataset Collaboration (ISIC) dataset with 7 classes Melanocytic nevi,

Melanoma, Benign keratosis-like lesions, Basal cell carcinoma, Actinic keratoses,

Vascular lesions, Dermatofibroma

and

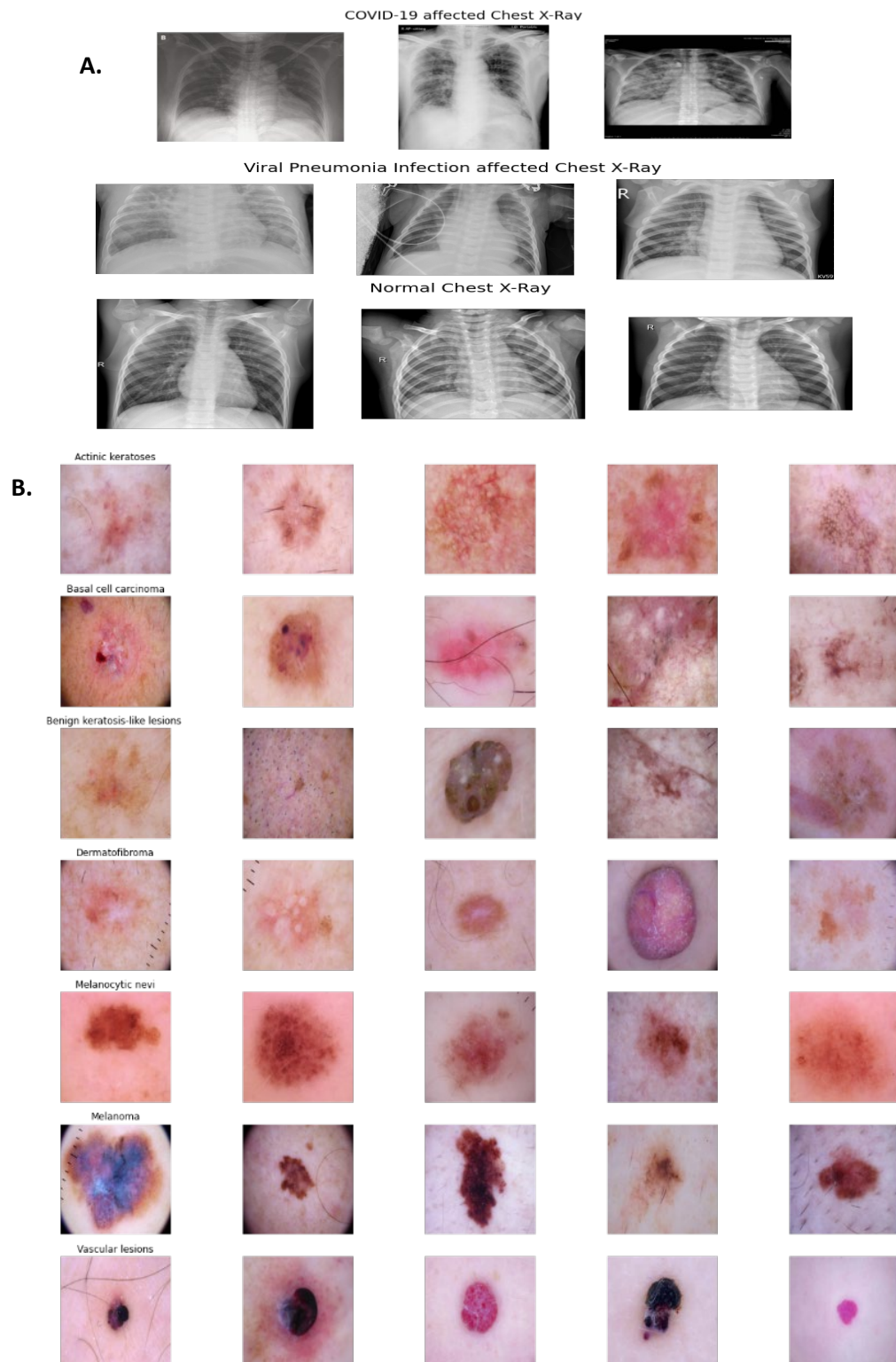


Figure 1 A. Group of 3 images for COVID-19 Chest X-Ray, Viral pneumonia and Normal chest X-Ray.  
 B. Group of 4 random images for Melanocytic nevi, Melanoma, Benign keratosis-like lesions, Basal cell carcinoma, Actinic keratoses, Vascular lesions, and Dermatofibroma.

