Pneumonia Classification

Pneumonia classification from chest X-rays using deep learning

Hakim Chekirou

Why does it matter ?

- Pneumonia accounts for over 15% of all deaths of children under 5 years old.

- It requires highly trained specialists and confirmation through clinical history.

- Other conditions such as fluid overload (pulmonary edema),
  bleeding, volume loss, lung cancer, or post-radiation
  or surgical changes make the prediction harder.

- A number of factors such as positioning of the patient and depth
  of inspiration can alter the appearance of the CXR.

- In addition, clinicians are faced with reading high volumes of images every shift.

Dataset

- Pneumonia detection challenge

- 26,684 Chest X rays in the form of Dycom files.

- Unbalanced dataset (20672 negative vs 6012 positive examples.

- High resolution 1024*1024 in gray scale
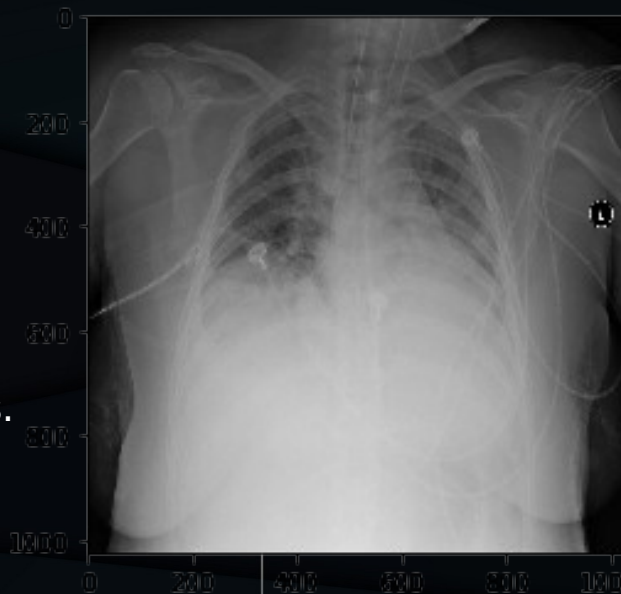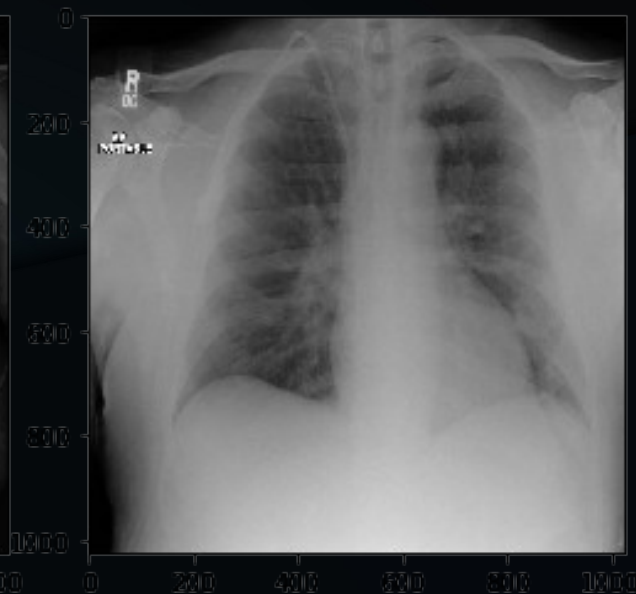
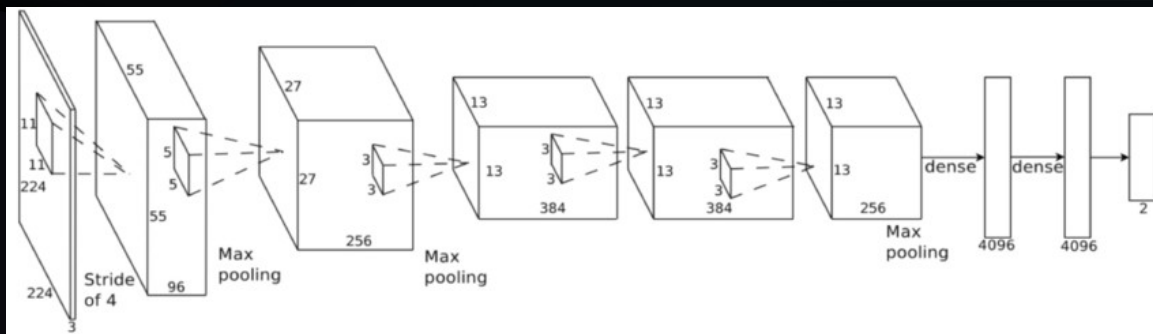Positive example   Negative example

Data preparation

Preprocessing

- Normalizing.

- Train and test split (90 %-10%).

