Classification of Pneumonia from chest X-rays using Transfer Learning

MSc Research Project
MSc in Data Analytics

Tushar Shailesh Dalvi
Student ID: x18134301

School of Computing
National College of Ireland

Supervisor: Mr. Noel Cosgrave

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Tushar Shailesh Dalvi
Student ID:	x18134301
Programme:	MSc in Data Analytics
Year:	2019
Module:	MSc Research Project
Supervisor:	Mr. Noel Cosgrave
Submission Due Date:	12/12/2019
Project Title:	Classification of Pneumonia from chest X-rays using Transfer Learning
Word Count:	6567
Page Count:	20

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:	11th December 2019

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):	

Classification of Pneumonia from chest X-rays using Transfer Learning

Tushar Shailesh Dalvi
x18134301

Abstract

Pneumonia can be life-threatening for people with weak immune systems, in which the alveoli filled with fluid that makes it hard to pass oxygen throughout the bloodstream. Detecting pneumonia is from a chest X-ray is not only expansive but also time-consuming for normal people. Throughout this research introduced a machine learning technique to classify pneumonia from Chest X-ray Images. Most of the medical datasets having class imbalance issues in the dataset. The Data augmentation technique used to reduce the class imbalance from the dataset, Horizontal Flip, width shift and height shift techniques used to complete the augmentation technique. Used VGG19 as a base architecture and ImageNet weights added for the transfer learning approach, also Removing initial layers and adding some more dense layers helped to discover new possibilities. After testing proposed model on testing data, we are able to achieve 98% recall and 82% of precision. As compare with state of the art technique, the proposed method able to achieve high recall but that compromises with Precision.

1 Introduction

Pneumonia is one of the common diseases throughout human history, it is derived from the Greek word (pneúmōn) meaning "lung". Cough, fever, breathing difficulties are the symptoms of pneumonia which are Initially described by the Hippocrates in 460-370 BC. Pneumonia is more common in old and very young people because they have weaker lungs. Even mild viral pneumonia also can be life-threatening in those peoples whose immune systems are weak. Pneumonia primarily caused by bacteria and viruses and less commonly by fungi and parasites. Pneumonia usually begins as a higher respiration tract contamination that movements into the lower breathing tract. In this, the alveoli (air sac) which is inside the lungs are filled with the fluid, which makes harder for the lungs to pass oxygen into bloodstream. As per the national vital statistics report Pneumonia is on 8th rank among 10th leading death cause (Heron; 2018). In 2016, 51537 death caused by pneumonia in Alone USA. And according to World Health Organization 15% of all deaths of children under 15 years old are caused by pneumonia, which means Pneumonia killed 808694 children in 2017. In the term of antibiotic treatment to treat pneumonia, it took around US$ 109 million per year. Of course, the advantage of this treatment can only take those patients who knows they are affected with pneumonia in early stages. It takes lots of efforts and manpower to detect pneumonia in patients. Day by day for pneumonia patients getting hard to cure the diseases, and for that we need experienced and well-trained doctors so they can identify the pneumonia and help to cure the illness.

Identifying pneumonia from a chest x-ray is one of the famous techniques developed by doctors, in which fluid is identified in lungs. It is hard for any layman to find fluid in lungs and workload on specially trained practiser increase the risk of pneumonia infected people. To reduce this risk, we need to develop a computer-based system that can identify pneumonia from a chest x-ray. Currently, very few systems can identify the organs and tissue in the medical image analysis, and still makes an issue for implementing AI-based Solution. this research is conducted to develop a system that can identify pneumonia from a chest x-ray, also the primary goal to achieve high recall on classification of disease. In this research custom Vgg19 model developed to carry out this research.
Hence From the above discussion, the research question will be:

Can the custom VGG19 model help to improve the recall of pneumonia detection from chest X-ray images as compare state of the art technology?

In this study, we will use the CNN based model which will train for the image classification process. In section 2, we will see the detail description of previously done researches and find out the best possible technique for classification technique. In section 3 explained the detailed methodology with subsection like Business Requirement, setup, Data understanding, modelling, and In section 4 overall flow of research is explained. section 5 represented proper implementation method. After that section 6 provides insight of our model performance and in next section 7 result will be discussed. finally in section 8 conclusion and future work is explained.

2 Related Work

In the Field of Pneumonia classification, so many researchers contributed their work to take motivation. Everyone used different techniques, different methodology and different datasets to get state of the art output in same field. most of the Researchers used CNN based Models on different datasets, hence we dividing the literature review into sections with Different available datasets.

2.0.1 RSNA Dataset:

Researcher Liang and Zheng (2019) performed pneumonia detection on Chest X-ray dataset which contained a total of 5856 X-ray images of pneumonia and normal as well, also the number of images of normal as compare pneumonia was low, the researcher used image augmentation to reduce the overfitting. The researcher used CNN techniques in which developed their own CNN Architecture with 49 convolutional Layers and 2 dense layers researcher used residual network-based models to differentiate the actual observed values and estimated values this helps the model to learn on small changes. In this research, the researcher used other models like VGG16 (Simonyan and Zisserman; 2015), Densnet121, Xception and inception V3 used to compare the result with the custom model. The result of this research got the highest recall and precision from custom model, 0.967 and 0.891 respectively as compare with another algorithm.

In previously researcher Liang and Zheng (2019) developed its own CNN architecture with 49 Convolutional layers and 2 dense layers for pneumonia classification but another researcher O'Quinn et al. (2019) used the Alexnet network on the same the RSNA dataset. Alexnet network created to classify the 1000 different classes, which consist of 650,000 neurons and 60 million parameters including 5 convolutional layers followed by max-pooling layer and 3 fully connected layers with 1000-way SoftMax. To carried out this research, the researcher used the fine-tune pre-trained model using transfer learning and replaced the last 3 layers replaced by 3 fully connected layers for classifying two categories. For further execution, this program researcher divided data in the ratio of 70-30%, in which 70% used for training purposes and remaining used as validation dataset. With the 7922 training images researcher set batch size of 128 images for 20 epochs it took around 62 iterations to complete single epoch. With this setting Researcher able to achieve 72% initial accuracy on validation dataset which used separately. In conclusion, researcher proposed the image augmentation on dataset to improve the result. The same approach used in another research but with image augmentation researcher Stephen et al. (2019) developed own network with convolution, max pooling, and classification layers combined together. The researcher created several dataset sizes to check different approaches in which Researcher able to achieve highest of 95% training accuracy and 93.73% accuracy on validation dataset which can show us the importance of image augmentation.

In spite of the previous two researchers, this researcher used a different approach, researcher Islam et al. (2019) used compressed Sensing based on deep learning framework for pneumonia detection from X-ray images. In this paper, the researcher used a simple multilayer network with 4 convolutional and 3 fully connected layers as hidden layers also for non-linearity ReLU (Rectified Linear Unit) activation function used. The whole dataset was distributed in 70%, 25%, 5% ratio for training, validation and testing purpose. The researcher used Peak Signal to Noise Ratio (PSNR) and mean Structural SIMilarity (SSIM) measures to calculate the reconstructed X-Ray images. The user runs this CNN model for 500 epochs and after 400 epochs model achieved 98.82% and 99.80% accuracy on training and validation. Throughout this whole research researcher used 3 more methods to compare against Proposed Method, in which F-cMeans got the least prediction accuracy 85.87%, another method DL (Deep Learning) and ChexNet algorithm got 91.16% and 95.28% accuracy, and the proposed model able to achieve 97.34% Prediction Accuracy.