### 3. Methods:

#### 3.1. Data pre-processing:

For Chest X-Ray dataset, the images were enhanced using the same approach following in COVIDLite [10] white balance and CLAHE.

White balance is done to ensure that the final image represents the true image while enhancing the visibility of specific features that the CNN would capture.

Contrast Limited Adaptive Histogram Equalization (CLAHE) is done to increase the contrast of the image. This is done by stretching the intensity values of the input image that changes the natural brightness.

Both white balance and CLAHE has shown significant improvement in enhancing CT and X-ray images.

#### 3.2. Data augmentation:

Both chest X-Ray dataset and skin cancer dataset were passed data augmenter to increase the number of samples to avoid the issue of overfitting in the neural network.

Option	Value
Rotation Range	10
Zoom Range	0.1
Width shift range	0.12
Height shift range	0.12
Horizontal flip	True
Vertical flip	True

Table 1 Parameters set for data augmentation for both the datasets.

#### 3.3. Classification models:

##### 3.3.1. Depth-wise separable deep neural network:

The DSCNN (Depth-wise separable CNN) [11] divides regular convolution to 2 parts: (a) depth wise or spatial

convolution and (b) sequential point-wise convolution.

The depth wise convolution operation to each of the input channels 2D filters are applied. The number of computations is less as a single filter is applied to each input channel. Furthermore, a new feature map is generated by running a pointwise of 1x1 convolution.

Only the Chest X-ray dataset was trained on this model.

Layer	Output Shape	Parameter #
InputLayer	None, 224, 224, 3	0
Conv2D	None, 224, 224, 16	448
Conv2D	None, 224, 224, 16	2320
MaxPooling2D	None, 112, 112, 16	0
SeparableConv2d	None, 112, 112, 32	688
SeparableConv2d	None, 112, 112, 32	1344
BatchNormalization	None, 112, 112, 32	128
MaxPooling2D	None, 56, 56, 32	0
SeparableConv2d	None, 56, 56, 64	2400
SeparableConv2d	None, 56, 56, 64	4736
BatchNormalization	None, 56, 56, 64	256
MaxPooling2D	None, 28, 28, 64	0
SeparableConv2d	None, 28, 28, 128	8896
SeparableConv2d	None, 28, 28, 128	17664
BatchNormalization	None, 28, 28, 128	512
MaxPooling2D	None, 14, 14, 128	0
Dropout	None, 14, 14, 128	0
SeparableConv2d	None, 14, 14, 256	34176
SeparableConv2d	None, 14, 14, 256	68096

BatchNormalization	None, 14, 14, 256	1024
MaxPooling2D	None, 7, 7, 256	0
Dropout	None, 7, 7, 256	0
SeparableConv2d	None, 7, 7, 512	133888
SeparableConv2d	None, 7, 7, 512	267264
BatchNormalization	None, 7, 7, 512	2048
MaxPooling2D	None, 3, 3, 512	0
Dropout	None, 3, 3, 512	0
Flatten	None, 4608	0
Dense	None, 512	2359808
Dropout	None, 512	0
Dense	None, 128	65664
Dropout	None, 128	0
Dense	None, 64	8256
Dropout	None, 64	0
Dense	None, 32	0
Dropout	None, 32	2080
Dense	None, 3	99
Total params: 2,981,795 Trainable params: 2,979,811 Non-trainable params: 1,984		

Table 2 Model architecture for DSCNN

### 3.3.2. ResNetV2:

Residual network (ResNet) [12] is built on top of the principal of skip connections. For each branch, the model learns the difference between input and output residual activation maps. This difference is then added together with previous activation maps to form a "Collective knowledge" of the residual network. As deep neural networks are hard to train, ResNet come in handy as they skip connections which allows a deeper learning than ever before.

Pretrained ResNetv2 with ImageNet weights was used to build a model to train on Skin-cancer dataset.

