A picture containing graphical user interface

Description automatically generatedMSc in Data Analytics

Research Project Report

Akshen Doke

Student ID: x18191592

School of Computing

National College of Ireland

Supervisor : Prof. Rashmi Gupta

A picture containing text, clipart

Description automatically generated

**National College of Ireland**

**MSc Project Report**

**School of Computing**

|  |  |  |  |
| --- | --- | --- | --- |
| **Student Name:** | Akshen Doke | | |
| **Student ID:** | x18191592 | | |
| **Programme:** | MSc in Data Analytics | **Year:** | 2019-2020 |
| **Module:** | Research Project | | |
| **Supervisor:** | Prof. Rashmi Gupta | | |
| **Submission Due Date:** | 17th December 2020. | | |
| **Project Title:** | Image Classification: Using Ensemble method to classify covid19, normal and pneumonia x-ray images. | | |
| **Word Count:**4388 | **Page Count** 21 | | |
|  |  | | |

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

|  |  |
| --- | --- |
| **Signature:** | ……………………………………………………………………………………………………………… |
| **Date:** | 17h December 2020 |

**PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST**

|  |  |
| --- | --- |
| Attach a completed copy of this sheet to each project (including multiple copies) | □ |
| **Attach a Moodle submission receipt of the online project submission,** to each project (including multiple copies). | □ |
| **You must ensure that you retain a HARD COPY of the project**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. | □ |

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

|  |  |
| --- | --- |
| **Office Use Only** | |
| Signature: |  |
| Date: |  |
| Penalty Applied (if applicable): |  |

Image Classification: Using Ensemble method to classify covid19, normal and pneumonia x-ray images.

**Abstract**

Covid 19 is a lung disease which originated from Wuhan, China and has spread across the globe resulting in loss of human lives and also causing economic loss in various countries. This paper aims to talk about how a method was developed using deep neural networks which will help us detect this disease using just chest x-ray. Since Covid 19 chest x-rays are somewhat similar to that of pneumonia, they are also being taken as a set for training the models so that they can help us detect and distinguish covid19 from pneumonia accurately without any confusion. In total 7 models were trained some created from scratch while most adopted using transfer learning methodology on the dataset available publicly and ensemble techniques of two kind i.e., voting based and weighted were used to get the final output for the given input. Overall, the output obtained was close to 91% based on popular evaluation metrics used for multi-class classification which is very good but still it can raise some concerns as even a single inaccurate diagnosis might result in a loss.

*Keywords:* Transfer Learning, Ensemble Methods, Deep Neural Networks, Chest X-ray

# Introduction

Soon after the discovery of x-rays in 1895, radiology i.e., use of x-ray technology for medical image analysis became very popular as it opened unseen doors in the world of human health treatment. Treatments of illnesses in chest region and bone damage detection became more accurate and precise. Then after the invention of computers and machine learning techniques whereby using statistical formulas and functions computers are taught to perform work like doing repetitive tasks, identification, prediction of required data etc. became prominent. Many machine learning researchers especially in the field of CNN have worked on developing a machine learning model which identifies any object presented to it. In this paper using similar principles of machine learning for classification of images we would be trying to get a highly precise model for identification of chest related diseases like pneumonia, covid-19 etc. (Wang, et al., n.d.) The advantage of using deep neural networks lies in the very word deep were in there are many layers through which the data passes and on each passing layer, we identify the features of that data and make a note of it via updating our weights and biases accordingly. To have a better performance at every stage we can train it with large-scale datasets which are unique in some ways which will also help to avoid overfitting. (Krizhevsky, et al., 2012)

In this paper, we will make use of basic exploratory data analytics i.e. data visualization and pre-processing which involves cleaning up the data and plotting charts and graphs to get a gist of data we would be using on our machine learning models, later on we will divide that dataset into train and test sets and pass the former over our machine learning models to train them and then valid their performances by checking their output on test set, this process can be repeated N number of times depending upon our satisfaction as we increase this process of train and test, the model’s performance also known as accuracy measure tends to improve overtime. One thing to keep in mind is that no matter the number of loops or epochs as known in the machine learning world, we can’t keep it running forever as at certain point it reaches a stage where it gets overfitted for our training data so we need to try permutations and combinations from the dataset to have a diverse train and test cases and also make sure that the model we are trying to train has the required layers which works very well on our dataset when it comes to extraction of features and processing or adjusting the weights in those layers accordingly. Also, we will be using Gradient-weighted Class Activation Mapping (Grad-CAM) which will highlight the region of interest which are identified by our trained model to classify the input image.