According to Sirazitdinov et al. (2019) ensemble method is also becoming handy for pneumonia classification hence researcher used an ensemble of RetinaNet and Mask R-CNN to detect pneumonia from chest X-ray. Initially, the base network trained with the COCO network and use image augmentation like Random Brightness, random degree rotation. In the training phase both models trained to predict the pneumonia region then NonMaximum Suppression (NMS) was used to set threshold in predicted region. Afterword depending on similarity of pneumonia region model getting selected and if both models have different areas of region in that case ensemble method uses to select area with maximum confidence. With this method Sirazitdinov et al. (2019) got 0.790 accuracy for pneumonia prediction. The usefulness for ensemble method is further stated in the research which is published by Ko et al. (2019) with the help of RetinaNet and Mask R-CNN networks ensemble model created and for the accuracy measurement mean average precision (mAP) used. To improve the mAP models trained individually and adjusted ratio of weight for more accuracy. This process helps to achieve the highest mAP of 0.21746. Jaiswal et al. (2019) used same Mask RCNN method with little changes initially set the threshold on ResNet50 and ResNet101 to get optimum results, and both models combined as ensemble method to get maximum prediction value. In this research image augmentation, dropout and L2 Regularization used to reduce overfitting but that cause weaker results on training set as compare testing dataset.

2.0.2 ChexNet:

In the year 2018, Jaipurkar et al. (2018) used DensNet based model for classification of Chest X-ray dataset with 13 more diseases including pneumonia, the main purpose of this model is to reduce loss and reuse of features. One of the key characteristics of DensNet is it creates a shorter connection from the first layer and that helps to reuse the features. The creating of bounding boxes on the sample dataset helps to achieve more accuracy, F-Score, recall and precision. Chest X-ray data was having class imbalance to reduce class imbalance data augmentation is used like Stephen et al. (2019) and Sirazitdinov et al. (2019), also initial layers trained on ImageNet dataset to identify curves and boundaries. All these techniques helped to achieve the accuracy of the validation dataset as compare to CheXnet network, proposed network got a high accuracy of 76% and 96% for pneumonia disease.

Using deep convolutional Neural network gives advantages for feature selection and data training process it helps to make task easy but, every network have its own feature selection techniques and different numbers of layers Researcher Islam et al. (2017) used six different neural that is Alexnet, VGG16, VGG 19, ResNet-50, ResNet-101, ResNet-152 to get the best result. and on the top of that also used ensemble method to compare the result with pre-trained models. This research helps to find out how each network is different from each other, throughout this research ensemble method able to achieve the highest accuracy with 94% but in terms of sensitivity VGG19 network overpowered other networks and in the context of specificity Alexnet got the highest accuracy of 93%. In another research Rahmat et al. (2019) used Faster Regional Convolutional Neural network (Fast R-CNN) for detection in which Reginal proposal Network predicts the object bound and score and also can be used as fully convolutional network, to handle loss three different loss function used which helps to handle total loss and classification loss, throughout this result researcher able identify pneumonia faster than general practitioner and medical student, with 62% accuracy.

With the use of CNN Abiyev and Ma'aitah (2018) introduced to another method of pneumonia detection using supervised and unsupervised learning, the researcher used Back Propagation neural network with supervised learning and Competitive neural network with unsupervised learning, also used CNN for comparative purpose. The result is discussed in terms of accuracy, error rate and training time between networks. Multilayer feedforward neural network used in BPNN as an error BackPropagation algorithm which consists of input hidden and output layer in the network which helps to repeat backward and forward computation for training pairs. In CpNN only the Input and Output layer are fully connected to each other. the main reason to use this algorithm because these networks show efficiency in computer vision and biological computation. The third network is used as CNN which consists of a convolutional layer, fully connected layers and sub-sampling layer for activation function ReLU and sigmoid used. After testing all network CNN outperformed with the highest accuracy rate of 92% followed by CpNN and BPNN with 89.57% and 80.04% respectively. Abiyev and Ma'aitah (2018) also used VGG 16 and VGG19 network and able to achieve 86% and 92% accuracy rate. For further study researcher suggested using VGG16 and VGG19 network on different large datasets to get the highest result.

Initially, CNN was helped in image classification with start to end approach but now researchers finding different ways of utilizing CNN for the same purpose. Heo et al. (2019) used 1000 images of pneumonia infected people and another 1000 images of nonpneumonia patients. like previous researcher Abiyev and Ma'aitah (2018) VGG19, InceptionV3, ResNet50 used for feature extraction also in addition age, height and gender used as a demographic variable. But in images there are other structures like spine and heart are included, to reduce this structure researcher generated mask which is included lung area only to generate this mask U-Net algorithm used, on the basis of this masked output remaining model got trained, in training researcher used different CNN algorithm on both image-based demographic-based dataset. In the testing phase, image-based model outperformed on a demographics-based model in terms of AOC, but combination of both provided more efficient results. In which a combination of images and all demographic variables outperformed different combinations with 0.9213 in the scale of AUC.

Till now we saw so many used CNN model approaches in which well-known algorithms are used to get the highest output but still some algorithms like Multilayer extreme learning machine (MLELM) and Online Sequential Extreme Learning Machine (OSELM) rarely used in the researches. Vijendran and Dubey (2019) used both algorithms and the ensemble of these algorithms are used for pneumonia classification. MLELM and OSELM both consist of an input layer hidden layer and output layer. Training accuracy for both models was 91.1% and 95% respectively. but an ensemble of MLELM and OSELM which was known as Multilayer OS-ELM outperformed them with 96% accuracy. Even on the testing dataset Multilayer OS-ELM outperformed with 91.7%, which helps to recognize the less known algorithms are also useable for image classification.

Researcher Liu et al. (2019) used a novel technique called Segmentation Based Fusion Network (SDFN) in this research two CNN based classification models were used as feature extractors and the Lung region generator used to identify the lung regions. for the CNN model Fine-tuned Densnet is used like Pan et al. (2019) and Heo et al. (2019), and in the last FC layer, the sigmoid activation function used to normalization. With this technique approximately 95.84% of lung region detected and 0.719 AUC Score obtain for pneumonia detection.

In image classification or detection parameters like image size and noise in image always plays a crucial task in result, Sharma et al. (2018) completed research to identify how the quality and size of the image effect on image classification result. The researcher uses a different size of images with different noise levels to justify the research. Mainly chest X-ray images are used in this study. The researcher took different sizes of images with image resolution from 2048x2048 to 500x500 and performed the Otsu thresholding method to identify the pneumonia clouds from pneumonia affected patients. Totally 40 images used for this process and studied that the images with high resolution having high chances to detect pneumonia as compare to low images with a low resolution like 500x500.