Layer	Output Shape	Parameter #
InputLayer	None, 125, 100, 3	0
ResNet50v2	None, 4, 4, 2048	2356480

GlobalAveragePooling2D	None, 2048	0
Dropout	None, 2048	0
Dense	None, 512	1049088
Dropout	None, 512	0
Dense	None, 7	3591
Total params: 24,617,479 Trainable params: 1,052,679 Non-trainable params: 23,564,800		

Table 3 Model architecture for using ResNetv2

### 3.4. Explanation module:

#### 3.4.1. LIME:

Local Interpretable Model-agnostic Explanations (LIME) [13] tries to interpret outcome of a deep neural net by the following:

- For an input image, LIME will generate several samples which closely resemble the input by turning off or on some of the super-pixels in the image.
- Next, LIME will predict which class the generated data point belongs to in the model.
- Then the weight of the generated data point is used a metric to find the distance between generated and input image, bigger distance means more weight and importance.
- Lastly using linear regression fit the generated data. Sorted coefficients show us the features with utmost importance.

#### 3.4.2. Grad-CAM Heatmaps:

Gradient weighted Class Activation Maps (Grad-CAM) [14] is weakly-supervised localization algorithm which is trained on the whole image rather than annotated locations and also without pixel wise training.

Grad-CAM is built on the idea that exploits the spatial information preserved through convolutional layers, this allows it understand sections of an image that stand out the most for classification of an image.

It works following the 3 steps:

- Compute the gradient for image with respect to activation map of the convolutional layer.
- Perform a global average over the width and height to get the weights important.
- Lastly, take the sum of weights of the feature maps for a particular activation.

### 3.4.3. Saliency Maps:

Saliency maps [15] in an image is a topographic representation which pertains to unique features like pixels, resolution, etc in context of visual analysis. Colour temperature is one of most commonly used to show important features in an image.

It is calculated as follows:

- Three features: colours, intensity and orientation are extracted from an image.
- Colours in the image are converted from red-green-yellow-blue colour space.
- Gabor filters are used to obtain the orientation filters with respect to four angles.
- Feature map is created from the three features and saliency maps is mean of these features.

## 4. Results and discussion:

### 4.1. Classification:

#### 4.1.1. Chest X-ray Dataset:

DSCNN was trained with data split of 75% train- and 25% test- data. Accuracy plots are shown in Figure 2.

The test data set achieved a classification accuracy of 90.96%.

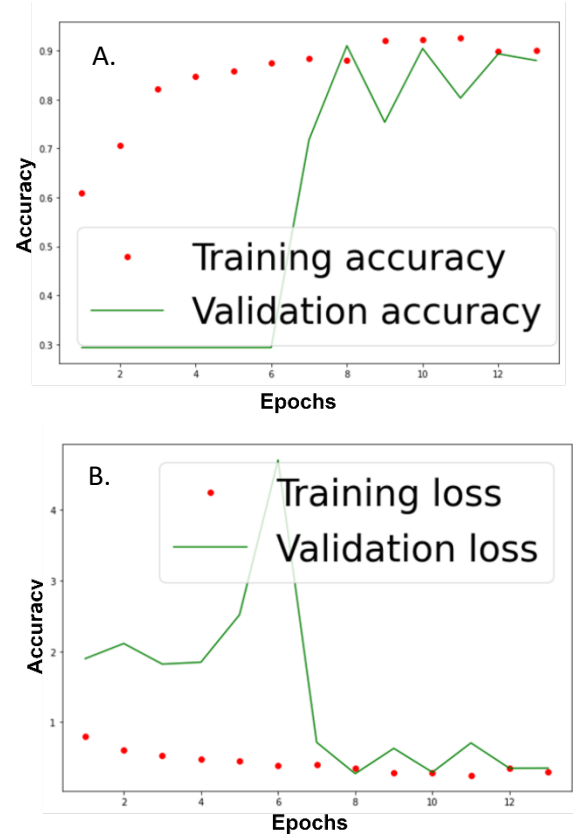


Figure 2 A. Training and validation accuracy and B. Training and validation loss for DSCNN for Chest X-Ray dataset.

