Ollscoil Teicneolaíochta an Atlantaigh

Atlantic Technological University

COVID - 19 Automated Detection using Convolutional Neural Networks and Generative Adversarial Networks

A thesis submitted
by

Ultan Kearns

in partial fulfillment of the requirements for the degree of
Master of Science in Computing in Big Data Analytics and Artificial Intelligence

Supervisor: Dr Paul Greaney

Submitted to Quality and Qualifications Ireland (QQI)
Dearbhú Cáilíochta agus Cáilíochtaí Éireann
October 2022

Contents

Declaration

I hereby certify that this material, which l now submit for assessment on the programme of study leading to the award of Master of Science in Computing in . . . , is entirely my own work and has not been taken from the work of others except and to the extent that such work has been cited and acknowledged within the text of my own work. No portion of the work contained in this thesis has been submitted in support of an application for another degree or qualification to this or any other institution. I understand that it is my responsibility to ensure that I have adhered to LYIT's rules and regulations.

I hereby certify that the material on which I have relied on for the purpose of my assessment is not deemed as personal data under the GDPR Regulations. Personal data is any data from living people that can be identified. Any personal data used for the purpose of my assessment has been psuedonymised and the data set and identifiers are not held by LYIT. Alternatively, personal data has been anonymised in line with the Data Protection Commissioners Guidelines on Anonymisation.

I give consent for my work to be held for the purposes of education assistance to future Computing students at LYIT and it will not be shared outside the Department of Computing at LYIT. I understand that my assessment may be shared with any other third party and will be held securely in LYIT in line with the Institute's Records Retention Policy.

Signed: Ultan Kearns

Date: Wednesday 26^{th} October, 2022

Acknowledgements

I would first of all like to thank my supervisor during this project Dr. Paul Greaney, he was a fantastic help throughout the course of writing this thesis and this work could not have been completed without his input and help.

I would also like to thank the lecturers who taught me so much during my postgraduate course, both Doctors Karla Muñoz-Esqueival and Shagufta Henna provided a fantastic introduction into many areas of Artificial Intelligence and the knowledge they imparted has helped me a lot throughout the course of conducting this research.

I would also like to thank Andrew Ng, his deep learning courses provided a great foundation into the realm of deep learning and Artificial Intelligence.

Acronyms

- AI - Artificial Intelligence
- ANN - Artificial Neural Network
- CNN - Convolutional Neural Network
- GAN - Generative Adversarial Network
- CT - Computed Topography

List of Figures

Abstract

This paper aims to analyze the applications of Generative Adversarial Networks in overcoming issues of data-shortages in relation to COVID-19. There are many COVID-19 data-sets compiled but some suffer from lack of data-quality and data shortages[1][2]. In this paper I aim to create and train a convolutional neural network or CNNs to analyze X-Rays of patients lungs to automate the detection of COVID-19. The CNN will be trained with a number of images generated from different GAN architectures to determine which will prove most efficient in automating the detection of COVID-19. I also aim to use the GANs in conjunction with one and other to try out different combinations to see if feeding images generated by one GAN to other GANs will produce more accurate results when training the model.

Chapter 1

Introduction

1.1 Generative Adversarial Network (GAN)

A generative adversarial network or GAN for short first appeared in a 2014 paper by Ian Goodfellow et al[3]. In this paper Goodfellow et al propose a new way to generate data via an adversarial process. The GAN essentially works as follows: two models are trained, a generative model G which will generate the content from the data and another model D which will be the discriminator, judging if data created by the model came from the dataset rather than G. The goal of this training is to ensure data generated from G is realistic enough to fool the discriminator D into believing that the generated content came from the training set. It is in this way that we can create realistic "fake" data from the generative model.

There are a number of GAN architectures which are useful in different scenarios, such as CycleGans[4] which are useful for translating images from a source domain $X \rightarrow Y$ in which Y is the target domain, StyleGan which was created by NVIDIA which allows more control over the generative process[5] and PixelRNN which can recreate images when given a fraction of the original and can generate new images based on probability[6].

In this dissertation I will examine a number of different generative adversarial network architectures and will use them in conjunction with each other, by feeding content generated by one architecture into another to develop a more diverse training set for the final model.

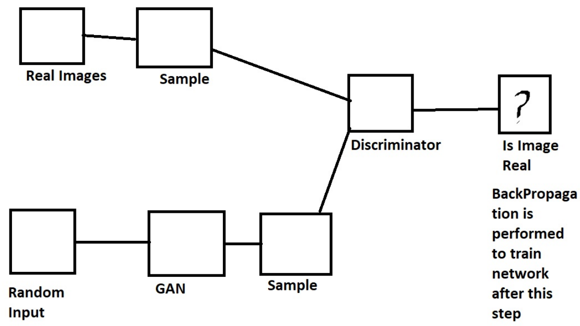

Figure 1.1: Basic Example of generative adversarial network

As we can see from the image above, we start the process by taking a sample of real images from the training data, then passing it to the discriminator. We also take a sample from the GAN created images and pass that to the discriminator which will then determine if the images are real are fake. After the discriminator determines if the image is real or fake then backpropagation is performed to train the model so that it can differentiate better between samples which came from the training set and those which came from G.