2.0.3 Mendeley Dataset:

As we see in previous researches researcher Islam et al. (2017), Liang and Zheng (2019) and Jaipurkar et al. (2018) used the DensNet, VGG16, Inception and ResNet50/101/152 models to classify pneumonia with state of the art techniques using multiple algorithms not only provide broad view about respective networks but also helps to choose the perfect model for solving a problem. Furthermore, Ayan and Ünver (2019) used VGG16 against the Xception algorithm, which not used throughout the lit review. While using the Xception model pre-trained ImageNet weights used for initial layers also researcher added 2 fully connected layers with two-way output layers with a SoftMax activation function. This model trained on more than 5000 images which included both normal and pneumonia images while testing on test dataset VGG16 model outperformed in accuracy and specificity against Xception with 0.87 and 0.91 scores but Xception able to achieve the highest sensitivity with 0.85 and recall was 0.94.

Like Sirazitdinov et al. (2019), Jaiswal et al. (2019) and Ko et al. (2019) in this study Livieris et al. (2019) uses a statistical method with the new ensemble self-labeled algorithm to detect pneumonia from X-ray images. With the use of selection and combination, an ensemble of classifier generated. The self-labeled classifier used by applying different algorithms and every prediction of algorithm are selected through voting methodology.in the training phase dataset is labeled with L and U notation. After that voting fusion techniques used for the final hypothesis and the output are measured by sensitivity, specificity, and F-Measure.

2.0.4 Other Dataset:

Researcher Taylor et al. (2018) used VGG 19, Xception and inception models to carry pneumonia detection quests Islam et al. (2017), with the dataset with 13292 images including Pneumonia and normal cases. In which the researcher split data into 70:15:15 ratio for training, testing and validation, after running all this model researcher able to achieve 0.79 sensitivity for the VGG19 model and the Inception model is measured on specificity with 0.97. with the same VGG16 Network Islam et al. (2017) able to achived 76% Recall and 84% Precision. during this research Taylor et al. (2018) used hyperparameter tuning to achieve a result.

As we see from above Section 2 there are different datasets available for Pneumonia classification study and so many approaches still used including Transfer learning, Deep Neural Network, Statistical Method. Based on the previous studies we can clearly say that on Mendeley dataset appeared twice in research, still, there are lots of new aspects that can explore with this dataset, hence for this study Mendeley dataset is used. In further report Methodology, Evaluation and Results are explained.

3 Methodology

From the above section 2, we can clearly say that it is hard to identify pneumonia from chest X-ray images for any layman, it takes lots of practice and proper domain knowledge to identify disease. There are several methodologies developed to perform the Data mining task, in which Crisp-DM, SEEMA, and KDD process are the most famous methodologies. The main reason to use this methodology is portability, in CRISP-DM sequence of the project is not single directional also it can be customized easily. also we don't know which factors from our dataset will help to classify pneumonia, hence we are using CRISP-DM methodology (Marbán et al.; 2009).

CRIPS-DM methodology stands for Cross Industry-standard process and data Mining[1]. This Methodology contains 6 steps which are Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Deployment. Below Fig. 1 shows the working structure of CRISP-DM Methodlogy.

[1]https://www.ibm.com/support/knowledgecenter/en/SS3RA7_15.0.0/com.ibm.spss.crispdm.help/crisp_overview.htm

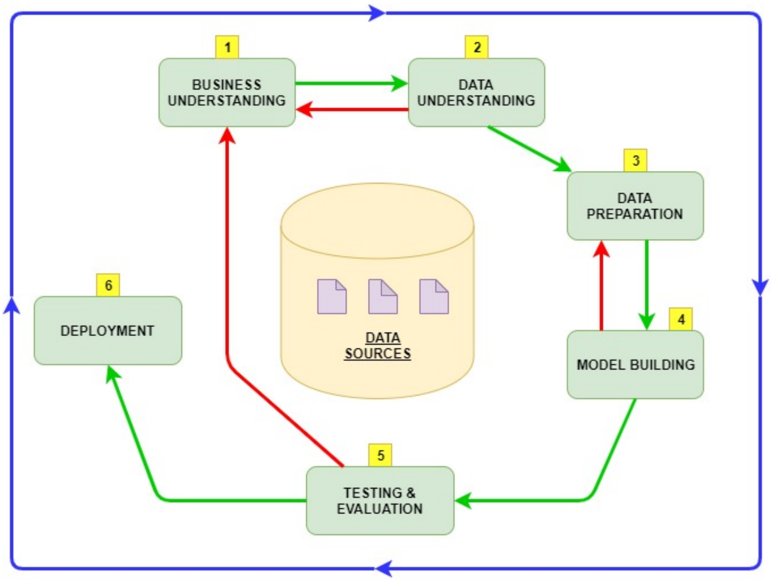

Figure 1: CRISP-DM Workflow

3.1 Business Understanding

Before working on any project, it is important to have a clear objective. The main objective of this study is to classify pneumonia from chest X-ray images with maximum accuracy. The researcher who tried to classify pneumonia from images got a competitively strong output. Their work has gained proficiency in their work but still there is chance for improvements. In this study, we trying to produce better results. This model will be major advantage for those people who are still not able to afford the expansive medical treatment and not able to visits the medical Practitioner for diseases.

3.2 Setup

Deep learning applications like natural language processing, search engines, recommender systems totally rely on the heavy computation on datasets, to perform this tasks parallel computing is considered as one of the best options, parallel computing traditionally totally rely on the Graphical Processing Units (GPUs) which leads to faster process and reliable solutions. It's always hard to maintain this kind of heavy system locally and performing heavy tasks on normal computers may lead to system or hardware failure, also wastage of time and human energy also this hardware is relatively very expensive than a normal computer. To overcome this drawback cloud-based solutions can be a option. As per the researcher Carneiro et al. (2018) A cloud-based platform like intel, azure, amazon is providing these services for fair pricing. Also, pre-configured CUDA based environments are provided by all cloud-based service provider. In this study we are going to perform classification of the images, which include processing image files, to perform this task requirement of heavy computation power is a must. To overcome this drawback Google Collaboratory (a.k.a. colab) is one of the reliable solutions used.

3.3 Data Understanding

To continue this study, we need to perform image classification and all data is image-based. The most common issue faced in images-based data is an imbalance, and it will lead to not utilizing a neural network with full potential. In the medical domain, datasets always have data imbalance as compare to another domain. For this study, we collected data from Mendeley[2] like Researcher Ayan and Ünver (2019) and Stephen et al. (2019). Below Fig.2 shows the Chest X-ray images that is normal and pnumonia affected, This dataset is freely available for research purposes. In June 2017 Mendeley awarded by Data Seal of Approval which indicates, this data is retrieved from trustworthy sources.