A picture containing film, sitting, laying, person

Description automatically generatedBackground pattern

Description automatically generated

Figure 1: Regular Covid X-Ray, Covid X-ray Image with Grad-CAM

After the models are built, we would club them and form an ensemble which will act as a single black box to each input given to it. Inside this ensemble each model we have trained acts independently and gives the prediction based on its own interpretation which is then counted for voting, in another case models with high performance and more reliability are weighted more i.e., they’re given a chance to vote twice and whatever is the major is considered as our final output. Such ensembles which have more than two models increases the confidence in the output and reliability.

# Literature Review

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Author(s) | Research Question/Purpose | Title | Search Terms | Data Synthesis | Finding |
| Gianluca Maguoloa, Loris Nannia | To prove that several testing protocols for recognition are always right for neural network to identify covid19 | A Critic Evaluation of Methods for COVID-19 Automatic Detection from X-Ray Images  (2020) | Covid-19; Covid-19 Diagnosis; Convolutional Neural Networks; X- Ray Images | 108,948 images of 32,717 different patients, classified into 8 different sectors |  |
| T. Rahman, M. E. H. Chowdhury, A. Khandakar, Khandaker R. I., Khandaker F. Islam, Z. B. Mahbub, M. A. Kadir and S. Kashem | Report on advances in accurate detection of pneumonia using transfer learning methods | Transfer Learning with Deep Convolutional Neural Network (CNN) for Pneumonia Detection Using Chest X-ray.  (2020) | pneumonia; bacterial and viral pneumonia; chest X-ray; deep learning; transfer learning; image processing | 5247 chest X-ray images consisting of bacterial, viral, and normal chest x-rays images were pre-processed and trained for the transfer learning-based classification | Authors used 4 models out of which DenseNet201 outperformed rest of the models in classification. |
| R. Selvaraju, M. Cogswell, Abhishek Das, R. Vedantam, D. Parikh, D. Batra | Proposed a Novel way to make any CNN-based model more transparent in visual explanation | Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization  (2019) | CNN,  VGG16,  ResNet |  | Results show that the authors were able to achieve what they intended to do. |
| Ivo M. Baltruschat, Hannes Nickisch, Michael Grass, Tobias Knopp, and Axel Saalbach | Building a model using transfer learning and testing with and without fine tuning it. | Comparison of Deep Learning Approaches for Multi-Label Chest X-Ray Classification  (2019) | X-Ray, Deep Learning, Convolutional Neural Networks | ChestX-ray14 consists of 112,120 frontal chest X-rays from 30,805 patients |  |
| X. Wang, Y. Peng, Le Lu, Z. Lu, M. Bagheri, Ronald M. Summers | To demonstrate that thoracic diseases can be detected and even spatially-located via a unified weakly-supervised multi-label image classification and disease localization framework | ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases  (2017) |  | 108,948 frontal view X-ray images of 32,717 unique patients with the text mined eight disease image labels | Authors conducted quantitative performance on the ChestX-ray8 db using transfer learning techniques. |
| Mohammad Tariqul Islam, Md Abdul Aowal, Ahmed Tahseen Minhaz, Khalid Ashraf | DCNN based classification and localization on the publicly available datasets | Abnormality Detection and Localization in Chest X-Rays using Deep Convolutional Neural Networks(2017) | Abnormality detection | 7284 CXRs, both frontal and lateral images. Another one with 247 chest X-rays, among which 154 have lung nodules and 93 have no nodules. | The DCNN architecture they used didn’t perform well on all abnormalities. |
| P. Rajpurkar, J. Irvin, K. Zhu, B. Yang , H. Mehta, T. Duan, D. Ding, A. Bagul, Robyn L. Ball, Curtis L. , Kate Sh., M. Lungren, Andrew Y. Ng, | Develop an Algorithm which can detect pneumonia from chest x-ray. | CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning  (2017) | CheXNet | Chest Xray dataset of 100,000 frontal view X-ray images with 14 diseases. | Authors created a 121 Dense Convolutional Network called CheXNet which was able to perform better than radiologist given the data under certain circumstances. |
| Juan Manuel Carrillo-de-Gea, Ginés García-Mateos, José Luis Fernández-Alemán and José Luis Hernández-Hernández | Objective of the paper is to perform an automatic normality classification of posteroanterior chest radiographs. | A Computer-Aided Detection System for Digital Chest Radiographs  (2016) | Medical Imaging | DICOM images of chest radiographs (23 women and 25 men) were provided by HGURSM, Spain to perform the test. | A new approach for detection of normality in chest radiographies was given which is based on LBP. |
| Kaiming He,Xiangyu Zhang, Shaoqing Ren, Jian Sun |  | Deep Residual Learning for Image Recognition  (2015) | Deep neural network, ILSVRC 2015,  Residual Learning,  COCO object detection. | ImageNet dataset,  CIFAR-10 |  |
| Saad, Mohd Nizam; Muda, Zurina; Sahari, Noraidh; Hamid, Hamzaini Abdul | Segmentation of Lungs based on object detection technqiues. | Image Segmentation for Lung region in Chest Xray Images using Edge Detection and Morphology(2014) | Lungs, Image edge detection, shape, noise, biomedical imaging. | 247 CXR images with  standard size of 2048 x 2048 pixel | Researcher did built a method for lung segmentation however there are some changes and improvements needed as pointed out by the authors themselves. |
| Alex Krizhevsky,  Ilya Sutskever,  Geoffrey E. Hinton | Create a Model which works efficiently over huge datasets | ImageNet Classification with Deep Convolutional Neural Networks  (2012) |  | 1.2 million high-resolution images were trained for ImageNet LSVRC-2010 contest into the 1000 classes | Large networks tend to achieve good results on huge datasets with supervised learning. Authors want to try the same on video sequence. |

