COSC 2673/2793 - Assigment 4

Syed Haider Saleem (s3796258)

I. Introduction

COVID-19 is a respiratory illness caused by a new virus. Symptoms include fever, coughing, a sore throat and shortness of breath. The virus can spread from person to person. Currently there is no treatment for COVID-19 [1]. This is the reason why it is of such prime importance to detect it early on and handle it accordingly by implying proper social distancing and isolation from others. Currently the setup in place to detect virus is not as efficient as you would like to test millions of people in a day and this is more prominent in third world countries where sufficient equipment and recourses are not available to deal with this. It is believed that chest radiography does contain small portions that well trained deep neural networks can focus on and find points that are otherwise not notable to the human eye [2]. The purpose of this paper is to explore this hypothesis by going through existing work in machine learning using machine and deep learning models, their strengths, limitations and try to propose a solution that could hopefully help in furthering the efforts to tackle this pandemic.

II. Literature Review

We believe that chest x-rays contain small portions of patterns that could help in identifying COVID-19 in patients. Deep learning has time and time proved that it has the capacity to perform well on image classification problems among which Convolutional Neural Networks has the front seat, hence we believe that trying to understand and extend already developed deep convolutional neural networks will be the best way to move forward hence our proposed solution is based on this. Hamdan and his team [3] discusses how Deep Neural Networks (abbreviated to DCNN) are one of the most powerful deep learning models and they have been widely used in image classification and pattern recognition problems. They also state that DCNNs are previously been deployed to help successfully diagnose common chest disease such as Tuberculosis screening [4] and mediastinal lymph nodes in CT images [5]. Hamdan discusses different state-of-the-art deep learning models that have performed exceptionally well over the years and how they have used each one of them to accomplish the clinical purpose of the COVIDX-Net framework which they propose in their paper. These deep learning models includes VGG19, DenseNet201, ResNetV2, InceptionV3, InceptionResNetV2, Xception and MobileNetV2. All deep learning classifiers used stochastic gradient decent as their optimizer. Their data set includes 50 X-ray images divided into two classes with 25 belonging to normal cases and 25 positive COVID-19 images. 80% of the data (images) were used for training while 20% were used for testing purposes. VGG19 and DenseNet201 models performed best with an accuracy of 90% and an astonishing recall of 100% on COVID-19 positive case classification. recall in this case is particularly important as you don't want your model to miss a positive case (patient). Some key contribution of the paper includes building a new framework (COVIDX-Net) to automatically assist in early diagnosis of COVID-19 patients. Proposed framework support to help researchers further their study based on these results. The limitations of the paper are in the size of the data set, the author used only 50 images for the task among which 45 images were used for training and only 5 images were used for testing which clearly states that the results are not stable. Other issue is that this small data set was trained for 50 epochs which could cause the model to overfit on such few examples hence the model might not perform as expected in the real-world setting. A better built data with more examples needs to be taken, different COVID-19 data repositories needs to be checked and some data enhancing techniques could be tried as well like data augmentation to avoid the current limitations.

A paper by Ozturk [2], further builds on our initial claim as how important of a role can Artificial intelligence play if coupled with radiological imaging in automated accurate detection of COVID-19. Ozturk further disuses how deep neural networks could pave the way especially Convolutional Neural Networks. They propose a model DarkCovidNet that is constructed using already proven model DarkNet classifier which is taken as a base model. The proposed model (DarkCovidNet) has 17 convolutional layers and LeakyReLU was used as an activation function. An important issue that the author discusses is how pneumonia is a disease associated with COVID-19 and since the chest x-ray for pneumonia patient could look similar to COVID-19 patient this could confuse the model and hence a separate dataset for pneumonia was also collected. The paper also provides detailed domain specific information that helped understand the problem in a bit more detail. The dataset used to train the model was taken from two different sources having 127 COVID-19 x-ray images and 500 images from no-findings and pneumonia class. The author acknowledges the problem of limited COVID-19 images. The model was then trained in two different scenarios, firstly, DarkCovidNet was used to do multilabel classification and then it was used for binary classification using 5-fold-cross-validation for both. The experiment was repeated 5 times and the model was trained for 100 epochs each time. The model produced an average classification accuracy of 87.02% in multi class classification while the average was much high in binary classification, as high as 98.08%. The author then comments how the model was not only tested on the test set but it was reviewed by radiologists who had a very positive response on the model saying that the model performed exceptionally well in detecting COVID-19 cases for binary class tasks, it was successful in finding COVID-19 findings, model is sensitive in detecting pneumonia disease, heat map is useful in detecting COVID-19. The paper also establishes some common distinguishing facts between the COVID and pneumonia case images which would be really helpful for future work. Some limitations of the paper even though a few were that model made incorrect predictions in case of poor-quality images and on patients with acute respiratory distress.