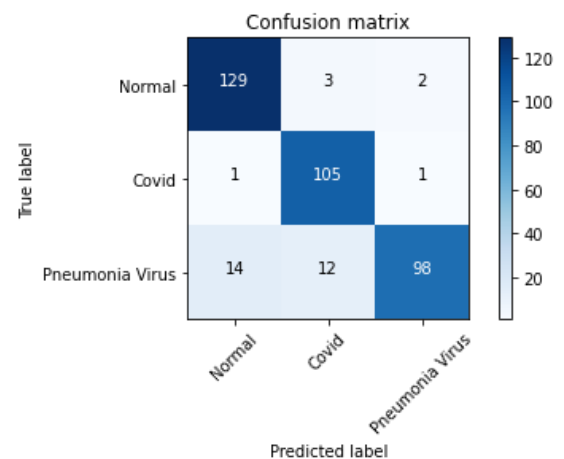


Figure 3 Confusion matrix for model predictions on test Dataset for DSCNN for Chest X-Ray dataset.

#### 4.1.2. Skin Cancer dataset:

Pretrained ResNetV2 with ImageNet weights was trained with data split of 75% train- and 25% test- data. Accuracy plots are shown in Figure 4.

The test data set achieved a classification accuracy of 70.04%. Figure 5 shows that the dataset wasn't perfectly balanced as Actinic keratoses had more number of images compared to others.

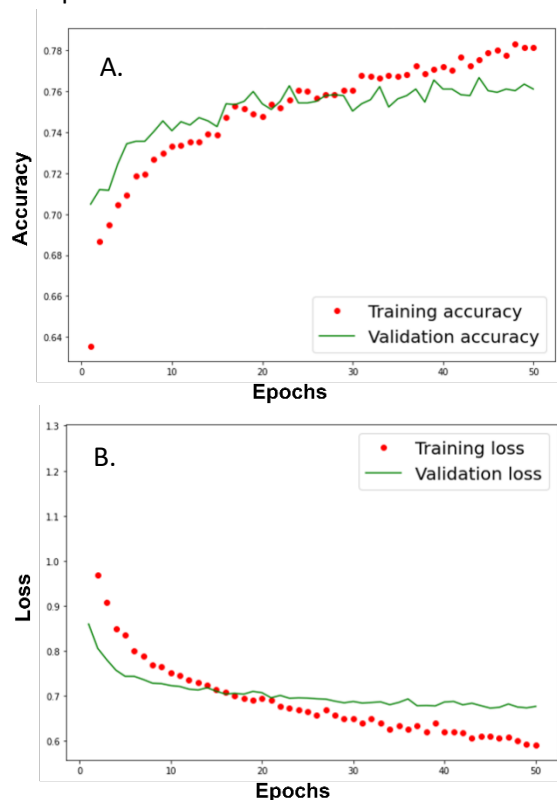


Figure 5 A. Training and validation accuracy and B. Training and validation loss for ResNetV2 for Skin-Cancer Dataset.

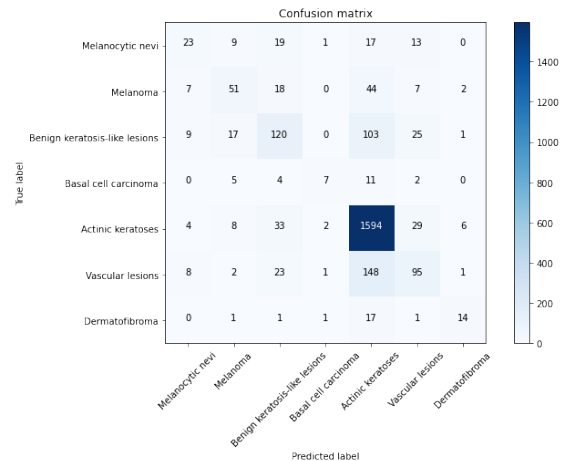


Figure 4 Confusion matrix for model predictions on test dataset for ResNetV2 for Skin-Cancer dataset.

#### 4.2. Explanations:

##### 4.2.1. Chest X-ray dataset:

Figure 6 under LIME explanations the shaded areas are called super pixels, the shaded green is region contributing the most towards correct classification and red region is contributing to incorrect classification. We see that LIME exPlainations for viral pneumonia shows a lesion in the left lung which is highlighted as green by LIME explainer.