**INTRODUCTION**

To begin with as we know that the use of deep neural network technique called convolutional neural networks (CNN hereafter) in medical image diagnosis is not a new phenomenon also a lot of work has been done in the world of computer vision which has helped us develop models which can be able to classify various objects properly. In this section we would be discussing some work of previous researchers and how we can understand, learn and fill-in the gaps of their work and help reach the common objective of uplifting medical image diagnosis.

Use of Transfer Learning Methodologies in Chest X-ray classification

As per (Tawsifur Rahman, 2020)’s research, one of the greatest causes for child fatalities around the world is pneumonia because of which more than 1 million children die among which 18% of the total deaths is of children less than 5 years of age at the same time globally around 2 billion people suffered from pneumonia. They also found that chest x rays among others is the best method to detect pneumonia even though it’s not very clear and sometimes leads to misclassification to other diseases. In their paper there objective is to provide a CNN-based transfer learning approach using 4 pre-trained models which were ResNet18, DenseNet201, SqueezeNet and AlexNet to detect and classify two forms of known pneumonia which are bacterial and viral. They worked on a dataset which comprised of mixed cases of pneumonia i.e. viral and bacterial infected along with normal pictures of around 5247 chest x-ray images, it didn’t have any cases where both viral and bacterial infection was found in one patient. They also made use of data augmentation so as to increase the number of datapoints and improve the quality of the dataset as it’s a better and inexpensive alternative to collecting new data. Since they were focusing on a very niche disease i.e. pneumonia, they were able to get a pretty decent accuracy for their models DenseNet201 leading the way.

Recent work on Covid-19 data

Covid-19 the pandemic which has stopped the wheel of the world on all fronts also has some window of opportunities for research. To address this (Maguolo & Nannia, 2020) compared different methods and testing protocols used to automatically diagnose covid-19 from x-ray pictures and they were able to showcase that identical outcomes can be achieved using those pictures which did not had most part of the lungs. They made use of four different datasets available publicly online. They also performed a unique approach wherein they first covered the images with a black box in the centre which covered most of the lung region and then it was sent to AlexNet for classification based on labelled set to those images before they were transformed with black box in front. The model was trained to recognize the source of the dataset which in turn tried to prove that models are biased as they learn to classify more on the source dataset rather than on correct medical diagnosis. This was a very novel way of approaching not just covid-19 diagnosis but any serious medical image classifying problem in general.