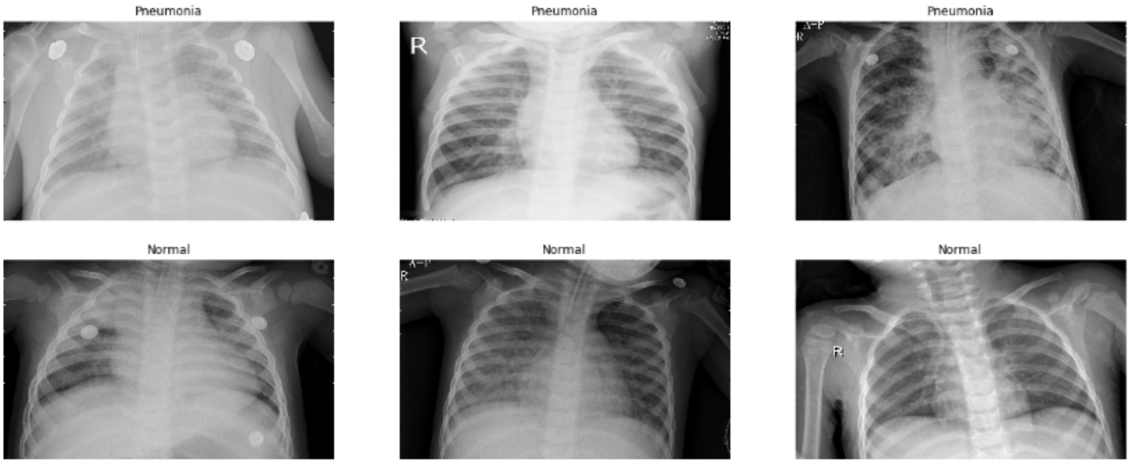

Figure 2: Normal & Pneumonia Images

The data is already divided in the group of training, testing and validation category. Each category having two folders called Normal and pneumonia which having images. from the below fig. 3 the training folder normal class having a total of 1323 images and pneumonia class having 3875 images, which clearly indicates that training data is highly imbalanced. In testing folder total 622 images are available in which 389 images are from pneumonia class and remaining 233 are from a normal class.

Figure 3: Training and Testing Dataset Count

[2]https://data.mendeley.com/datasets/rscbjbr9sj/3

3.4 Data Pre-processing

As mentioned earlier in Subsection 3.3, data is highly imbalance which can lead research to unwanted output, because the Machine Learning algorithm is developed to maximize the accuracy and reduce the error and in the case of data imbalance, we will not get proper predicted output. There are several ways to remove data imbalance in which oversample minority class, under-sample majority class and Synthetic Minority Oversampling Technique (SMOTE) are involved, but SMOTE is not very practical for high dimensional data [3] and under-sampling technique may lead to data loss which may reduce the feature collection for pneumonia images. To avoid this issue data augmentation technique has been performed for data pre-processing. As per Farhadi and Foruzan (2019) Data Augmentation technique can used to artificially increase the volume of the dataset. training Deep Neural network model on a large number of data can help to increase the accuracy of the result and the data augmentation technique can create the modified version of same data. To perform the image augmentation keras providing the library that performs augmentation while training model, which reduces the excess space to store new images, In below Figure 4 shows the example of Augemnted images genrated while trainning model, although offline augmentation helps you to create a new image and which can store on the system if required. We procced the task using ImageDataGenerator library which provides good range techniques, for this research specially used image rescaling, horizontal shift, image shift, and image flip techniques. the reason to use these techniques is, this method can use on most type of image Fujita et al. (2019) and other techniques like denoising, segmentation and marker labeling may reduce the number of features from images.

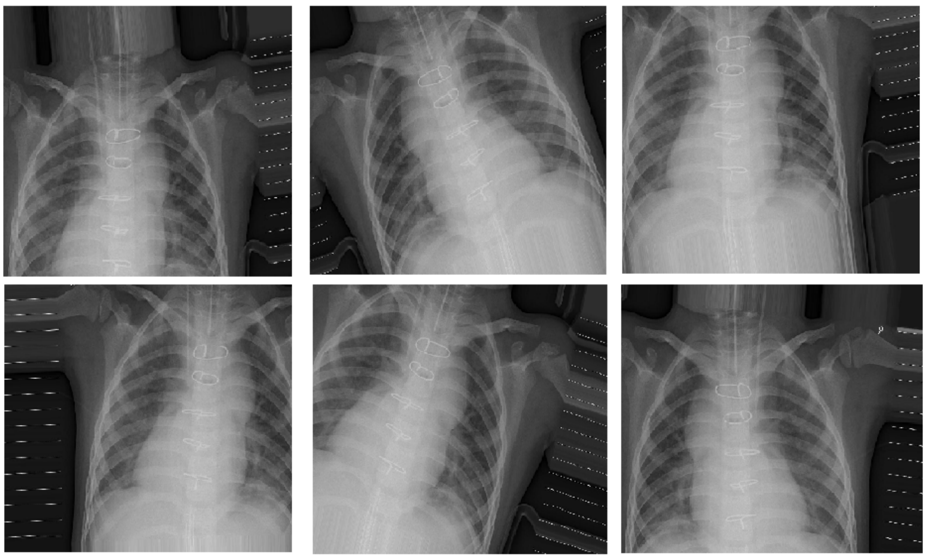

Figure 4: Augmented Images

In the ImageDataGenerator we set the values for techniques, in which for rotation_range we set to 20 degrees, for width_shift_range and Height_shift_range used 0.2 which indices the percentage (between 0 to 1), and set horizontal_flip true to flip image randomly. In the next step provided the input files using flow_from_directory which in-

[3]https://www.datacamp.com/community/tutorials/divingdeepimbalanceddata

cludes batch size target size and enabled shuffle mode to randomly shuffle images from a directory.

3.5 Modeling

The aim of this model to classify the difference between normal chest X-ray and pneumonia chest X-ray. To continue this study Transfer learning method is selected. Previous studies like Ko et al. (2019), Hwang et al. (2016), Taylor et al. (2018) and Islam et al. (2017) used the Convolutional Neural Network and got state of the art result. For this study using Islam et al. (2017) and Taylor et al. (2018), as a base paper in which the VGG19 model is used. VGG network basically introduced by Simonyan and Zisserman (2015) for the classification process, even though the VGG19 model having heavy parameters and long inference time, still this network characterized by simplicity. this model using a 3x3 Convolutional layer to increase the depth and with Rectifier Linear Unit (ReLU) activation function and volume size are reduced by max-pooling layer and SoftMax classifier used, the main reason to use this network is portability and this network rarely used on Mendeley dataset Ayan and Ünver (2019). Basically, CNN is a version of a multilayer perceptron but in regularized form, in this network each layer is connected to other layers with neuron using a fully connected layer [4]. Each CNN model consists of the convolutional layer + pooling layer and followed by fully connected layer. In CNN, Convolutional layer fold the one layer's input and pass as output to the next layer, where each pixel is treated as a relevant variable. Basically, it's a mathematical operation that combine the information of two sets. In our study, a convolutional filter creates a feature map of the image. That means it stacks all the features together and produces an output of layer. Another layer involved in CNN is the Pooling layer, the Main task of the pooling layer is to reduce dimension. Every time it reduces the parameter numbers which helps to reduce the training time and encounter overfitting. The most common pooling is max pooling which takes maximum value from the pooling window. Lastly, a fully connected layer is used to connect every neuron from one layer to another and to classify the image, flattened matrix passes through a fully connected layer layer[5]. While creating a model for our research dropout function is also used to prevent overfitting.