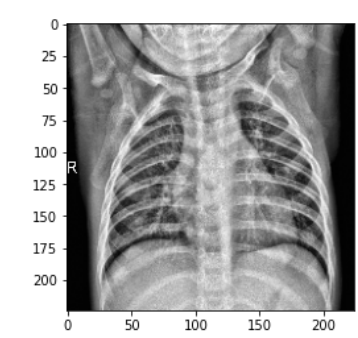
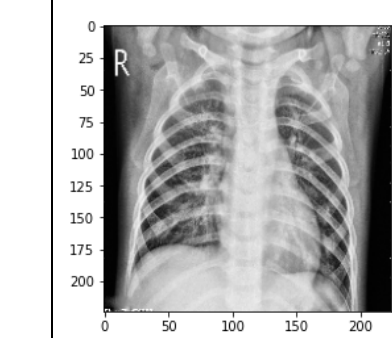
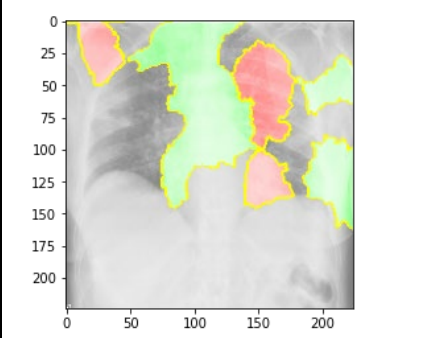
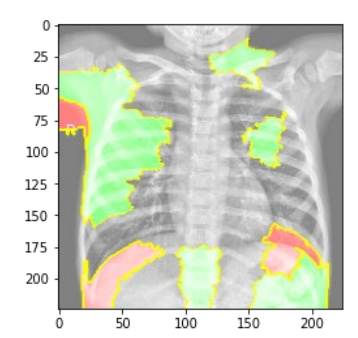
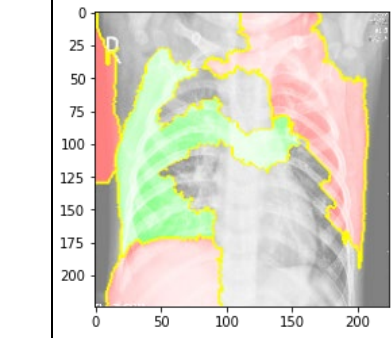
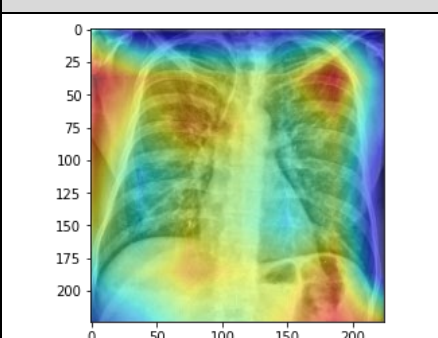
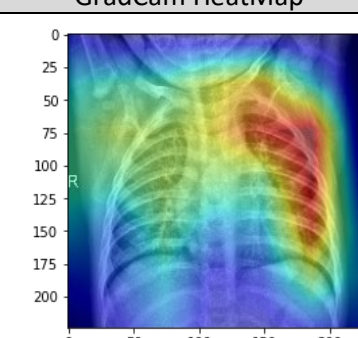
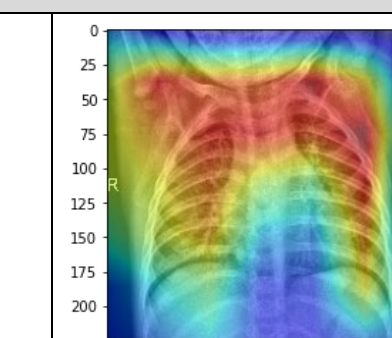
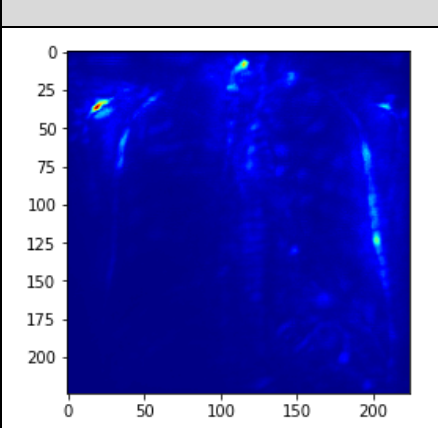
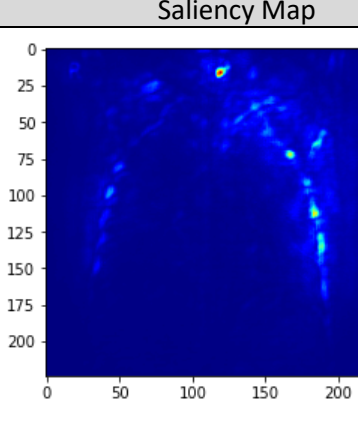
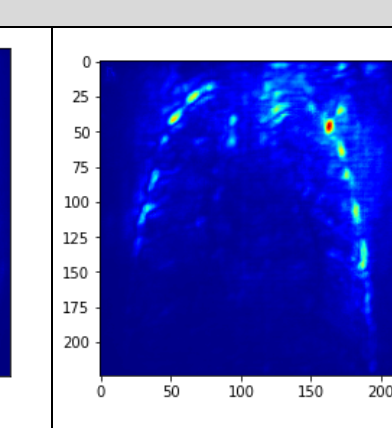
Figure 6 under GradCAM heatmap we can see that red regions show area with highest level of activation for that image. For Normal lung we can see that right lung shows high level activation towards the centre which enables the model to classify the image correctly.