AlexNet

Geffrey Hilton is regarded as one of the most pioneering figure in the area of research related to deep neural networks, in his paper along with his colleagues they developed a deep convolutional model named AlexNet (Kaiming He, 2015) which was able to classify more than 1 million high- resolution pictures in a contest to differentiate up to 1000 different classes. They did this by training almost 600,000+ neurons which consisted of 5 convolutional layers. For training their model they made use of a sub-set of ImageNet dataset which around 1.2 million training pictures and roughly 50,000 to 150,000 for validation and testing. Considering the size of dataset and length of network, they spread out the network across two GTX 580 GPUs each of 3GB in memory using parallelism. Since their network architecture was dense, they had to face an issue of overfitting and to combat that they made use of two techniques namely Data Augmentation and Dropout. Advantage of using data augmentation is that the transformed images which they’ve generated using the existing ones don’t need to be stored on the disk instead they’re generated via code on CPU while the GPU trains on the preceding array of pictures. Also, in the technique called Dropout, every neuron in the hidden layer with a probability of 0.5 was set to zero in this way they didn’t contribute to the backpropagation. In this way they achieved good output on an immensely challenging dataset using just supervised learning (Krizhevsky, et al., 2012).

Some previous work on pneumonia detection using chest x-ray images

As per (Rajpurkar, et al., 2017)’s study, more than a million people are affected with pneumonia every year in the United States alone out of which nearly 50,000 people die due to this disease. To tackle this issue they built a 121-layer dense convolutional network model called CheXNet which has dense connections and batch normalization for optimization of deep network tractable, which they claim can detect pneumonia in par with radiologists and in some cases even exceed their predictions also it outperformed all other best known published results for 14 different pathologies in the dataset called ChestX-ray14. Although while validating their claims there were some limitations which are worth considering, firstly they made use of only frontal radiographs to test their model and same was given to radiologists also the radiologists involved in this project and the model were not given permission to use current or past patient information in any way or manner which impacts the performance of radiologists in general. Overall the model generated performed quite well and the results obtained showed a way to the machine learning community that medical imaging technology can be automated to improve the healthcare delivery where access to radiologists is not possible.

Image Segmentation

Image Segmentation has always been an interesting area of research which distinguishes objects area from one and another in pictures, early edge detection methods like Laplacian, Prewitt and Sobel were used but the problem with that was that it failed in cases where they acted as high pass filter and were affected by noise present in the picture, to overcome this issue (Saad, et al., 2014) proposed a better algorithm which is good at edge detection and can cope up with lower and upper threshold units for image noise and combined it with morphology methods for fined outputs. In order to address this issue, they made use of Euler number method.

Multi-Label Chest X-Ray Classification

As per (Baltruschat, et al., 2018) over 23,000 chest x-ray images didn’t went through certified radiologist at Queen Alexandra Medical Institution alone and a handful of patients suffered because their x-ray images weren’t being properly accessed. This is not just the only medical institution which has these kind of issues as aging population is growing due to increase in life expectancy, the requirement for healthcare workers and clinicians has also driven up. This observation made them to build a computer vision model which makes use of transfer learning and classifies images based on the features learned by their model. The model they created is loosed based on ResNet-50, they fined tuned some layers of that model to suit their needs at the same time they made use of Grad-CAM which helped them get insights their model. The dataset they used was of 14 different chest related diseases which we’ve also seen in other researches cited above. There optimized model ResNet-38 achieved better results in 5 out of 14 classes compared to the works they’ve cited also they observed substantial variability when different splits of the same dataset were considered.

Class Visualization

(Selvaraju, et al., 2019) devise a method to identify neurons which were important through the Grad-CAM technique as even though computer vision models enable us to classify things, detect objects and do semantic segmentations, it’s still hard to interpret individual components. In order to get a reliable model we need to know how and why the model came to the predicted conclusion. This interpretability is not just useful when it comes to understand the conclusion part of the model but also while training it as it can help us make the training process smoother by helping us debug and fine tune the model efficiently, diagnosing classification models for considerable biased errors. As shown in the example given that the model to classify doctors and nurses was trained on a dataset which had 78% images of doctors who were men and 93% images of nurses who were women so the model instead of checking for props like stethoscope went for features like facial dimensions, hair length etc and classified all women as nurses while all men as doctors. This issue was identified using grad-cam technique which highlights the region of interest based on which the model made it’s classification. Further the model was retrained with more images this time balancing it out with more female doctors and more male nurses to lower the bias results. This helped to demonstrate that grad-cam technique can help not just visualize the output of the trained model but also fix issues if any.

Detection of Abnormality in Digital Chest Radiography with the aid of Computer Detection Systems

(Juan Manuel Carrillo-de-Gea, 2016) proposed a method to perform automatic normality classification of posteroanterior digital chest radiographs which is able to detect anything which can be classified as different from normality. Initially images of 3000 by 3000 pixels with depth of 12 bits per pixel were taken with an average age of 55. These images were reduced in pixel depth by 4 bits i.e. from 12 to 8 after which decimation is applied to the image using super sampling interpolation which reduced the size by 2000 i.e. to 1000 x 1000 which is considered the standard resolution for further steps.