A model building or selecting model is one of the crucial tasks in CNN based classification method, selecting the proper model not only enhance the result of the classification but also reduces extra efforts. In this study, we used transfer learning to build the model. As per Researcher Shin et al. (2016) and Ling Shao et al. (2014) In the Transfer learning method pre-trained model weight is used to solve other problems by training them. In this research, the Mobile net model is used as base transfer learning. MobileNet is a small, low powered model that is specially designed for visual classification on top of that MobileV2 is faster than the previous version MobileNetV1 with the ability to produce the same accuracy. MobileNetV2 contains fully convolutional layer with 32 filters followed by 19 bottleneck layer that helps to improve the performance of the model. Initially VGG19 model is imported for model building purpose but for novel approach initial top 3 layers were excluded and added 2 more dense layers. VGG stands for Visual Geometry Group from Oxford University and 19 are the number of layers in CNN model. usually VGG16 uses the 224x224x3 dimensions for input images but we changed that with 128x128x3

[4]https://medium.com/@ksusorokina/imageclassificationwithconvolutionalneuralnetworks-496815db12a8

[5]https://cvtricks.com/tensorflowtutorial/trainingconvolutionalneuralnetworkforimageclassification/

dimensions, also ImageNet weights are used for pre-training weights and lastly disabled the training option so no new weights will be updated while model getting trained. While adding extra Dense layer ReLU and SoftMax activation function used, the ReLU function doesn't create back propagation errors like sigmoid and it is also nonlinear. And the reason to use SoftMax function is that SoftMax generates output in the range of 0 to 1(less than 0.5 considered as 0 and greater than 0.5 is considered as 1). Also, to reduce the overfitting effect dropout rate is set. with the help of this setting, we build our model to train for pneumonia classification. The Neural network like Xception, VGG16 and Custom Neural network is already used on the Mendeley dataset. Thus VGG19 selected to continue this research.

4 Design

In the Designing flow of the process is explained. Below Figure 5 clearify the steps involved in this study. In which initially dataset is downloaded from Mendeley to carry out this research, afterwords Data augmentation technique is used for Data preprocessing which helps to reduce data imbalance from data. The new model is developed with base VGG19 architecture in which Extra Dense layer and ImageNet weights added also initial 3 Layers are removed. After creating the model using training data training started and save this model for reusability. to calculate the measurement testing data used on the previously saved model and In the term of recall and precision used to check the measurement.

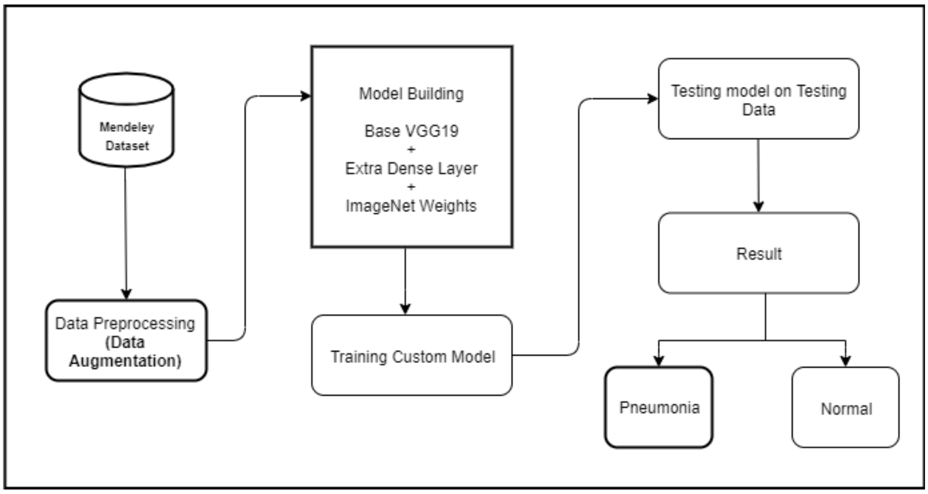

Figure 5: Process Flow

5 Implementation

In the previous section 4, we discussed the process flow of this study, but total detail implementation will discuss in this section. as mentioned earlier in subsection 3.2, we need high computational power with the python based working environment and to fulfill these requirements google collaboratory is used. Google colab is a cloud-based platform that provides CUDA based platform, which is perfect for the heavy task. Google provides a runtime python environment that is fully configured with all the latest Artificial Intelligence (AI) libraries and robust GPU for heavy computations. Google provides a Jupyter notebook which supports both Python 2 and python 3 environment without any prior setup. Google colab is configured with Intel(R) Xeon(R) CPU @ 2.20GHz with 13 GB of Memory and NVIDIA Tesla (K80) 2496 CUDA cores @560Mhz, 12GB GDDR5 VRAM with 353 GB Storage space and all this hardware configuration is provided without any cost. To complete this research further google colab is used and imaged based dataset is uploaded on google drive for maximum accessibility.

As mentioned earlier subsection 3.3 Dataset is already separated in training, testing and validation form, hence no need to used other techniques like K-fold cross-validation. But dataset having an imbalance with a 3 to 1 ratio, which can lead to unwanted results hence in data reprocessing, the Data augmentation technique is used. To complete data augmentation TensorFlow's *ImageDataGenerator* function is used in which horizontal flip, width shift, and height shift are used[6]. Also, *Flow_From_Directory* is another function from keras is used to set the batch size of the data and the dimension of the images is set to 128x128, with the same function we set to shuffle data on to shuffle data for each epoch.

While building a model transfer learning technique is used, in which the base VGG19 model used as base architecture also initial 3 layers of the VGG19 are removed and *model.trainable* function set to False so the model will not add extra weight while training. Also while creating model 3 more dense layers added and activation function ReLU and Softmax are used respectively. Also, to reduce the overfitting effect dropout rate is set to 0.7.

to Compile the model a *model.compile* function is used, while compiling model optimizer, loss function and metric need to define[7]. In this study, Adam optimizer algorithm is used because Adam optimizer uses less memory and for the different parameter it computes different learning rates. Secondly, loss function used for parameter estimation, for this study *categorical_crossentropy* is used to detect the single pneumonia class and for the metrics, Accuracy is used.

From the figure 6 model layers can be seen. Till now, the model is built for training purposes. now, we will train the model for classification, To initiate the training keras providing *fit_generator* function which helps to train the training data for N number of iteration N will be any number or how many times model should be trained.

Figure 7 helps to illustrate model training logs, Furthermore, we set epochs for 20. For every iteration, the model creates new data and discards the previous data. Also to checks that how a model is performing after each iterations. validation dataset is provided to check loss and accuracy against the validation set and all training logs are stored for evaluation purposes.

Using *evaluate_generator* function we check how much accuracy will get on the testing dataset. *save* function from keras is utilized to save the model for future use. To find out the recall of the model on unintroduced data. initially, stored the images from the training dataset to their every class i.e. 0 is for normal and 1 is for pneumonia, so we get to know which image is infected with pneumonia and which not. Furthermore using *predict* Function from Keras we provided all test data to model so the model can perform classification. At the end of this we can compare the result with the Confusion Matrix.