1.2 What is An Artificial Neural Network? (ANN)

An artificial neural network, or ANN for short, is a network of neurons or nodes which are used for training a model to perform a certain task. They are made up of an input layer, N hidden layers, and finally an output layer. Each layer has its own activation function, and will adjust its weights and biases to determine the final output of the model.[7] These networks are heavily inspired by biological processes which occur in the brain.
artificial neural networks are a general purpose model used to solve a number of common problems.

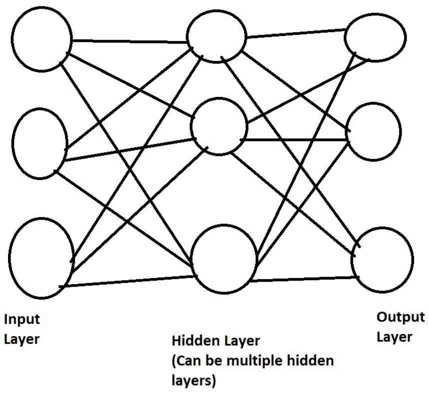

Figure 1.2: Basic Example of Artificial Neural Network

A basic example of an artificial neural network is shown in Figure 1.2. As shown in the figure, the network has an input layer, a hidden layer, and an output layer. Generally when creating these networks we determine the number of neurons in both the input and the output layers based on the different classifications we are trying to predict. The above network could be used to predict if an image is of a cat, a dog or a fish for example. There can multiple hidden layers in an ANN and the number of neurons in each layer can be adjusted. In reality Artificial Neural Networks will typically be far bigger than the example given above in terms of neurons and hidden layers but for illustrative purposes the above diagram will suffice. Each neuron will also have its own weights, biases and activation function which will determine whether a neuron fires or not. Common activation functions include ReLU (Rectified Linear Units), Sigmoidal function and tanh.

1.3 What is A Convolutional Neural Network? (CNN)

A convolutional neural network, or CNN for short, is a type of neural network which is primarily used for tasks involving image and pattern recognition[7] The structure is similar to an ANN in which we have an input layer, N hidden layers, and finally an output layer. As with the Artificial Neural Network each of these layers will have an activation function and it's own weights and biases to determine the final output for a given input. The model will take an image as input, the image is made up of vectors (RGB) or a similar format and from that image the model will determine certain patterns. For example, the output might be a classification of whether or not Covid-19 is present or not. This application will be discussed in more detail later in the dissertation.

There are a few ways in which CNNs differ from ANNs, in that they are comprised of three types of layers which are the convolutional layer, the pooling layer and fully connected layers[7]. The convolutional layer is responsible for determining the output of a given input (for example if we were trying to detect images of cats then this layer would activate 1 for a cat and 0 for images which were not cats), the pooling layer will reduce the parameters of a given input by means of downsampling, and finally the fully connected layers will then determine and classify the output for a given input. The convolutional layers parameters utilize learnable kernels, and this layer also produces a $2D$ activation map which will be used to determine if a neuron fires or not for a given input. We can adjust hyper parameters in the convolutional layer to greatly reduce the complexity of the model through optimization, which can be achieved by adjusting the following hyper parameters: depth, stride and zero padding.

Depth is related to the output volume produced by the convolutional layers in the model which can be manually set by adjusting the number of neurons in each layer. Reducing the depth of the model can greatly decrease the training time but at the expense of performance. Stride is related to the spatial dimensionality of the input which will determine the receptive field (every neuron is connected only to a small region of the input this region is referred to as the receptive field[7]), if the stride is set to a low integer we will produce extremely large activations, and if it is set too high we won't produce enough activations.

Finally, zero-padding this will pad the border of the images ingested by the model reducing their dimensionality, padding is useful for increasing the accuracy of the model as it can possibly eliminate areas of the image which are not useful for the model and can also improve training time times in some use cases.[8]

Through the adjustment of the hyper parameters mentioned above, and through the utilization of different activation functions, the accuracy of the convolutional neural network can be improved through a process of trial and error.

1.4 Supervised Learning

Supervised learning is a type of learning involving labelled data to train the model[9]. The data is labelled manually by a data scientist which can be a long and laborious process depending on a number of factors (size of the data, number of classes, etc.), but offers many benefits when it comes to training models. Supervised learning performs extremely well at tasks involving classification (classifying data into a given category), and regression (understanding the relationships between independent and dependent variables).