In the next step of segmentation, image is segmented to locate the position of both lungs which then helps them to determine the region of interest. Samples of both left and right lungs were extracted and the location with maximum correlation is selected as the expected position of each lung after which a grid of 3 by 4 region is generated. For feature extraction, they made use of LBP histogram for each reach obtained, later these features were classified based on distances between histograms i.e. using Bhattacharyya distance two histograms are computed. Later on the experimental results obtained from the classification were 90% for the best classifier speaking about the disadvantage, the method implied by the researchers relies mainly on texture information which in case of some diseases which affect only the intensity of the images would be hard to detect.

Detection of Abnormality and it’s Localization using DNN in Chest X-ray region.

(Mohammad Tariqul Islam, 2017) made use of ensemble models to improve the classification accuracy for abnormality detection in chest region for x rays compared to a single model. They made use of some open datasets like JSRT dataset which contained around 250 images of which 154 had lung nodules i.e. malignant and benign cases and 93 didn’t have any nodules each of size 2048 x 2048 pixels and a gray-scale color depth of 12 bits, Shenzhen Dataset from China which had two classes i.e. normal and tuberculosis and Indiana chest X-ray dataset the largest amongst the 3 with 7284 chest x-ray images of both frontal and lateral with diseases like cardiomegaly, pulmonary and pleural effusion on which the first studied performances of some already built deep convolutional network over different abnormalities. For their experiment, they trained many models via transfer learning like AlexNet, VGG based and ResNet and for each they found some model performing better than other on certain diseases while some giving high sensitivity and specificity etc. For the ensemble of models, they trained the variants of the same mentioned model and used a simple linear averaging of probability on individual model as bagging and boosting are implied it might result in a biased model as the number of dataset is low for such huge model with multiple layers. They also tried to increase the number of models in the ensemble and found that gradually a consistent performance was seen after 9 models. For better understanding, like other researchers listed above they too went to visual depiction of model prediction to actually understand how and why the models were making the classification they were giving rather than just blindly treating it as a black box. Speaking in simple words they wanted to identify the features which contributed more to the output of the model. They took the localization approach for cardiomegaly abnormality and highlighted the 20% area which was more sensitive to the region where the heart is larger than normal heart. They performed the same experiment on around 50 samples of cardiomegaly and normal images and found the result to be mostly consistent. Based on the localization observation approach seeing sensitive region they came to a conclusion that characteristic features in the shape of heart and its surrounding regions is alone to detect cardiomegaly, the lungs are less important when it comes to detecting it. However while applying similar methods on pointed features for that of nature like bone fracture and lung nodule, the localization method failed.

Conclusion

After going through the literature mentioned above, it was seen that ample of research has been carried out in the area of image detection, medical image analysis etc even though some researchers in the domain of chest x- ray analysis and classification claim to have developed a model which performed better than a human radiologist, it’s interesting to note that the model built was only trained with frontal chest x-ray images also the human radiologists were not given any supporting data such as patient age, gender and previous health condition which still leaves a room for improvement and as a researcher myself it makes me curious to know what could have been the results if those kind of data would have been provided.

In the upcoming sections we would be seeing the proposed research methodology for our selected approach of building a model via transfer and ensemble learning.

# Background and Motivation

Due to rapid industrialization which leads to air pollution and unproductive human habit like smoking has always made way to rising number of cases related to lung disease also recently a new type of coronavirus called covid-19 which is spreading rapidly across the globe, demand for radiologist have increased but the supply and reach is low because radiology being a critical skill not many people can be trained quickly and also most of the radiologists are based in cities hence the demand in undeveloped rural areas isn’t met. This made many computer vision researchers work on a model which will help radiologists do their job in less time and if perfected even fill in the gap where required. (Krizhevsky, et al., 2012) (Maguolo & Nanni, 2020) (Wang, et al., n.d.) (Tawsifur Rahman, 2020)

# Research Question

*“With use of ensemble technique on machine learning models can we get a reliable and efficient output for chest x-ray image classification?”*

The important research question proposed here is to build, modify and study prominent deep learning methodologies for accurate, dependable model on a relatively large dataset.