[6]https://www.tensorflow.org/tutorials
[7]https://keras.io/preprocessing/image/

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_1 (InputLayer)         [(None, 128, 128, 3)]     0
_________________________________________________________________
block1_conv1 (Conv2D)        (None, 128, 128, 64)      1792
_________________________________________________________________
block1_conv2 (Conv2D)        (None, 128, 128, 64)      36928
_________________________________________________________________
block1_pool (MaxPooling2D)   (None, 64, 64, 64)        0
_________________________________________________________________
block2_conv1 (Conv2D)        (None, 64, 64, 128)       73856
_________________________________________________________________
block2_conv2 (Conv2D)        (None, 64, 64, 128)       147584
_________________________________________________________________
block2_pool (MaxPooling2D)   (None, 32, 32, 128)       0
_________________________________________________________________
block3_conv1 (Conv2D)        (None, 32, 32, 256)       295168
_________________________________________________________________
block3_conv2 (Conv2D)        (None, 32, 32, 256)       590080
_________________________________________________________________
block3_conv3 (Conv2D)        (None, 32, 32, 256)       590080
_________________________________________________________________
block3_conv4 (Conv2D)        (None, 32, 32, 256)       590080
_________________________________________________________________
block3_pool (MaxPooling2D)   (None, 16, 16, 256)       0
_________________________________________________________________
block4_conv1 (Conv2D)        (None, 16, 16, 512)       1180160
_________________________________________________________________
block4_conv2 (Conv2D)        (None, 16, 16, 512)       2359808
_________________________________________________________________
block4_conv3 (Conv2D)        (None, 16, 16, 512)       2359808
_________________________________________________________________
block4_conv4 (Conv2D)        (None, 16, 16, 512)       2359808
_________________________________________________________________
block4_pool (MaxPooling2D)   (None, 8, 8, 512)         0
_________________________________________________________________
block5_conv1 (Conv2D)        (None, 8, 8, 512)         2359808
_________________________________________________________________
block5_conv2 (Conv2D)        (None, 8, 8, 512)         2359808
_________________________________________________________________
block5_conv3 (Conv2D)        (None, 8, 8, 512)         2359808
_________________________________________________________________
block5_conv4 (Conv2D)        (None, 8, 8, 512)         2359808
_________________________________________________________________
block5_pool (MaxPooling2D)   (None, 4, 4, 512)         0
_________________________________________________________________
global_average_pooling2d (Gl (None, 512)               0
_________________________________________________________________
dense (Dense)                (None, 512)               262656
_________________________________________________________________
dropout (Dropout)            (None, 512)               0
_________________________________________________________________
dense_1 (Dense)              (None, 2)                 1026
=================================================================
Total params: 20,288,066
Trainable params: 263,682
Non-trainable params: 20,024,384
```

Figure 6: Model Summary