Another limitation is the uneven class distribution as they had comparatively less COVID-19 case images as compared to normal and pneumonia case images which could create a bias and may not work as well in real life. Data Augmentation using zoom, rotate and shift could have been experimented with the data.

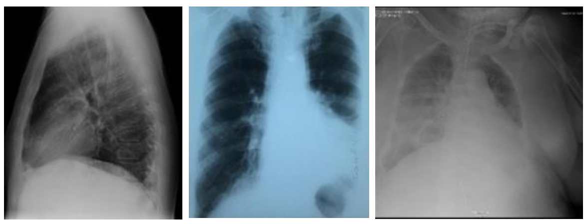
The paper by Apostolopoulos [6] also opens by stating the importance of deep learning applications over the last five years and how machine learning and deep learning have become established disciplines in applying artificial intelligence to recognize and learn patterns from images especially Convolutional Neural Networks. This paper was particularly interesting as it focused on the importance of transfer learn in case of Covid-19 problem due to low availability of Covid-19 case images for training model. Author discusses the effort put in to select a very particular and carefully picked data from multiple resources. The problem of model (CNN) getting confused between pneumonia and Covid-19 due to similarity in the images is also discussed and catered for by selecting pneumonia images in the dataset as well. The author then continues to explain how transfer learn helps the model in retaining the know ledged mined from a given larger dataset and then could be employed to process a new set of images of another nature and extract useful features out of it, two types of transfer learn are discussed where in the first type the original structure and weights are retained and the pre-trained model is used as a feature extractor while in the second type specific modifications are applied to pre-trained model. The author then discussed the five models shortlisted that were already trained on large dataset (transfer learnt) and proved to perform well in image classification and object detection. The evaluation metrics was defined, and two type of accuracy are discussed, first accuracy refers to the overall accuracy between the three classes known as 3-class accuracy and second was 2-class accuracy between Covid-19 and normal classes. Based on these metrics, accuracy, sensitivity and specificity was computed. VGG19 and the Mobile Net achieved best classification accuracy overall with Mobile Net outperforming VGG19 in terms of specificity hence the author describes Mobile Net as the most effective model for detection of Covid-19 from X-ray images. The author concludes the paper by discussing that the present work could contribute to the possibility of developing a low-cost automatic diagnosis of the disease and could help in reducing the manual work. Some major takeaways from the paper is the focus on transfer learn technique and how it could prove useful in such circumstances when we have such limited data but there are few limitations to this paper which are acknowledged by the author as the problem of not having enough patient data, even with transfer learn you need to have good enough problem specific data that would help in improving the accuracy of the model. Data augmentation was also not experimented with which could have helped in solving the problem of less data, even though its (data augmentation) importance is mentioned by the author in such scenario. Another limitation of the model mentioned in this paper like many others is its inability to distinguish between Covid-19 cases from similar viral cases, such as SARS and also from different variety of common pneumonia. Another important problem mentioned is only using a medical image for diagnosis rather than a more holistic approach in which other factors could also be covered. For future some of the major improvements that could be done includes augmenting and creating new images with more variations. More information could be incorporated along with the image to provide more background history for the model to make decision and another important aspect could be to train our model with more diseases that appears to be similar to Covid-19 in X-rays so that our model could be able to distinguish better between Covid-19 and others like for ex SARS or pneumonia.

III. PROPOSED METHODOLOGY

Literature review helped understand the Covid-19 image classification problem in a bit more detail which definitely helped in analysing the shortcomings with the current methods that we need to tackle in our approach. Firstly, it became evidently clear that the major issue that we are going to face is the very limited publicly available dataset which would cause a huge class imbalance if we try to train it with other class images (normal, pneumonia patients) hence we need to come up with solutions that might help in resolving this. Data Augmentation might help us in increasing the dataset, transfer learn will also play an important considering an already trained model on huge dataset would be better able to extract features with even small data. For transfer learn we would also need to be smart enough to pick models that are already trained on medical images specifically chest x-ray images as we already have Convolutional Neural Networks that are successful in diagnoses of common chest disease such as Tuberculosis screening [4] and mediastinal lymph nodes in CT images. Another major decision inspired by literature review is the decision of using Convolutional Neural Networks models as they have proved to perform exceptionally well in all papers. The proposed methodology will be discussed more in detail below.

A. Dataset

The dataset is collected by Dr. Joseph Cohen who keeps on updating the dataset with more publicly available data as through indirect collection from hospitals and physicians, it consists of about 200 X-ray images of Covid-19 patients and about 300 images of non-Covid patients (healthy once's as well as images of patients with other viral and bacterial pneumonias). These images can be observed in fig 3.1.