# Research Objective

Basically, we will train a considerable amount of dataset over certain pre-trained models and fine-tune it by changing certain layers in its architecture and try to optimize its performance. Also, we will use ensemble technique in which multiple models are used and their output is evaluated for reliable results. The fact that any misjudgement of radiologist or machine learning model in classifying the x-ray image in wrong group might lead to inefficient treatment of the patient so we need to train our model with as much data as possible and get a strong model which is dependable in times like pandemics where human healthcare workers are very much occupied with treating patients and such models will help reduce the time of identifying already known diseases.

# Research Methodology

In most cases whenever an x-ray image of patient is taken, first of all it is examined by a radiologist who writes a report of his/her interpretation and then the x-ray image and report are forwarded to the doctor who has asked for it. The research in this paper focuses on reducing this time by automating the process of detecting covid19 and helping doctors to take decisions quickly in order to save patients life. The work for this research is carried out by following simple process illustrated below.

Modeling

Evaluation

Data Collection and Preparation

Comparison and Interpretation of results

Transfer Learning and Ensembles

A picture containing text, screenshot

Description automatically generated

Source: Google Images

Figure 2

The Novel approach proposed in this paper is the use of ensemble technique to make the final predict more reliable and efficient.

## Data Collection, Exploration and Preparation

Since the beginning of January 2020, the virus has been spreading across the globe like wildfire, some medical institutions and researchers also published various data on this virus which ranged from its symptoms, measures to take for prevention and if highly suspected, how to contact. For collecting data over Covid 19 we made use of the following available dataset [COVID-19 RADIOGRAPHY DATABASE](https://www.kaggle.com/tawsifurrahman/covid19-radiography-database)**,** also while trying to detect covid19 from chest x-rays manual it was observed that radiologist got confused between covid19 and pneumonia hence we also take pneumonia images from [here](https://github.com/ieee8023/covid-chestxray-dataset) so that our models learn how to distinguish and don’t fall for the similarity trap.

Images downloaded from the provider were collected from various sources also the provider made sure to hide vital information of the patient while giving the x-ray images considering the privacy issues hence in order to make them consist in terms of name and extension the data were renamed, and their extensions were set to .png using a simple bash script. The results stored in the folder by the name of type they belonged to i.e., covid19, normal and pneumonia. Later they would be split into two categories one called training and the other for testing, the ratio for which is 70-30 respectively. The split is carried out using random shuffling of images. Also when the split is made, each class would be assigned a label which in this case would be the index of the folders occurrence i.e., 0, 1 and 2.

This newly created data folders were then uploaded on Google Drive as we need them to connect with our cloud infrastructure called Google Colab. This cloud environment allows us to be flexible with our use of resources on demand and also has all the packages required to achieve the objective of our project.

*Data Pre-processing and augmentation*

Overall, our data count is a bit low so in order to make most of our dataset we use a technique called data augmentation wherein we perform operations like rotation, transformation by flipping, zooming and scaling so that our model to be trained gets various inputs and also it doesn’t see the same image twice as this is helpful to avoid a training problem called over-fitting of data. Also, before setting the model to train on our dataset we set the number of batch size which makes sure that only that many numbers of images are given in trained at the particular moment by updating and validating the weights of our model. This feature setup can also be called as hyperparameter optimization.

## Modelling

The pre-processed data is then sent as input to our 7 models and output from each one is later combined to get our polling results.

A picture containing text, screenshot

Description automatically generated

[0,1,2]

model

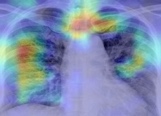


Figure 3

For first 5 models which we are going to use in our ensembles, we would be using transfer learning methodology which enables us to adopt a pre-trained model for our current dataset.

The models used in the ensemble network are as follows: -

* DenseNet

As the name suggests, it has many layers hence the name dense, the speciality about this neural network architectural model is that output from each layer is given as input to every layer coming after it.

Diagram, engineering drawing

Description automatically generated

Figure 4 Source: (Gao Huang, 2018)

* Resnet

This is one of the most successful neural network architectures ever built for computer vision problems as it showed that the more layers you add to a network, it is not the case that the performance will increase contrary to the popular belief reason being is that after each layer you have feature extracted which extract and update the weights of the model and downsize further but this can’t be done after a certain period as the input gets very small. (Krizhevsky, et al., 2012)

Chart, diagram

Description automatically generatedDiagram

Description automatically generated

Figure 5 Source: (K. He, 2016)

They also introduced a technique called skip layer where in a layer can jump directly by skipping it’s very next layer and pass on the output to the one after that, this helps to make more complex connections via just few layers.