```
Epoch 1/20
325/326 [============================>.] - ETA: 7s - loss: 0.3524 - acc: 0.8358 Epoch 1/20
326/326 [==============================] - 2526s 8s/step - loss: 0.3527 - acc: 0.8359 - val_los
s: 0.5513 - val_acc: 0.7500
Epoch 2/20
325/326 [============================>.] - ETA: 2s - loss: 0.2655 - acc: 0.8906Epoch 1/20
326/326 [==============================] - 762s 2s/step - loss: 0.2649 - acc: 0.8909 - val_loss:
0.5838 - val_acc: 0.8125
Epoch 3/20
325/326 [============================>.] - ETA: 2s - loss: 0.2445 - acc: 0.8985Epoch 1/20
326/326 [==============================] - 769s 2s/step - loss: 0.2448 - acc: 0.8982 - val_loss:
0.4172 - val_acc: 0.7500
•
•
•
Epoch 19/20
325/326 [============================>.] - ETA: 2s - loss: 0.1771 - acc: 0.9273Epoch 1/20
326/326 [==============================] - 776s 2s/step - loss: 0.1767 - acc: 0.9275 - val_loss:
0.6501 - val_acc: 0.8125
Epoch 20/20
325/326 [============================>.] - ETA: 2s - loss: 0.1723 - acc: 0.9327Epoch 1/20
326/326 [==============================] - 774s 2s/step - loss: 0.1725 - acc: 0.9325 - val_loss:
0.7670 - val_acc: 0.8125
```

Figure 7: Model Trainning Log

In the next section 6 will explain the evaluation technique for the model trained previously.

6 Evaluation

After the model gets trained, we need to check how the model is performed while training, in that case, Evaluation comes in a handy way. As per Hsueh et al. (2019) Maintaining logs of the training cycle helps a lot to understand model performance. In this study, we maintained the log of each epoch that is ran during model training. For evaluation purpose we will use that logs as Accuracy and loss plot, these logs will help to understand the growth of the model while training phase[8].

6.0.1 Accuracy plot:

An accuracy plot is one of the keras tools that help to identify how the model is performing while training. The below fig. 8 shows the accuracy plot our study, in which we can see that accuracy is for training data is started from 0.830 and it increasing till 5th epoch after little fluctuation accuracy plot again start to increase. During the training period accuracy plot for the training dataset is constantly fluctuating but also increasing. For validation data accuracy plot is showing constant development but during the training the model accuracy is decreased but again it comes to constant level. Even though accuracy is showing different patterns for both datasets the difference between both accuracy is least and they do not constantly increase together so we can say that data is not overfitting while training model.

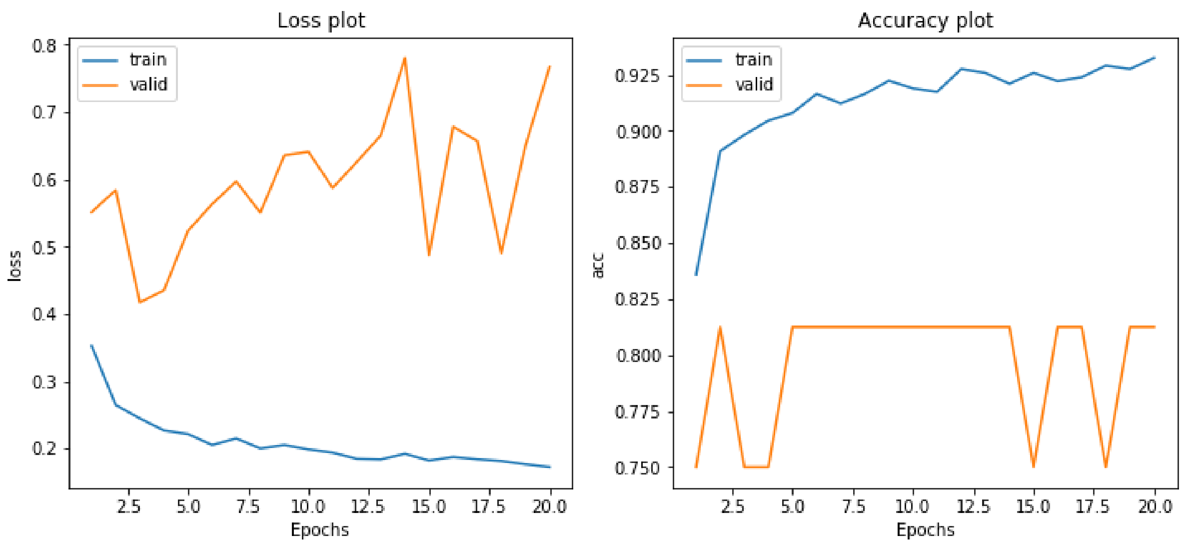

Figure 8: Loss and Accuracy Plot

6.0.2 Loss Plot:

Loss Plot tells the story about how much data is lost by the model while training period. Above Fig. 8 show how loss was continuously decreasing from beginning till last Epoch

[8]https://keras.io/visualization/

for training data. At initially data loss was in the range of 0.30 to 0.40 but after each iteration loss is decreased till less than 0.20. and for validation data loss is fluctuating thorough the whole training cycle. From the Above Fig. 8 image we can clearly say that loss for training and validation suggests that training data is not overfitting.

6.0.3 Evaluate Generator

Until now Accuracy plot and loss plot help us to find that how the model is performing while training phase on training data, but to check the accuracy on testing data keras providing evaluate_genrator function. This function helps to find out how the model will work on testing data, throughout this step our model got almost 0.86 accuracy on testing data. But evaluate generator predict the performance of testing data and for measure accuracy is considered. As this study is performing on medical dataset accuracy is not the finest measure that we can consider for measurement because accuracy is based on overall performance, hence recall and precision are the measure we are considering to find out how accurate is model predicting. To calculate the recall, model need to get unintroduced data that are not known to model yet. In that case, the test dataset used for testing purposes. Initially, all images from the test dataset are converted into 128x128x3 dimensions and labeled each image with their respective class, that is 0 for normal and 1 for pneumonia infected class, which will help to compare the prediction result with original labels. With help of Predict function provided by keras the model can test on testing dataset and store that result and to compare the predicted label with the original label.

7 Result

As mentioned earlier measuring criteria for healthcare domain is used as Recall and precision for more accurate classification (Deng et al.; 2016). The overall performance of the Custom VGG19 model performed on 624 images of pneumonia and normal class. It can be seen from the figure 9 out of 469 pneumonia cases, 383 cases are able to classify accurately, and 148 normal cases out of 155 classified as normal.

The mathematical representation of recall and Precision will give the proper measurement for the model[9].

basically,

True Positive (TP), is the case where a person has pneumonia and even model predicted Yes.
True Negative (TN), is the case where a person doesn't have pneumonia and model predicted No.
False Positive (FP), is the case where a person doesn't have pneumonia and model predicted Yes.
False Negative (FN), is the case where a person has pneumonia and model predicted as No.

[9]https://en.wikipedia.org/wiki/Confusion_matrix

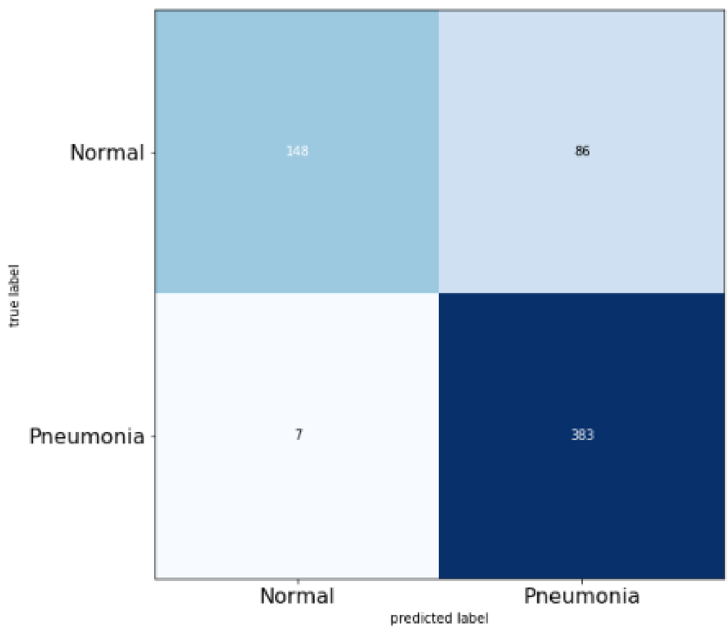

Figure 9: Loss and Accuracy Plot

Based on discussed Scenarios above below are the mathematical representation of Precision, Recall, Accuracy and F1 score.

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$Accuracy = \frac{TN + TP}{TP + TN + FP + FN} \tag{3}$$

$$F1score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{4}$$

where,

TP= True Positive, FP = False Positive, TN = True Negative, FN= False Negative

From the above formulas Recall is calculated as 0.98 and precision is calculated as 0.83, 0.851 is accuracy and 0.89 F1 Score calculated. In this study we able to get High recall as 98% but precision is measured 83%, this means whenever we get high Recall model compromise with Precision (Sokolova et al.; 2006).

A comparison of Previous work with the same VGG19 model is explained in Table 1, indicates that Recall for both researchers is not able to exceed more than 80% but in terms of precision Taylor et al. (2018) achieved more precision with 91% as compared Islam et al. (2017). The newly proposed method helps to find new result with increasing Recall percentage as compared Taylor et al. (2018) and Islam et al. (2017). New method able to achieve 98% recall, but it compromises Precision, during this research precision is reduced to 83% which lower than both previous studies, Thus it's possible to identify that

Researcher	Model	Recall	Precision
Taylor et al. (2018)	VGG-19	79%	91%
Islam et al. (2017)	VGG-19	76%	84%
Proposed Method	Custom VGG-19	98%	83%

Table 1: Comparison table between Previous studies and Proposed Studies

there is a trade-off between these two measures which need to maintain as per depending on the industry requirement.

8 Conclusion and Future Work