1.5 Unsupervised Learning

Unsupervised learning is a type of machine learning which involves using unlabelled data to train machine learning models[9]. This type of machine learning requires no human interven-

tion since the data is unlabelled and will detect relationships between data based on the raw data fed in to the model. This type of machine learning is used for the following tasks: clustering(grouping data together based on shared characteristics or features), association(Finding relationships between features), and dimensionality reduction (Reducing the number of features in a given dataset without compromising the integrity of said data). The key differences between supervised and unsupervised learning are: labelled vs unlabelled datasets, and finding relationships in data (unsupervised) or trying to predict and classify data (supervised).
In this dissertation we examine the use of both labelled and unlabelled datasets to train and test the model.

1.6 Tensorflow

Tensorflow is an open-source library used for machine-learning and artificial intelligence research worldwide[10]. Tensorflow provides numerous modules and classes which form the foundation of building both the generative adversarial network and the convolutional neural network. There have been numerous case studies proving the efficacy of Tensorflow in solving many AI / ML problems and the library is used by research teams in organisations such as Google, Airbnb, ARM, Coca-Cola, Intel, and many more[11].
Given the reputation and widespread use of Tensorflow, and the vast amount of documentation around the framework it seems an ideal library for the implementation of GANs and CNNs for this study.

1.7 Keras

Keras is a deep-learning framework for Python which provides a number of helpful functions and methods for creating and training the CNN[12]. Keras is built on top of Tensorflow and simplifies data loading, pre-processing and the overall building of the model. Keras is commonly used by data-scientists and researchers due to the powerful methods it offers and the time it saves. The additional classes and modules Keras provides on top of Tensorflow will help to reduce the time taken to build and develop of building both the convolutional neural network and the generative adversarial network.
Like Tensorflow, Keras has been used by a number of companies and is well recognised in the Artificial Intelligence community. Its uses include Computer Vision, Natural Language Processing, Generative Deep-Learning and Reinforcement Learning amongst others[13].

1.8 Background of Problem & Aims of This Paper

COVID-19 is a highly transmissible virus which has caused a worldwide pandemic and has claimed many lives. There have been 616,951,418 cases worldwide and 6,530,281 deaths as of the 4th of October 2022[14]. During the pandemic, Ireland alone had a total of over 1.6 million confirmed cases and nearly 8,000 deaths[15]. This has led many researchers to pursue the goal of automating the detection of COVID-19 to partially relieve the immense pressure put on medical staff throughout the pandemic.

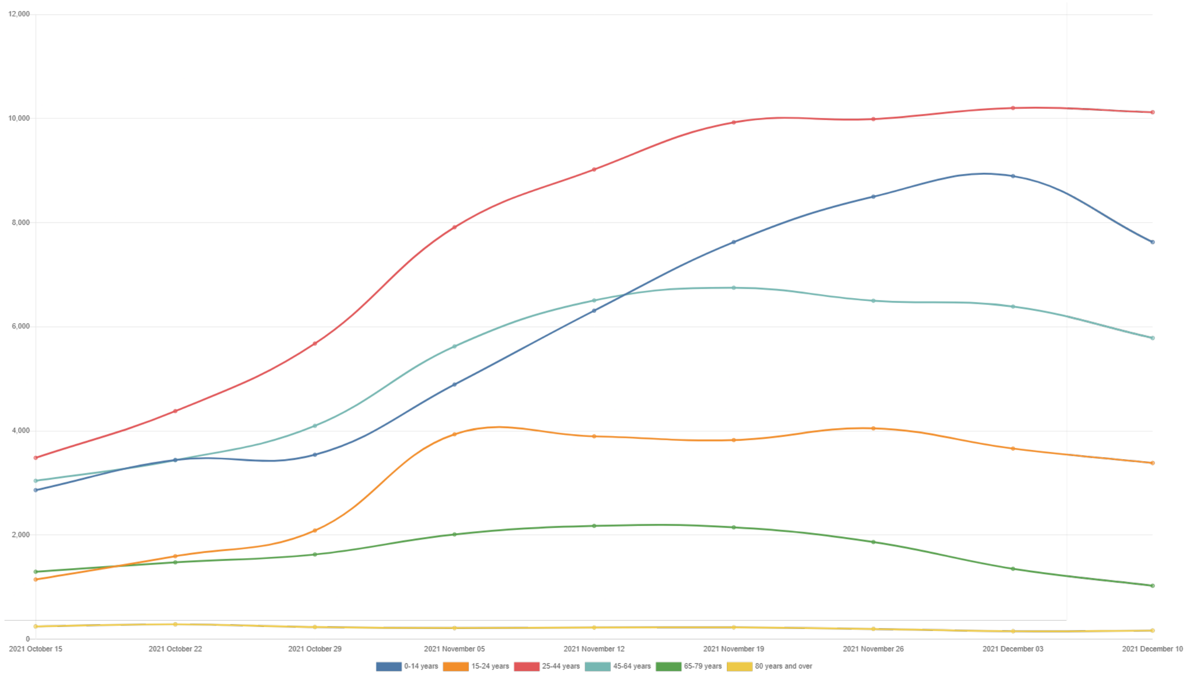

Figure 1.3: Graph of COVID-19 Statistics by age-range Ireland from October 2021 - December 2021 Courtesy of CSO[16]