* Alexnet

As discussed earlier in the literature review section, this network was built by Geffrey Hilton along with his team, it is bigger and deeper than LeNet and includes ReLu activation, Dropout layers etc to avoid overfitting issues. It was also the winner of an Image Classification problem called ILSVRC-2012 (Krizhevsky, et al., 2012).

* VGG16

VGG stands for Visual Geometric Group and 16 is the number of layers present in the model it is also present in a 19-layer variant but for this project we use the 16 layer one. It is bigger and deeper in comparison to Alexnet. This model is also available in keras package and it takes 224 \* 224 size of image as input (Simonyan & Zisserman, 2015).

* Xception

This model was inspired by Inception model where in the modules of inception are replaced by depthwise separable convolutions, it has the same number of parameters (Chollet, 2017)like that of Inception V3 but it still slightly outperforms V3 on the ImageNet dataset.

* NASnet

“NAS” stands for Neural Architecture Search which makes use of a reinforcement learning search methodology in order to optimize configuration of the architecture. It is mostly used for object detection. Basically, it is kind of an automated way to create the most efficient model for the dataset we have, each model has a controller which has certain blocks and cells, each blocks consists of filters, pooling layers called operators and each cell is a collection of 5 blocks. When the data is feed, the controller generates a child model which is trained on our input data, it tried all permutation and combinations of cells to get the highest validation accuracy and at the end the controller is rewarded. NASnet often gives unimaginable neural architecture and works efficiently for the particular dataset which it was provided (Zoph, 2018).

* MyModel

This model was built from scratch by combining some layers from AlexNet and DenseNet, it was an experimental model, yet it performed as good as the previous ones when trained on the given dataset. Input for this model was standard 224 by 224 with 6 convolutional layers and 2 fully connected layers along with dropouts at each end of the fully connected layers finally with a dense layer having SoftMax as activation function.

*Class Visualization Tool*

As presented in (Selvaraju, et al., 2019), keras the package used to implement the project also provides us an option to visualize the last layers and see which is the region of interest or area in the image based on which the model is making the given prediction. This technique helps us to understand the behaviour of our model and if require further fine-tune it by optimizing and updating the hyperparameters of that model.

A picture containing blur

Description automatically generated

Figure 6 Source: NASNet output

The above image is the interpretation of NASNet for a covid19 input image, the model classified it correctly as a covid19 image and we can also see the regions based on which the model has made such prediction. However, note that ensemble being a collection of models, the layer configuration of each model present in the ensemble network is different hence getting such a mapping for an output from the ensemble would be impossible

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Epochs** | **Origin** | **Input Size** | **Accuracy** |
| AlexNet | 15 | Coded from scratch | [150, 150] | 0.9268 |
| Vgg16 | 25 | Transfer Learning | [299, 299] | 0.9790 |
| NASNet | 15 | Transfer Learning | [331, 331] | 0.9705 |
| Xception | 15 | Transfer Learning | [299, 299] | 0.9732 |
| DenseNet | 25 | Transfer Learning | [224, 224] | 0.9677 |
| ResNet | 25 | Transfer Learning | [224, 224] | 0.8999 |
| MyModel | 15 | Coded from scratch | [224, 224] | 0.9502 |

Table 1

As given above, the accuracy of each of our 7 models trained on the training set, they were run for the mentioned epochs i.e., number of training time and the size of input image for each model depended on the size of input layer which is the first layer it had. The reason why epochs are different for each model is because it was observed that for some models after a certain number of epochs, the resultant model was overfitting, or it had no accuracy improvement and was just consuming GPU time. After the training, each model was saved with “.h5” extension and downloaded.

Now these trained models will be clubbed together to form an ensemble network and an input would be given to the ensemble network which would go to each of these models present inside the box.

*Ensembles*

NasNet

AlexNet

DenseNet

Xception

Resnet

MyModel

VGG16

A picture containing text, screenshot

Description automatically generated

0, 1, 2

Figure 7

As stated in the data preparation stage that each class would be assigned a label as per the index of their occurrences, we will also get the output as class indexes from the ensembles based on the two ways which are polling and weighted polling.

