**Performance Comparison of Deep Neural Ensembles On Chest X-ray Images for Pneumonia Detection.**

Sai Srinath Alla

Dept of Electrical and Computer Engineering

The University of New Mexico

Albuquerque, NM 87131-0001, USA

[ssrinathalla@unm.edu](mailto:ssrinathalla@unm.edu)

Abstract:

Ensemble is the most stable and stand approach in the machine learning to ensure more stable and best predictions are made. This Paper will discuss the problem of chest x-ray image classification for pneumonia detection. In this study, I evaluate pre-trained CNN (Transfer learning) performance and construct an optimal ensemble (voting based) towards the Pneumonia detection challenge. The models I used to build the ensemble are VGG 16, InceptionResNetV2 and Inceptionv3. The results are validated based on accuracy, precision, recall and Confusion Matrix

Keywords—CNN, Feature Extractor, VGG16, InceptionV3 and InceptionResNetv2.

1. **Introduction**

One successful approach to decrease the variance of deep neural network models is to train multiple models on the same data instead of a single model and combine these models' predictions. This is called ensemble learning and reduces the variance of predictions and can result in better forecasts than any single model. Ensemble learning is creating multiple models on the same training set and combining them for the desired output. Frequently an ensemble of models performs better than any individual model. There are various types of ensembles like a model averaging ensemble,

weight averaging ensemble and voting-based ensemble. The ensemble technique used in this study is a voting-based ensemble, a simple technique in which the output is the majority of all the output for the models in the ensemble model. This voting ensemble helps in better prediction on the testing dataset than the original models.

Diagram

Description automatically generated

**Figure 1. Optimal ensemble**

1. **Background**

In this study the ensemble model build is based on the three different based on transfer learning.

* **VGG16 based model**
* **Inceptionv3 Based model**
* **InceptionResnetv2 based model**

**VGG16:**

**VGG16 [1]** is the model that model won the 1st, 2nd place on the above categories in 2014 ILSVRC challenge. In this challenge this model classified a 1000 class image dataset. This model is helpful and can be used as transfer learning. Figure 2 shows VGG16 architecture.

Chart, waterfall chart

Description automatically generated

**Figure 2. VGG16 architecture**

**Inception v3 :**

Inception v3 is a convolutional neural network that started as a module in google primarily for image classification problems. Inception v3 is the runner of ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2015 with 78.8% accuracy and less than 5% error rate on ImageNet. The Inception v3 model pretrained with ImageNet weights is available in the python Keras library. **[2]** Figure 3 shows Inception v3 architecture

**![Diagram

Description automatically generated]()**

**Figure 3. Inception v3 architecture**

**InceptionResnetv2:**

**Inception-ResNet-v2** is a convolutional neural architecture that builds on the Inception family of architectures but incorporates residual network (replacing the filter concatenation stage of the Inception architecture).**[3]** Figure 4. InceptionResnetv2 architecture

Chart

Description automatically generated

**Figure 4. InceptionResnetv2 architecture**

1. **Pseudo code**
2. **Preprocessing:**

In Preprocessing stage, I will work on the training images to prepare the training set for the model to train better by applying data augmentation.

**Data Augmentation:**

**Operations On Training dataset:**

rescale = 1/255,

rotation\_range = 30,

zoom\_range = 0.2, width\_shift\_range = 0.1, height\_shift\_range = 0.1

**Operation on test and validation dataset:**

rescale = 1/255,

**Input Size of the Deep learning model**

**VGG16** input shape is (300,300,3)

**Inception v3** input shape is (150,150,3)

**InceptionResnetv2** input shape is (150,150,3)

1. **Model:**

In this study there are three neuralnetwork using transfer learning

**VGG16 based model:** In this model the VGG16 is used as feature extractor followed by the flatten, Dense(512),dropout(0.2),Dense(512),dropout(0.2)and Dense (1).

**Inception v3 based model:**

In this model Inceptionv3 is used as feature extractor followed by the Dropout (0.5), GlobalAveragePooling2D (), Dense (128), Batch Normalization () and Dense as the final layer.

**InceptionResnetv2 based model:**

In this model inceptionv3 is used as feature extractor followed by the Flatten (), Dense (512), Dropout (0.5), Dense (512), Dropout (0.7) and Dense (1) as the final layer.