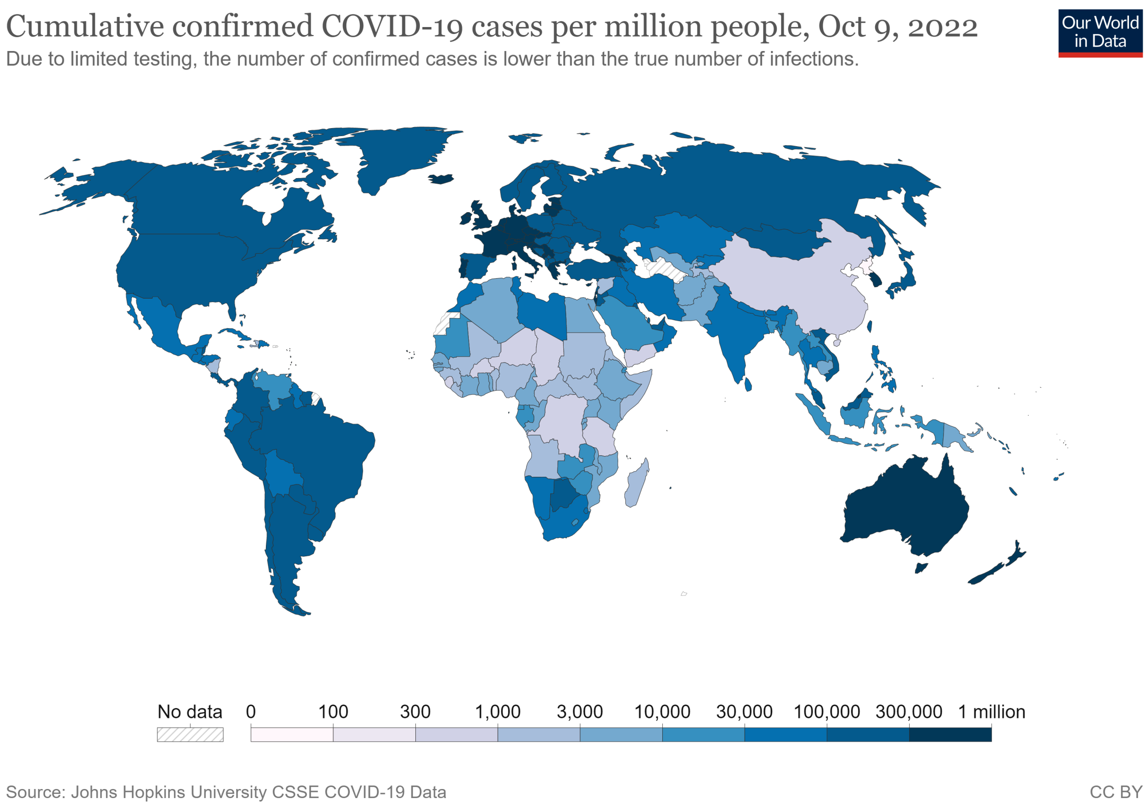

Figure 1.4: Cumulative cases of the COVID-19 virus world-wide courtesy of Our World in Data[17]

The main objective of this research is to develop a robust model which can accurately analyze X-Rays of patients and determine from said X-rays if the patient is afflicted with COVID-19. This will be achieved by utilizing a number of different GAN architectures which will create realistic "fake" data which will then be used to train a number of models. From this training I plan to compare and contrast the results when generating data with different architectures to determine the best configuration for data generation to train the CNN model. There has been some success in utilizing convolutional neural networks to automate the detection of the virus[18][19]. Through the use of data-augmentation utilizing a variety of GAN architectures that such Convolutional Models will be improved upon and made more accurate.

I plan on utilizing existing data sets which I will list in the next section when training the Generative Adversarial Models, through trial and error I plan on determining the best architecture of GANs to use for training the model for this use-case.

1.9 Datasets

1.10 About the datasets

Before beginning the training of the model it is important to explore and understand each of the datasets. There are a total of three datasets which will be used in the course of this research, I will explain more about these datasets below.

1.10.1 COVID-19 Chest X-ray

The COVID-19 Chest X-ray data set is a data set which is comprised of labelled X-Ray Images taken from a number of patients. This dataset contains 357 X-ray images of COVID positive patients, and Chest X-Rays of those afflicted with another disease (MERS, SARS, and ARDS). This dataset also includes a metadata file listing the diagnosis of the patient along with a number of other features.