*Polling*

One of the easiest to understand and simple to implement method of ensemble network is polling/voting based, here each model in the ensemble gives its output which is then counted to check which class has been voted the most and the answer to that is considered the answer to the input of the ensemble. Since each model present in the ensemble can only vote once so in this case, the number of models in the ensemble network plays a very crucial role in order to increase or decrease the overall accuracy of the ensemble as a whole. In some cases, to maintain reliability only the models with high accuracy are appended in the ensemble network, since in our case all models are trained well as in Figure 5, we include all the models in this ensemble network the performance of which is discussed in the results section.

*Weighted Polling*

In this method of ensemble network, certain models in the ensemble are given multiple chances to vote or check the input image and their output is considered 2x times in count. In our case since models with complex architectures and good efficiency like NasNet, Xception and DenseNet were taken into consideration for these perks and their predictions were counted twice. This method of weighted polling also known as weighted averaging helps to increase the overall reliability and tends to perform better for very large and complex datasets.

## Evaluation Metrics

Later we evaluated both the methods with popular evaluation metrics for classification problems which are confusion matrix, F1 score and Matthews Correlation Coefficient. Also, the library which was used to implement these called ‘sklearn’ had a metrics function to create a report for classification problems which gave the precision, recall, macro average and weighted average.

Confusion Matrix for Polling based Ensemble is plotted below.

Chart, treemap chart

Description automatically generated

Figure 8

Around 690 images from all three classes were included for the test, the calculated accuracy for polling-based ensemble is 0.9058, individual F1 score for covid19, normal and pneumonia is as follows 0.87, 0.81 and 0.94 respectively. F1 score as the formula goes

F1 = 2\* (Recall \* Precision)

(Recall + Precision)

helps us to strike a balance among Recall and Precision.[[1]](#footnote-1) The Precision and Recall for covid19 was 1 and 0.77, for normal it was 0.68 and 1.00 and for pneumonia it was 1.0 and 0.89, although the precision for covid19 and pneumonia was very efficient but for normal it was not up to mark which raises concern.

Chart, treemap chart

Description automatically generated

Figure 9

For our weight-based polling ensemble, the accuracy only improved by a margin of 0.5% but considering the amount of data we had it was predictable, also since deep neural networks tend to improve their performance overtime with new information, with this rate we can expect the weight-based ensemble to outperform the former when trained with more input data.

# Results and Discussion

In this section we would be discussing the outcome of the project based on the evaluation metrics seen above. Foremost, we loaded the dataset and performed certain image augmentation techniques as to have good amount of input data to train, the training time for each model used in the ensemble was around 2 hours on an average, considering the accuracies of our models trained individually on the dataset we got most of the models with accuracy of close to 90% and above, this is an indication that our models are able to understand and distinguish the difference between covid19 and pneumonia in most cases. As seen in the evaluation metrics, our ensemble networks gave an accuracy of about 90%.

Ensemble networks being the novelty of this project over chest x-ray classification for covid19, pneumonia and normal images we successfully implemented the mentioned approach and built a reliable model. However, since we are working with medical data and considering the fast-changing nature of medical research we might have some ethical concerns which are stated in the coming sections.

# Conclusion and Ethical Concerns

Automated classification of covid19 and pneumonia x-rays can help save a lot of time and in return can benefit the patient to recover quickly. In this project we successfully created a reliable and one-of-a-kind model for covid19 classification to the best of my knowledge as covid19 being a very new disease there aren’t any models with such approach available.

In future, the as more x-ray image data related to covid19 gets available, the models can be retrained to improve the performance further also as mentioned earlier ensemble networks have a reputation of winning image classification competitions hence same approach can be applied to create automated detections for brain and heart related diseases. Even other lung related diseases can be included, and models retrained to make this very ensemble network a generic model for detection of any lung related disease. (Tawsifur Rahman, 2020). Considering we were working on medical data, the factor of ethical concerns remains a point as even a small misclassification or error can lead to loss of human life hence results from the automated model must be reported to a doctor who in the end will take final decision as to how to treat the patient for speedy recovery.

# Acknowledgement

I would sincerely like to thank and convey gratitude to my supervisor Professor Rashmi Gupta for her tremendous support and guidance throughout the project phase, also like to thank other faculties from National College of Ireland who helped me as an when required and authors and researchers who’s papers I’ve referred and cited for this project as development in science and research is an incremental step I’ve stood over shoulders of these giants and tried contributing my drop in the ocean of scientific endeavours.

1. <https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9> [↑](#footnote-ref-1)