Figure 6 Saliency Maps is hard to interpret as it appears that the ribs seem to be highlighted rather than the lesions in the lungs.

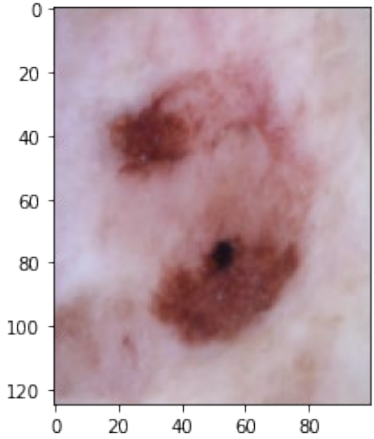
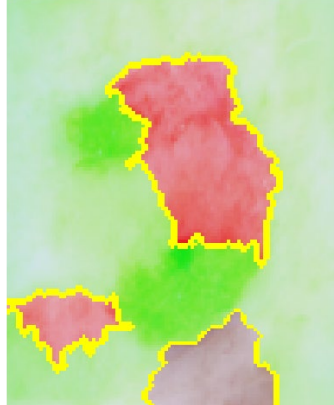
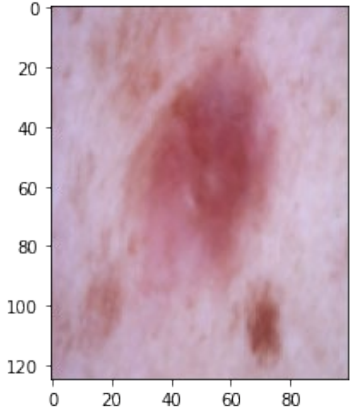
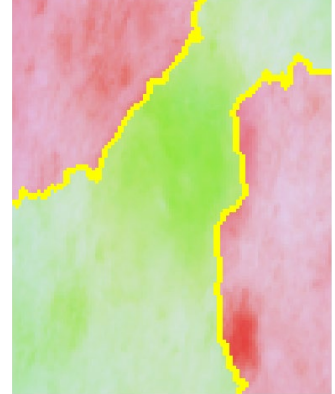
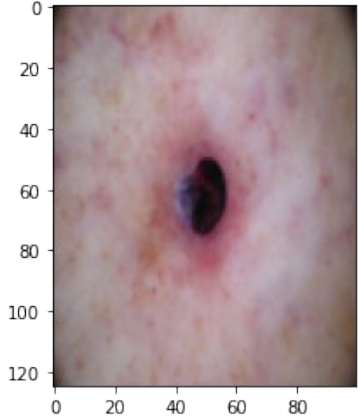

Figure 7 we couldn't explain the model with GradCAM or Saliency as we used a pretrained model and it had some compatibility issues which due to time constraints couldn't be resolved.