(a) (b) (c)

Fig 3.1 images collected from the available dataset. Fig (a) shows a side X-ray. Fig (b) shows a slightly different coloured X Ray. Fig (c) shows a front side poor quality X Ray

Some of the things we can straight away realise from viewing these images is that we have a varied dataset with images from different angles, different quality and in some cases even in different colours. Because of this we need to design our model keeping these dynamics in mind. We need to handle for these variations hence a thorough and specific data pre-processing is required to bring all of these images to a common scale before feeding to our model.

B. Data Pre-processing

Data pre-processing is the most crucial step in data science, and it is estimated to take about 80% of the overall data science project time. This clearly indicates the importance of this stage as the model is as good as the data it gets. After seeing our image data set some of the pre-processing steps require includes image normalization, Data Augmentation, resizing or cropping the image and converting all the images to grey scale to ensure consistency.

a) Image Normalization: one of the major tasks we need to do in data-pre-processing is to normalize the image to 0-1. If we do not normalize then the ranges of our distributions of feature values would likely be different for each feature, and thus the learning rate would cause corrections in each dimension that would differ (proportionally speaking) from one another. For that we have to divide each pixel value to the maximum value, which is 255, this helps in normalizing the pixel values to a range of 0-1.

b) Grey Scale Conversion: And also we need to convert our images to a single scale and grey scale would be a good idea considering it could significantly reduce the amount of input features and even if we think about it, we can't expect to get any information out of colour from an X-ray hence its best to convert all images to grey scale.

c) Resizing, Cropping and Selecting Images: It's probably best to crop and resize all images to fit to a constant size. From the literature review we leaned that model performed badly on poor quality images hence we need to make sure that the images we feed are at their best by cropping them to a perfect size and maybe manually removing those images that could cause our model to learn unwanted noise.

d) Data Augmentation: A very important data pre-processing step in case of limited data. One of the things we saw in other papers from literature review was that most authors did not experiment with data augmentation. We will try to augment and generate more data by using zooming, rotating, shifting and by varying the brightness of the image. This helps in generating a good amount of dataset which could help our model learn some variations and extra features that could be useful, but at the same time we also need to be careful to not overdo the data augmentation as it may cause our model to get confused by general unrealistic images. This also requires some manual handling as it won't make sense to apply some of data augmentation techniques to some category of images, for example you might not want to flip a traffic sign image as that would never exists in reality, so we have to be careful. Image data generators seems like a good option for us to do this.

e) Train Test split: It is necessary to split our data into train test split early on to make sure we have kept a good amount of data for our testing purposes. 80% of our data will be used for training purposes while the rest 20% will be used to validate/evaluate our models and make judgment based on that. Another important thing to keep in mind while creating a train test split is to keep in mind the class imbalance and make sure both train and test splits contain equal class representation.

C. Feature Selection

Feature selection is the process of selecting features from our dataset that would be most beneficial for our model in completing the specific task. We have all heard of curse of dimensionality as the data and its features grew it becomes more and more computationally expensive for our model to be trained. We generally have a lot of features, most of which are not adding any useful information for completing the task and it is just overcomplicating our problem hence we need to make sure to pick the right features in machine learning. But, fortunately deep learning models like CNN we do not need to worry about this.

D. Feature Extraction

Feature extraction is the process of generating new features from the already selected features using convolutional layers, max pooling and non-linear activation function. Max pooling helps in down sampling an input representing reducing dimensionality while non-linear activation function helps in extracting non-linear combination of features that could help in solving a non-linear problem.

E. Proposed Model(s)

As discussed before Deep Convolutional Neural Network would be our choice for this type of problem as CNN have proved to be better than traditional machine learning and deep learning models for image classification and pattern recognition.

a) Experimental Setup: Since we are using transfer learning, we will be considering some of the proved DCCNs as discussed in some other papers based on their performance. Transfer learning basically means to use pre-trained models on data that closely relates to the data on which the model was pre-trained. As mentioned, before we will try to use a model that is pre-trained on medical Imaging data specifically those models that are pre-trained on chest X-ray data. The reason why transfer learn becomes so important in our case is due to our small data size while deep neural networks required a lot of data to learn the pattern. what transfer learn does is you don't need to feed such large amount of data as it has already extracted features using the data feed before. Some of the pre-trained models that we are will consider are VGG19, DenseNet201, Mobile Net and DarkNet classifier. These are all those models that were among the top performers on the Covid-19 dataset. For each of these classifiers we will tune adjust our input layer according to the number of pixels in our images (as in any convolutional network layer) and pass on the input values. Each model will then be trained on those images for 50 epochs, this is done to ensure that our model does not overfit as the data is small, also a training and validation graph would be used to find the optimal point for epochs. After training each of these four pre-trained models with our dataset we will then compare each of these based on common evaluation criteria that we have defined and the model performing best on our set criteria will be selected for the task.