1.10.2 COVID-19 Radiography Database

The COVID-19 Radiography Database is made up of 3,616 images of chest X-Rays taken from COVID positive patients, 10,192 Images of lung X-Rays taken from healthy patients, and 1,345 X-ray images of viral pneumonia positive patients. All images in this dataset are PNG (Portable Network Graphic) images and are at a resolution of height 299 pixels and width 299 pixels eliminating the need for pre processing of the images, the dataset also includes metadata for each of the images in this dataset showing a number of features with the diagnosis of the patient as well. This data in this dataset was gathered by a team of researchers from Qatar University, Doha, Qatar, and the University of Dhaka, Bangladesh along with their collaborators from Pakistan and Malaysia.

1.10.3 COVID-19 Pneumonia Normal Chest Xray PA Dataset

The COVID-19 Pneumonia Normal Chest X-Ray PA dataset is comprised of a train set containing 74 Normal X-Ray Images taken from healthy Patients and afflicted with Pneumonia and a test-set containing a Normal set containing 20 chest X-Rays taken from healthy patients and a Pneumonia set containing 20 images. The images in this dataset are unlabelled and no metadata is offered, however the images are segregated into separate files listing the diagnosis.

1.10.4 Use of datasets in This Project

I plan to use each of these datasets to train and test the model and use data-augmentation to increase the train and test-sets by utilizing Generalized Adversarial Networks. When using these datasets in conjunction it is my hope that the GAN will have enough data to be effective when generating new sample images to train the final model.

1.11 Structure of This Thesis

This thesis is broken into 6 chapters in total, this section will include the headings of the chapters and a brief summary of each chapter below:

1.11.1 Chapter 1 - Introduction

This chapter will offer the reader of this thesis a brief introduction to a number of core concepts which will be necessary to understand before diving deeper into this thesis. It is important that the reader has a basic understanding of generative adversarial networks, convolutional neural networks, artificial neural networks, supervised & unsupervised learning, and the overall question that this research proposes before discussing the implementation or discussing pertinent literature in this field.
In this section I will frame the research question, explain what a generative adversarial network is, it's function, and how it works, I will also explain artificial neural networks and convolutional neural networks, and I will discuss the basic methodologies relating to the implementation of this project. I will also discuss the libraries used to implement the practical artifact, datasets used to train the model and give the reader of this thesis a clear understanding of the key aims of this research.

1.11.2 Chapter 2 - Literature Review

In this section I will review pertinent literature related to the problem domain and discuss the ideas and concepts presented in these papers. I will also review the results from the research conducted in these papers and use them as a metric to gauge the performance of my own model. The papers will also be compared and contrasted and I will discuss the findings and how useful these papers were when conducting my own research. It is very important to understand the problem domain before beginning implementation of this project to ensure that I am not "reinventing the wheel". This section will also provide the reader of this thesis with the most up-to-date progress made within the problem domain.

1.11.3 Chapter 3 - Implementation

In this section I will discuss the architecture of the convolutional model, the various architectures of generative adversarial networks implemented, how the models were trained and the overall design of the code implemented and the rationale behind certain design choices. I will also show the results from training the models and discuss how through trial and error I was able to improve the various models and will include code samples so that the models can be reviewed by the reader or re-implemented by them.

1.11.4 Chapter 4 - Results

In this section I will review the results achieved from training the best models and suggest how they may possibly be improved. I will be showing lots of graphs / tables in this section to gauge each model's test / dev set errors and I will also be comparing and contrasting the effects of the different GAN architectures implemented as well as discussing the results of the convolutional model.

1.11.5 Chapter 5 - Further Research and Conclusions

In this section I will discuss further research that may need to be done by any researchers who would like to build upon this research. I will also review where the models could be improved and what I'd do differently if I were to conduct this research again. I will also discuss common issues I faced during the implementation of this project and how I overcame them. This section will be a summary of all the research conducted, the code, and my experience overall throughout the writing of this thesis.
This will be the final section of the paper and will tie the entire thesis together.

Chapter 2

Literature Review

2.1 Introduction to Literature Review

The first reported cases of COVID-19 occurred in Wuhan, China on December 12th, 2019[20], when a number patients began to exhibit ”symptoms of an atypical pneumonia-like illness that does not respond well to standard treatments”. It was not until December 31, 2019 that the World Health Organization(WHO) Country Office in China was informed of several more cases of this strange virus described as ”pneumonia of unknown etiology”, the symptoms of this new virus were shortness of breath, and fever. All the initial cases observed seem to have been connected to a market called Huanan Seafood Wholesale Market. On January 1st of 2020 the Huanan Seafood Wholesale Market was shut down amid concerns over the spread of this new virus. On January 3rd the Government of China alerted the World Health Organization that they had identified over 40 new cases of this pnneumonia-like disease, on the 5th of January, Chinese public health officials shared the genetic sequence of the new virus with the world through a database that could be accessed by the public. Following the release of this information the CDC(Centers for Disease Control and Prevention), which is a US Government funded heath-care research agency, began an investigation into the origins of this new virus. The origins of COVID-19 are not clear and are still being researched, the most likely explanation offered by scientists is that it originated in the Huanan Seafood Wholesale Market from animals sold there, a likely culprit is the Raccoon Dog which are used for fur and food in China[21] but other theories suggest that a lab leak at a biological weapons facility[22] may be responsible for the creation of the virus. Some researchers are currently suggesting that blood samples taken from animals sold at the Hunan Market and samples from the people who sold them may lead to definitive evidence of the diseases origins[21].