However, in Figure 7 LIME explanation did a excellent job of highlighting the regions contributing positively for classification. For example, Basel cell carcinoma was perfectly highlighted as green showing that the entire lesion contributed in positive classification.

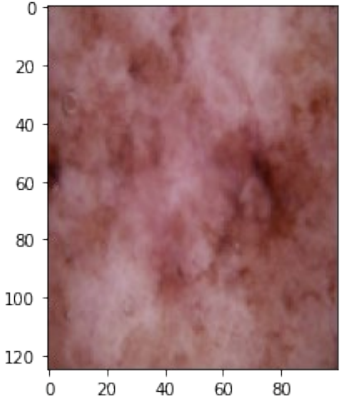
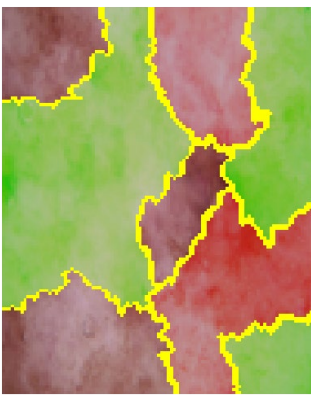
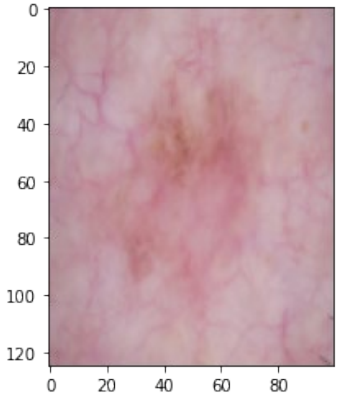
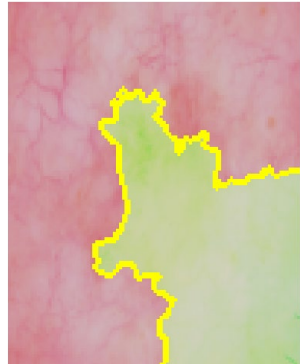
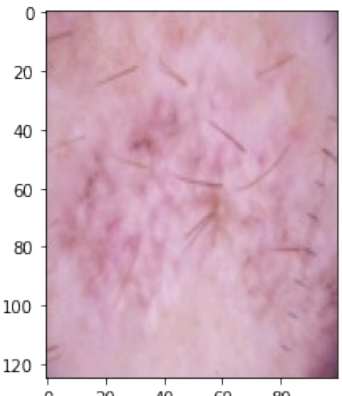

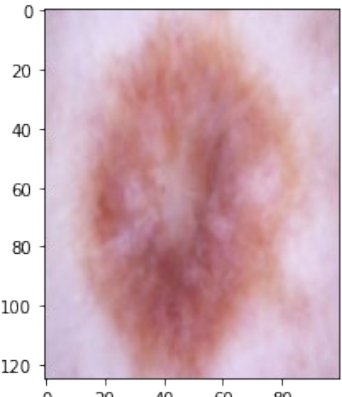



COVID19 affected lungs	Normal Lungs	Viral Pneumonia affected lungs
True image		
		
LIME explanation		
		
GradCam HeatMap		
		
Saliency Map		
		
Figure 6 DSCNN model explanation on Chest X-Ray dataset with LIME explanations, GradCam Heatmaps, and Saliency Maps.		