Data Augmentation :

- Random rotation up to 6°

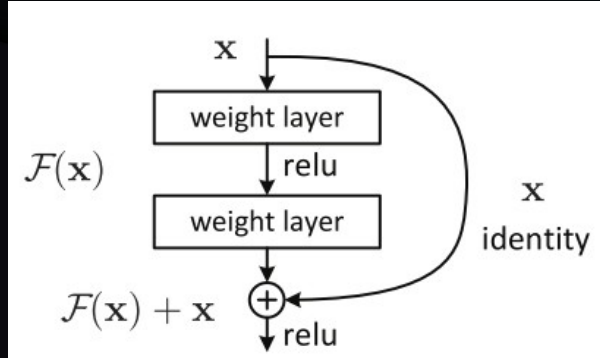- Random horizontal flipping

Convolution networks (alexnet)

Alexnet architecture

- Pretrained on imagenet
- Finetuned for 20 epochs
- SGD and weighted Cross-entropy loss

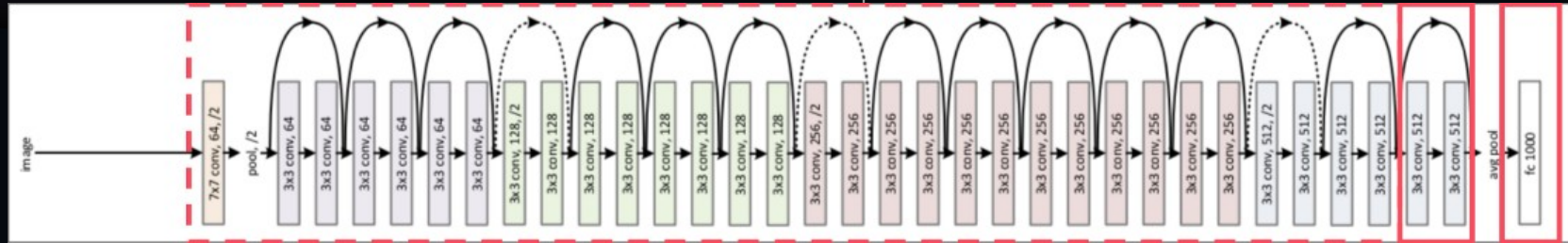| Class	Precision	Recall	F-1 score	ROC AUC	Accuracy
Negative	0.89	0.88	0.89	76.64%	83%
Positive	0.62	0.65	0.64		

Alexnet results

Residual networks

Residual block

- Pretrained on ImageNet and finetuned for 20 epochs.
- Using weighted cross entropy loss
- and SGD with momentum

Resnet50 architecture
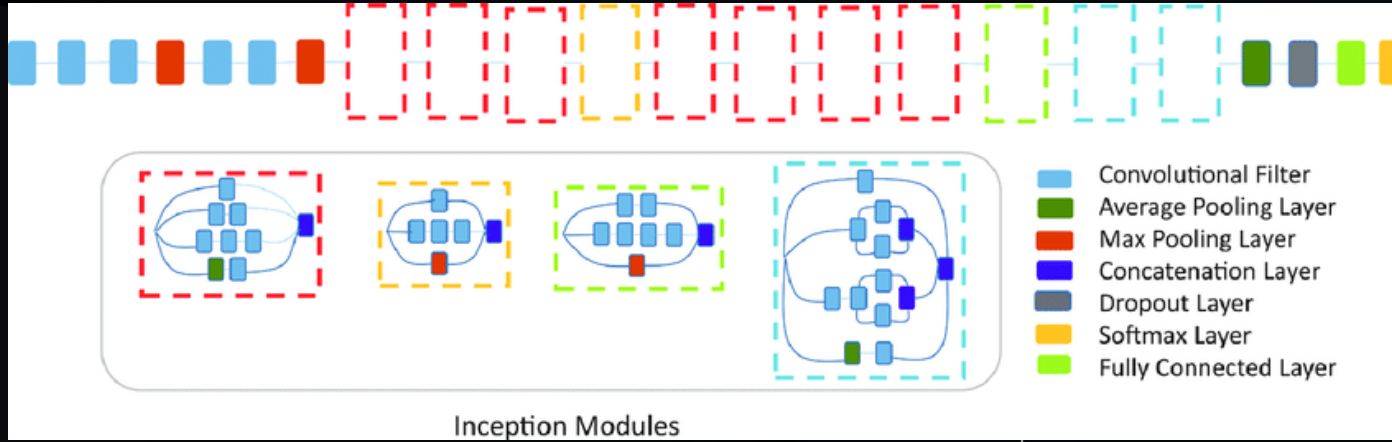
Residual networks results

Class	Precision	Recall	F-1 score	ROC AUC	Accuracy
Negative	0.90	0.87	0.88	76.84%	82 %
Positive	0.60	0.67	0.63		

Results for the Resnet50 architecture

Class	Precision	Recall	F-1 score	ROC AUC	Accuracy
Negative	0.94	0.75	0.84	79.71%	77 %
Positive	0.50	0.84	0.63		

Results for the Resnet18 architecture

Inception v3

Inception v3 architecture

- Pretrained on imagenet
- Finetuned for 20 epochs
- SGD and weighted CE loss

Class	Precision	Recall	F-1 score	ROC AUC	Accuracy
Negative	0.91	0.84	0.88	78.64%	82%
Positive	0.58	0.73	0.65		