Although the origins of COVID-19 still remain up for debate one thing is very clear when studying the virus, it’s high transmission rate and the speed at which it can spread made it one of the deadliest viruses in human history.[23][24].

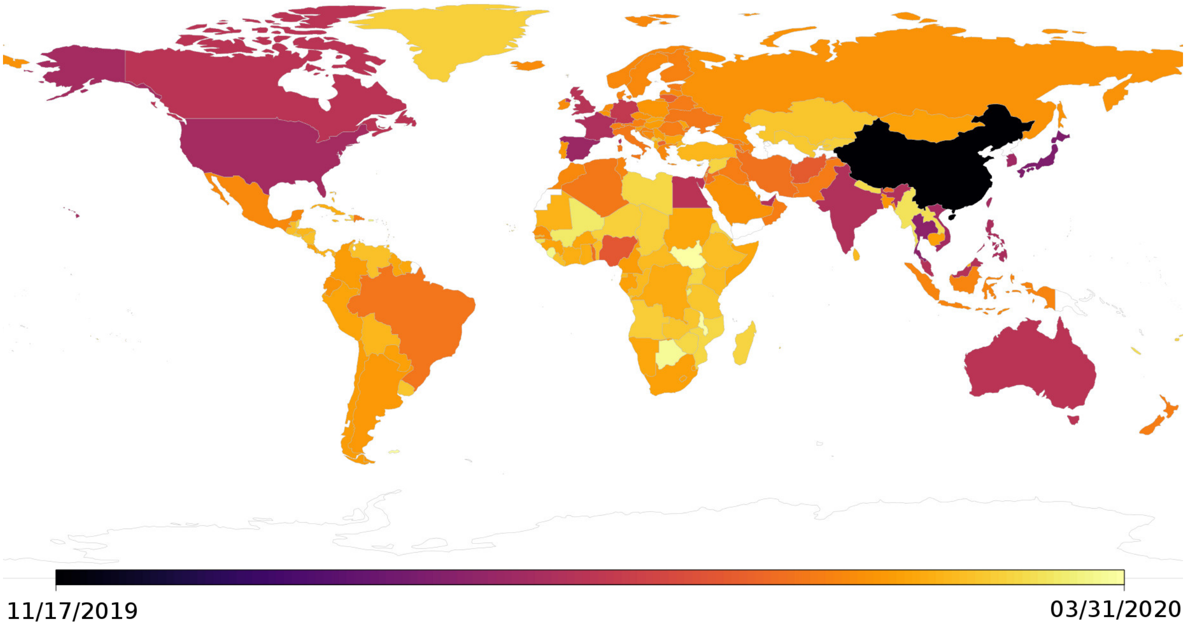

Figure 2.1: Estimated dates of first COVID-19 cases around the World, Image courtesy of Roberts,Rossman and Jarić[25]

Due to this rapid spread of the virus many governments were unprepared for dealing with such an outbreak. The Artificial Intelligence Community had published many papers and conducted much research into designing automated tools which could relieve medical professionals of the extreme stress they were under, unfortunately most of the models trained were of no use to medical professionals and some were even deemed harmful[26]. There were many limitations when it came to training automated diagnostic tools for COVID-19 such as incorrect assumptions about the data, lack of data quality and lack of data in general. Due to the lack of data and the urgent need for diagnostic tools many of the models were trained using poor quality data or incorrect data. Such poor models would have had drastic effects if patients who were COVID positive were diagnosed as negative by the model, models suffering from high false-negative rates would have drastic consequences for the patients afflicted with COVID.

The models trained on what has been termed as "Frakenstein Datasets" suffered immensely as some of the data came from the same source meaning the same data from the training set could be used in the test set, this would severely impact the performance of the model as it would have overfit on the data from which it was trained. These models which were overfit on the data would seem to have a high accuracy but ultimately would perform poorly on real world data.