**Optimizer:**

The optimizer used in

vgg16 based model is RMSprop(lr=0.001), Inception v3 based model is RMSprop(lr=0.001) and InceptionResnetv2 is RMSprop(lr=0.001)

**Ensemble:**

In this part of the code, the combinations of the model are being created and which is used to improve accuracy than the normal models. The benefit of the voting majority ensemble is that it needs no training since voting for output does not take any learnable parameters.

**Prediction:**

In this part, the prediction of ensemble model on the Test set is done and ensemble model are evaluated using Accuracy, Recall, Precision, and Confusion matrix.

1. **Dataset**

This Dataset is the Chest X-ray Images of pediatric patients from Guangzhou Women and Children’s Medical Center [3] is divided into 3 parts (train, test, validation). This dataset is classified into two categories (Normal/Pneumonia) with 5863 images in jpeg format.

**No of** **categories**: Two Categories of images (Normal/Pneumonia).

For Training Normal class has 1341 images and Pneumonia class has 3875 images. In this Training set 10%of images is used as validation set.

For Testing Normal(0) class has 234 images and Pneumonia(1) class has 390 images.

**Dataset Sources:** (Opensource Dataset)

**Chest X-ray 2017.zip [4]**

1. **Validation metrics**

Testing of the model using the test data helps in evaluating the model using various Performance Parameters. The model is evaluated using the following metrics:

1) Accuracy,

2) Confusion Matrix

3) Precision,

5)Recall

1. **Results**

In this Ensemble model the output is the majority of all the outputs from all the models based on different feature extractors**.** The performance of the all the neural and the ensemble model is validated on the test data set of 859 images with 234 of normal images and 624 of Pneumonia images.

| **VGG16:** | loss rate: 0.53 | accuracy rate:  0.70 | precision: 0.99 | Recall:  0.60 |  |  | Recall: 0.90 |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Inceptionv3:** | loss rate:  65.54 | accuracy rate: 0.93 | Precision: 0.93 | Recall:  0.91 |  |  | Recall:0.97 |
| **InceptionResnetv2:** | loss rate: 0.27 | accuracy rate: 0.91 | precision: 0.91 | Recall:  0.91 |  |  | recall:  0.91 |

Table: 1

**confusion matrix:**

|  |  |
| --- | --- |
| model | Confusion Matrix |
| Ensemble | [[ 54 180].  [ 154 471]] |
| VGG16 | [[128 180]  [ 154 471]] |
| Inceptionv3 | [[44 190]  [ 131 494]] |
| InceptionResNetv2 | [[55 179]  [150 475]] |

Table: 2

As seen in the above there is a small amount of improvement in the ensemble model compared to the other models and this can be clearly seen comparing the confusion matrix of different models on the test set.

Even though all the models have good accuracy, precision, and recall but the confusion matrix of all the models has a lot of false positive in classifying the normal images in the models. The main problem of all the model is that these models doesn’t work perfectly on the normal images (0) of the dataset. As shown in the table 1 in the VGG16 model has very less recall over the test dataset that was evident in the confusion matrix of the VGG16 in table 2. As shown in the table 2 the performance of the ensemble model is better than the VGG16, Inceptionv3 and at the same level as the InceptionResNetv2model.

1. **Conclusion**

As we all know that the main aim of the project is to create an ensemble model to show that there is certain amount of improvement in accuracy in classification compared to the initial models. In this project as the all the transfer learning models are not up to the mark does not work so, the ensemble model does not meet the expected requirement. In general Ensemble model with good initial model works better and improves the performance on the test set than the initial or transfer learning models. The main drawback is that building multiple models with good performance is time consuming and computationally expensive. This technique is very much helpful when we have the computational power to build multiple models then this will definitely improve the accuracy than the original model.

**References**

1. https://www.geeksforgeeks.org/vgg-16-cnn-model/

# Christian Szegedy, Wei Liu, Yangqing, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, “Going Deeper With Convolutions”. arXiV:1409.484 volume 1

1. https://paperswithcode.com/method/inception-resnet-v2#:~:text=Inception%2DResNet%2Dv2%20is%20a,stage%20of%20the%20Inception%20architecture).

1. <https://data.mendeley.com/datasets/rscbjbr9sj/2>