#### 4.2.2. Skin-Cancer Dataset:

Cancer type:	True image	LIME explanation
Melanoma		
Actinic keratoses		
Dermatofibroma		



Vascular lesions		
Benign keratosis-like lesions		
Melanocytic nevi		
Basal cell carcinoma		
<p>Figure 7 ResNetV2 model explanations on Skin-Cancer Dataset with LIME explanations.</p>		

## References:

- [1] W. Rawat and Z. Wang, 'Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review', *Neural Computation*, vol. 29, no. 9, pp. 2352–2449, Sep. 2017, doi: 10.1162/neco\_a\_00990.
- [2] T. R. Savera, W. H. Suryawan, and A. W. Setiawan, 'Deteksi Dini Kanker Kulit menggunakan K-NN dan Convolutional Neural Network', *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 7, no. 2, Art. no. 2, Feb. 2020, doi: 10.25126/jtiik.2020702602.
- [3] X. Zhang, S. Wang, J. Liu, and C. Tao, 'Towards improving diagnosis of skin diseases by combining deep neural network and human knowledge', *BMC Med Inform Decis Mak*, vol. 18, no. Suppl 2, p. 59, Jul. 2018, doi: 10.1186/s12911-018-0631-9.
- [4] N. Hameed, A. M. Shabut, and M. A. Hossain, 'Multi-Class Skin Diseases Classification Using Deep Convolutional Neural Network and Support Vector Machine', in *2018 12th International Conference on Software, Knowledge, Information Management Applications (SKIMA)*, Dec. 2018, pp. 1–7. doi: 10.1109/SKIMA.2018.8631525.
- [5] Y. N. Fu'adah, N. C. Pratiwi, M. A. Pramudito, and N. Ibrahim, 'Convolutional Neural Network (CNN) for Automatic Skin Cancer Classification System', *IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 982, p. 012005, Dec. 2020, doi: 10.1088/1757-899X/982/1/012005.
- [6] Y.-X. Tang *et al.*, 'Automated abnormality classification of chest radiographs using deep convolutional neural networks', *npj Digit. Med.*, vol. 3, no. 1, pp. 1–8, May 2020, doi: 10.1038/s41746-020-0273-z.
- [7] C. Barata, M. E. Celebi, and J. S. Marques, 'Explainable skin lesion diagnosis using taxonomies', *Pattern Recognition*, vol. 110, p. 107413, Feb. 2021, doi: 10.1016/j.patcog.2020.107413.
- [8] A. J. DeGrave, J. D. Janizek, and S.-I. Lee, 'AI for radiographic COVID-19 detection selects shortcuts over signal', *Nat Mach Intell*, vol. 3, no. 7, pp. 610–619, Jul. 2021, doi: 10.1038/s42256-021-00338-7.
- [9] J. P. Cohen, P. Morrison, L. Dao, K. Roth, T. Q. Duong, and M. Ghassemi, 'COVID-19 Image Data Collection: Prospective Predictions Are the Future', *arXiv:2006.11988 [cs, eess, q-bio]*, Dec. 2020, Accessed: Dec. 06, 2021. [Online]. Available: <http://arxiv.org/abs/2006.11988>
- [10] M. Siddhartha and A. Santra, 'COVIDLite: A depth-wise separable deep neural network with white balance and CLAHE for detection of COVID-19', *arXiv:2006.13873 [cs, eess]*, Jun. 2020, Accessed: Dec. 06, 2021. [Online]. Available: <http://arxiv.org/abs/2006.13873>
- [11] 'Xception: Deep Learning with Depthwise Separable Convolutions | IEEE Conference Publication | IEEE Xplore'. <https://ieeexplore.ieee.org/document/8099678> (accessed Dec. 07, 2021).
- [12] K. He, X. Zhang, S. Ren, and J. Sun, 'Deep Residual Learning for Image Recognition', *arXiv:1512.03385 [cs]*, Dec. 2015, Accessed: Sep. 28, 2021. [Online]. Available: <http://arxiv.org/abs/1512.03385>
- [13] M. T. Ribeiro, S. Singh, and C. Guestrin, '"Why Should I Trust You?": Explaining the Predictions of Any Classifier', *arXiv:1602.04938 [cs, stat]*, Aug. 2016, Accessed: Dec. 07, 2021. [Online]. Available: <http://arxiv.org/abs/1602.04938>
- [14] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, 'Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization', *Int J Comput Vis*, vol. 128, no. 2, pp. 336–359, Feb. 2020, doi: 10.1007/s11263-019-01228-7.
- [15] K. Simonyan, A. Vedaldi, and A. Zisserman, 'Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps', *arXiv:1312.6034 [cs]*, Apr. 2014, Accessed: Dec. 07, 2021. [Online]. Available: <http://arxiv.org/abs/1312.6034>