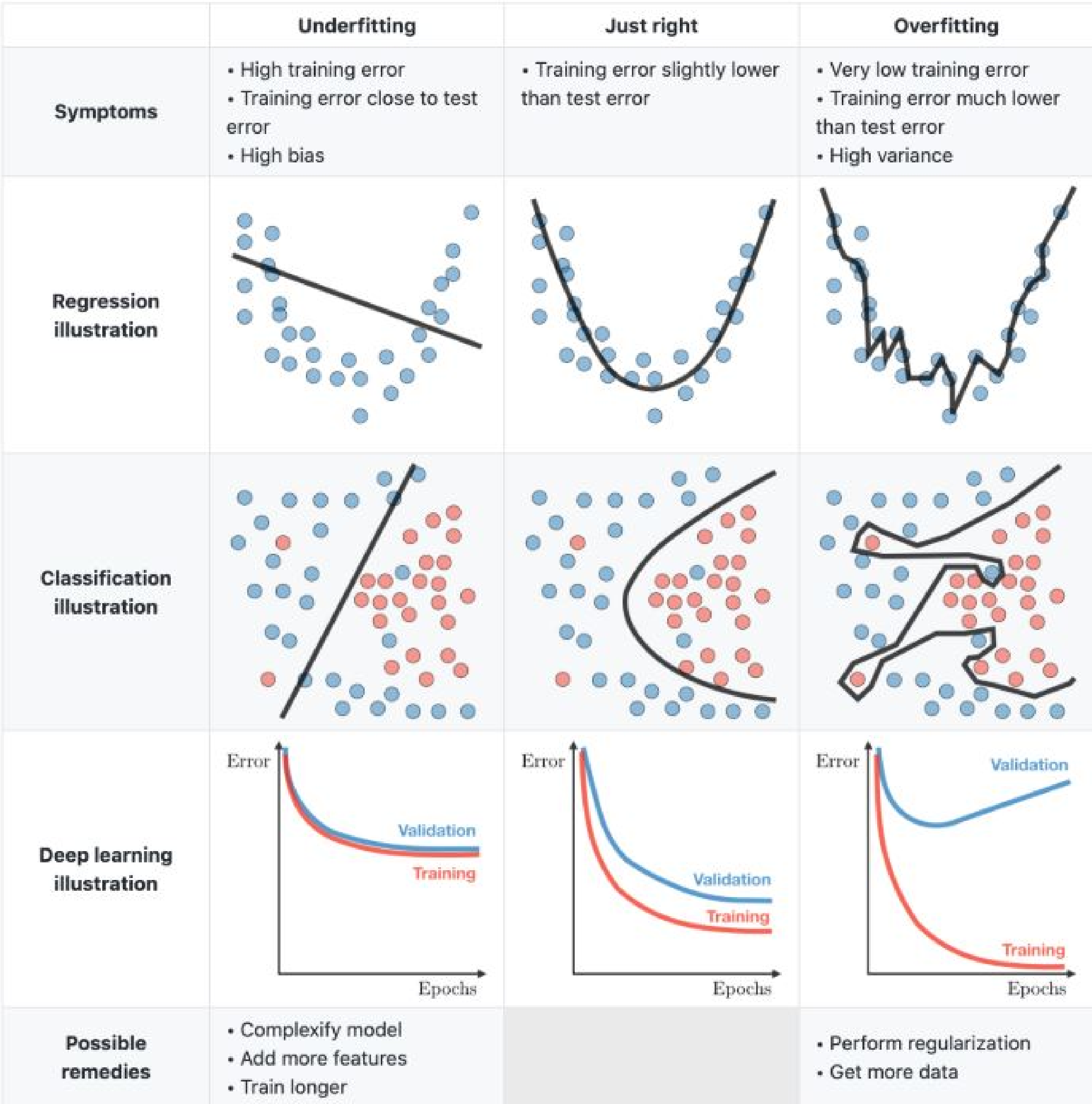

Figure 2.2: Examples of Overfitting, Underfitting and Optimal models, Image courtesy of Abhishek Shrivastava[27]

The figure above2.2 shows how a model's performance can be analyzed, underfitting yields a high training error and high bias meaning that the model will perform poorly on the training, test, and dev sets. Overfitting would have a very low training error which would be lower than the test, and dev set error and wouldn't be fit for purpose when analyzing real world data. The optimal model has a training error that is slightly lower or in and around the same accuracy as the test and dev sets.

The lack of medical experience also played a role in the poor performance of these models as many of the AI researchers training these models would be unfamiliar with flaws in the data. Bias of the radiologist labelling the X-rays of patients also played a role as the radiologist could have inaccurately diagnosed the patient as COVID positive or negative. Private Artificial Intelligence companies also played a role in poor model development as published models from researchers tied to the company also showed that these models had a high risk of bias. As the pandemic progressed more and more data was made available to researchers which was able to mitigate some of the problems stated above leading to more accurate and robust models which we will explore in the later sections.