In the section 2 as we saw so many researchers contributed their work in pneumonia classification with different techniques on different datasets, in which RSNA and ChexNet dataset is popular for pneumonia classification due to a vast quantity of data availability. Many researchers used Transfer learning, Deep Neural Network, Convolutional Neural Network and also Statistical method to carry out research. In this study, a novel approach has been developed to classify the pneumonia disease. The dataset that has been used in this study was image-based and also rarely used for research purposes. One of the major challenges faced in this dataset is the high-class imbalance. Thus, the data augmentation technique has been used to overcome this problem. using the help of a Pre-trained network and VGG19 Architecture, we used the transfer learning technique with Imagenet weights. A novel approach of adding extra dense layers and excluding the first three layers has been undertaken in our study. Based on the results, it has been observed that our proposed model achieved better classification performance than Islam et al. (2017) and Taylor et al. (2018)

In the future, we would like to extend this study on the same dataset and try to get more precision with maintaining recall percentages, so end-user get more precise results.

9 Acknowledgement

I would like to convey my thanks to my supervisor Mr. Noel Cosgrave for his supervision and guidance, his suggestions to read more papers helps a lot to find new possibilities throughout this study. I would also like to thanks my family and friends for all their support and encouragement, also grateful to Mendeley for sharing a good and rich dataset.

References

Abiyev, R. H. and Ma'aitah, M. K. S. (2018). Deep Convolutional Neural Networks for Chest Diseases Detection, *Journal of Healthcare Engineering* **2018**.

Ayan, E. and Ünver, H. M. (2019). Diagnosis of pneumonia from chest X-ray images using deep learning, *2019 Scientific Meeting on Electrical-Electronics and Biomedical Engineering and Computer Science, EBBT 2019* pp. 0–4.

Carneiro, T., Da Nobrega, R. V. M., Nepomuceno, T., Bian, G. B., De Albuquerque, V. H. C. and Filho, P. P. R. (2018). Performance Analysis of Google Colaboratory as a Tool for Accelerating Deep Learning Applications, *IEEE Access* **6**: 61677–61685.

Deng, X., Liu, Q., Deng, Y. and Mahadevan, S. (2016). An improved method to construct basic probability assignment based on the confusion matrix for classification problem, *Information Sciences* **340-341**: 250–261.
URL: *http://dx.doi.org/10.1016/j.ins.2016.01.033*

Farhadi, M. and Foruzan, A. H. (2019). Data Augmentation of CT Images of Liver Tumors to Reconstruct Super-Resolution Slices based on a Multi-Frame Approach, *ICEE 2019 - 27th Iranian Conference on Electrical Engineering* pp. 1783–1786.

Fujita, K., Kobayashi, M. and Nagao, T. (2019). Data Augmentation using Evolutionary Image Processing, *2018 International Conference on Digital Image Computing: Techniques and Applications, DICTA 2018* pp. 1–6.

Heo, S. J., Kim, Y., Yun, S., Lim, S. S., Kim, J., Nam, C. M., Park, E. C., Jung, I. and Yoon, J. H. (2019). Deep learning algorithms with demographic information help to detect tuberculosis in chest radiographs in annual workers' health examination data, *International Journal of Environmental Research and Public Health* **16**(2).

Heron, M. (2018). National Vital Statistics Reports Volume 67, Number 5 July 26, 2018, *National Vital Statistics Reports* **67**(6): 1–15.

Hsueh, Y., Ittangihala, V. R., Wu, W. B., Chang, H. C. and Kuo, C. C. (2019). Condition monitor system for rotation machine by Cnn with recurrence plot, *Energies* **12**(17).

Hwang, S., Kim, H.-E., Jeong, J. and Kim, H.-J. (2016). A novel approach for tuberculosis screening based on deep convolutional neural networks, *Medical Imaging 2016: Computer-Aided Diagnosis* **9785**: 97852W.

Islam, M. T., Aowal, A., Minhaz, A. T., Ashraf, K. and Sep, C. V. (2017). Abnormality Detection and Localization in Chest X-Rays, *arXiv* pp. 1–16.

Islam, S. R., Maity, S. P., Ray, A. K. and Mandal, M. (2019). Automatic Detection of Pneumonia on Compressed Sensing Images using Deep Learning, *2019 IEEE Canadian Conference of Electrical and Computer Engineering, CCECE 2019* pp. 1–4.

Jaipurkar, S. S., Jie, W., Zeng, Z., Gee, T. S., Veeravalli, B. and Chua, M. (2018). Automated Classification Using End-to-End Deep Learning, *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS* **2018-July**: 706–709.

Jaiswal, A. K., Tiwari, P., Kumar, S., Gupta, D., Khanna, A. and Rodrigues, J. J. (2019). Identifying pneumonia in chest X-rays: A deep learning approach, *Measurement: Journal of the International Measurement Confederation* **145**: 511–518.
URL: *https://doi.org/10.1016/j.measurement.2019.05.076*

Ko, H., Ha, H., Cho, H., Seo, K. and Lee, J. (2019). Pneumonia Detection with Weighted Voting Ensemble of CNN Models, *2019 2nd International Conference on Artificial Intelligence and Big Data, ICAIBD 2019* pp. 306–310.

Liang, G. and Zheng, L. (2019). A transfer learning method with deep residual network for pediatric pneumonia diagnosis, *Computer Methods and Programs in Biomedicine* (xxxx).
URL: *https://doi.org/10.1016/j.cmpb.2019.06.023*

Ling Shao, Fan Zhu and Xuelong Li (2014). Transfer Learning for Visual Categorization: A Survey, *IEEE Transactions on Neural Networks and Learning Systems* **26**(5): 1019–1034.

Liu, H., Wang, L., Nan, Y., Jin, F., Wang, Q. and Pu, J. (2019). SDFN: Segmentation-based deep fusion network for thoracic disease classification in chest X-ray images, *Computerized Medical Imaging and Graphics* **75**: 66–73.
URL: *https://doi.org/10.1016/j.compmedimag.2019.05.005*

Livieris, I., Kanavos, A. and Pintelas, P. (2019). Detecting Lung Abnormalities From X-rays Using an Improved SSL Algorithm, *Electronic Notes in Theoretical Computer Science* **343**: 19–33.
URL: *https://doi.org/10.1016/j.entcs.2019.04.008*

Marbán, O., Segovia, J., Menasalvas, E. and Fernández-Baizán, C. (2009). Toward data mining engineering: A software engineering approach, *Information Systems* **34**(1): 87–107.

O'Quinn, W., Haddad, R. J. and Moore, D. L. (2019). Pneumonia Radiograph Diagnosis Utilizing Deep Learning Network, *Proceedings of 2019 IEEE 2nd International Conference on Electronic Information and Communication Technology, ICEICT 2019* pp. 763–767.

Pan, I., Agarwal, S. and Merck, D. (2019). Generalizable Inter-Institutional Classification of Abnormal Chest Radiographs Using Efficient Convolutional Neural Networks, *Journal of Digital Imaging* **32**(5): 888–896.

Rahmat, T., Ismail, A. and Aliman, S. (2019). Chest X-Ray Image Classification Using Faster R-Cnn, **4**(1): 225–236.

Sharma, A., Raju, D. and Ranjan, S. (2018). Detection of pneumonia clouds in chest X-ray using image processing approach, *2017 Nirma University International Conference on Engineering, NUiCONE 2017* **2018-January**: 1–4.

Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D. and Summers, R. M. (2016). Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning, *IEEE Transactions on Medical Imaging* **35**(5): 1285–1298.

Simonyan, K. and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition, *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings* pp. 1–14.

Sirazitdinov, I., Kholiavchenko, M., Mustafaev, T., Yixuan, Y., Kuleev, R. and Ibragimov, B. (2019). Deep neural network ensemble for pneumonia localization from a large-scale chest x-ray database, *Computers and Electrical Engineering* **78**: 388–399.
URL: *https://doi.org/10.1016/j.compeleceng.2019.08.004*

Sokolova, M., Japkowicz, N. and Szpakowicz, S. (2006). Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation, *AAAI Workshop - Technical Report* **WS-06-06**: 24–29.

Stephen, O., Sain, M., Maduh, U. J. and Jeong, D. U. (2019). An Efficient Deep Learning Approach to Pneumonia Classification in Healthcare, *Journal of Healthcare Engineering* **2019**.

Taylor, A. G., Mielke, C. and Mongan, J. (2018). Automated detection of moderate and large pneumothorax on frontal chest X-rays using deep convolutional neural networks: A retrospective study, *PLoS Medicine* **15**(11): 1–16.

Vijendran, S. and Dubey, R. (2019). Deep Online Sequential Extreme Learning Machines and its Application in Pneumonia Detection, *Proceedings of 2019 8th International Conference on Industrial Technology and Management, ICITM 2019* (3): 311–316.