Inception v3 results

Proposition : Vision Transformers
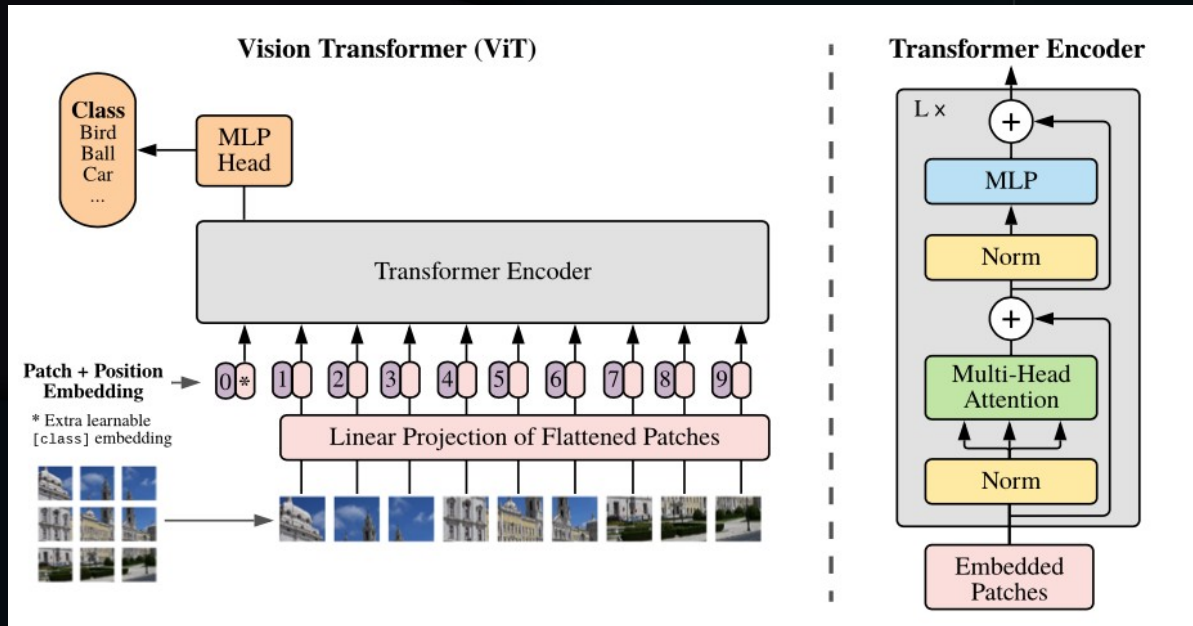
- Trained from scratch using two sets of parameters.
- Variation of number of patches : 64 vs 256
- Use of a CLS token or mean representation for classification.
- Depth fixed to 6 layers
- MLP dimension of 1024
- Use of dropout in the attention weights and positional embeddings.

An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, Dosovitskiy et al, 2020.

Vision Transformer results

Class	Precision	Recall	F-1 score	ROC AUC	Accuracy
Negative	0.89	0.75	0.82	72.51%	74%
Positive	0.46	0.70	0.55		

Interpretability : visualizing the attention weights
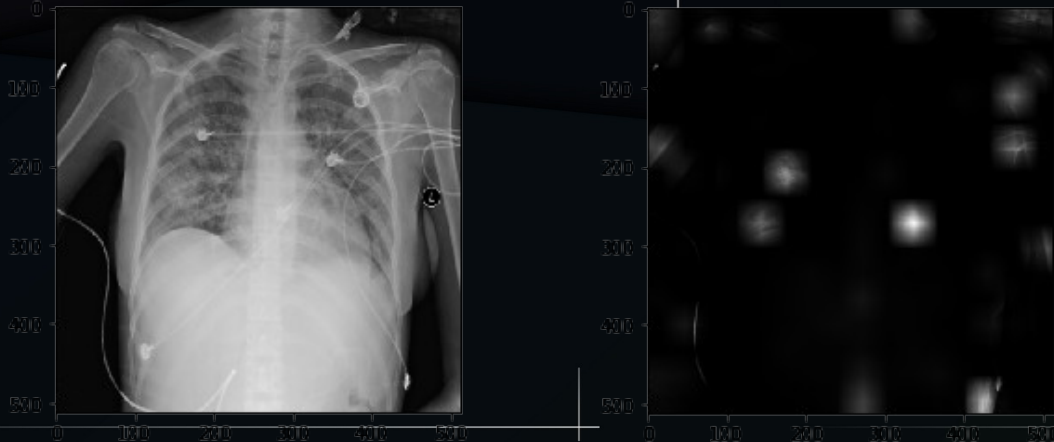
Ensemble Model

Combining Predicitons from all the mentionned models :
mean prediction of the 5 models considered.

Class	Precision	Recall	F-1 score	ROC AUC	Accuracy
Negative	0.92	0.85	0.88	79.28%	82 %
Positive	0.59	0.74	0.66		

Similar results to Resnet18 in terms of ROC AUC score but with better accuracy.

Conclusion

- The best ROC AUC score is achieved with the resnet18 model (79.71%).
- The ensemble model achieves a similar score in terms of ROC AUC (79.28%) Score but with a better accuracy and is thus prefered.
- The transformer model is more interpretable.