Comparison of our four selected models can be observed in table 3.1

Model #	Set of Models		
	Model Name	***Number of conv Layers***	***Pre-Trained Models***
1	VGG19	16	Yes
2	DenseNet201	201	Yes
3	Mobile Net	28	Yes
4	DarkNet	19	Yes

Table 3.1 Shows the comparison between our four models

b) Evaluation: For evaluation we will be using 20% of the data that we kept aside initially during train and test split. Evaluation metrics that we choose for this problem is a combination of few metrices that includes precision, recall and F1 score. Each of these metrics has their own importance in our judgement as they each gives a different type of information which could be essential in our judgment. We will however try to pick a model that has high recall (as we don't want to miss any positive case) and F1 score. F1 score is particularly important here as it is harmonic mean of both precision and recall and what it does is calculate result keeping in mind the class distribution, as we mentioned in the start that our problem is a class imbalance problem hence F1 score will be our primary evaluation metrics. Also, for now we are treating our problem as a binary classification problem meaning we would only be classifying the data as Covid-19 image or healthy person image (it is worth mentioning that healthy person in this case could have other virus and diseases).

c) Hyperparameter Tunning: One of the key things in machine learning and deep learning is to tune our parameters according to our data, these are unknown's that we need to tune for our model to give optimal performance. Some of the important hyperparameters that we need to keep in mind for CNNs are learning rate, batch size, momentum, decay and type of optimizer (i.e. stochastic gradient decent or gradient decent). This hyperparameters tuning process helps us find a balance between underfitting and overfitting. Grid search or Random search can be applied to our models to find the optimal hyperparameter's.

d) Independent Evaluation: Independent evaluation means to use data from outside the training and validation world. This is real life data which in our case, I believe will play an important role in telling us whether our model can perform practically well in real life situation or not.

e) Technical Difficulties: Some of the obvious technical difficulties we can presume straight away is loading and training such huge models. We would require a lot of space and computational power to perform these tasks. From the table 3.1 we can clearly see that our biggest model has up to 201 layers that would for sure take a huge effort to be trained and tested. Other models are comparatively smaller that is they have less convolutional layers. In case we are not able to train DenseNet201 or if it is taking excessive time, we might have to drop it and replace it with another less dense convolutional neural networks. Some other issues we might encounter handling this problem is the domain specific knowledge as we won't be able to work with 100% efficiency if we can't understand the domain of the project properly. Also, for final judgment we need to have expert opinion which will be difficult to get and most importantly a tough challenge would be to collect data for our independent evaluation.

f) Considerations and Biases: Since we have an extremely small dataset it is very much possible that we overfit our model causing it to not perform according to our expectations on real life dataset. As we are augmenting the same images it could be that the model starts to learn same features and noise and it would become impractical. Dataset labeling and authenticity could also not be as creditable since it's a very new dataset and is constantly been updated.

g) Ethical Declaration: The most important thing to keep in mind is that the data we are using is sensitive as it is a patient's data and HIPPA regulates strict guidelines in using this. Also, we need to be aware of the fact that our model is still very premature and is expected to perform unpredictably in real-life scenario hence this should be kept in mind and the decisions made by system should be reviewed by professionals for confirmation.

REFERENCES

[1] Australian Department of Health. (2019), *What you need to know about coronavirus(COVID-19). Available at:* https://www.health.gov.au/news/health-alerts/novel-coronavirus-2019-ncov-health-alert/what-you-need-to-know-about-coronavirus-covid-19 (Accessed at 08 June 2020).

[2] Tulin Ozturk , K., et al. " Automated detection of COVID-19 cases using deep neural networks with X-ray images" Disease Modelling 5 (2020): 256-263. Comput Biol Med 2020

[3] Ezz El-Din Hemdan, Marwa A. Shouman, and Mohamed Esmail Karar, "COVIDX-Net: A Framework of Deep Learning Classifiers to Diagnose COVID-19 in X-Ray Images." arXiv:2003.11055v1

[4] F. Pasa, V. Golkov, F. Pfeiffer, D. Cremers, and D. Pfeiffer, "Efficient Deep Network Architectures for Fast Chest X-Ray Tuberculosis Screening and Visualization," Scientific Reports, vol. 9, no. 1, p. 6268, 2019/04/18 2019.

[5] Y. Miki et al., "Classification of teeth in cone-beam CT using deep convolutional neural network," Computers in Biology and Medicine, vol. 80, pp. 24-29, 2017/01/01/ 2017.

[6] Apostolopoulos, Ioannis D., and Tzani A. Mpesiana. "Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks." *Physical and* Engineering Sciences in Medicine (2020): 1.