2.2 Analysis of Existing Models for Automated COVID-19 Detection

In a paper entitled "A Deep Learning-Based Diagnosis System for COVID-19 Detection and Pneumonia Screening Using CT Imaging"[28] researchers investigated a deep-learning approach to creating a diagnostic tool for COVID-19. The research involved utilizing data taken from Computed Topography scans. These scans segmented the infected regions of a patients lungs to determine if said patient was afflicted with the COVID-19 virus. The researchers also used a technique called "Contrast Limited Adaptive Histogram Equalization" which is a pre-processing method which removed noise and intensity to create a homogeneous data set. The researchers also removed black slices from the images so that only the region of interest was highlighted, further enhancing the performance of the model. U-Net architecture, which is based on Convolutional Neural Network encoders and decoders, was used in the creation of this model to allow for a more timely and accurate image segmentation to generate the lung and infection segmentation models. Four-fold cross validation was then used to analyze the performance of the model along with a three-layered CNN architecture which was comprised of additional fully-connected layers followed by a softmax output layer which was used for classification of the images.

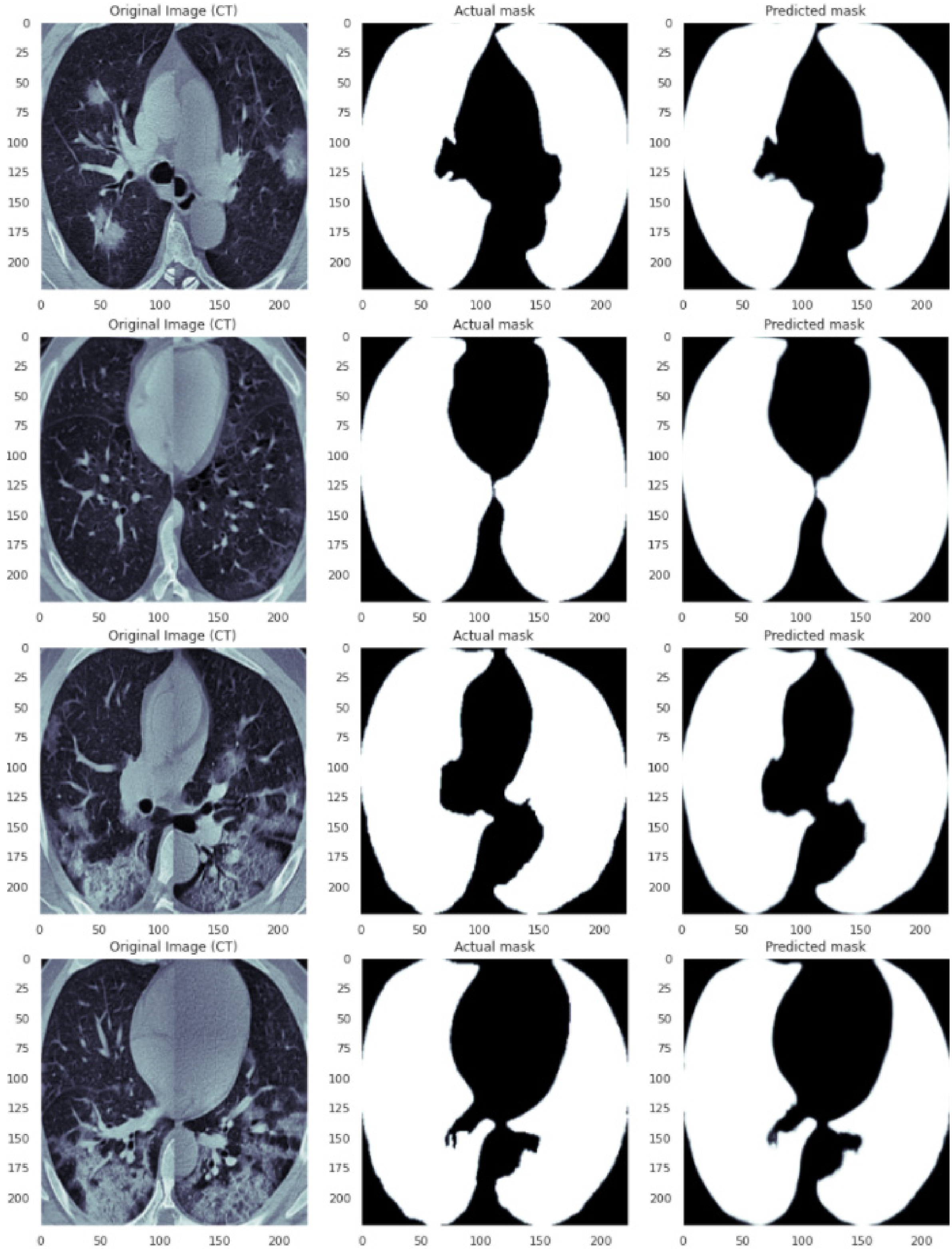

Figure 2.3: Examples of CT Qualitative images lung segmentation[28]

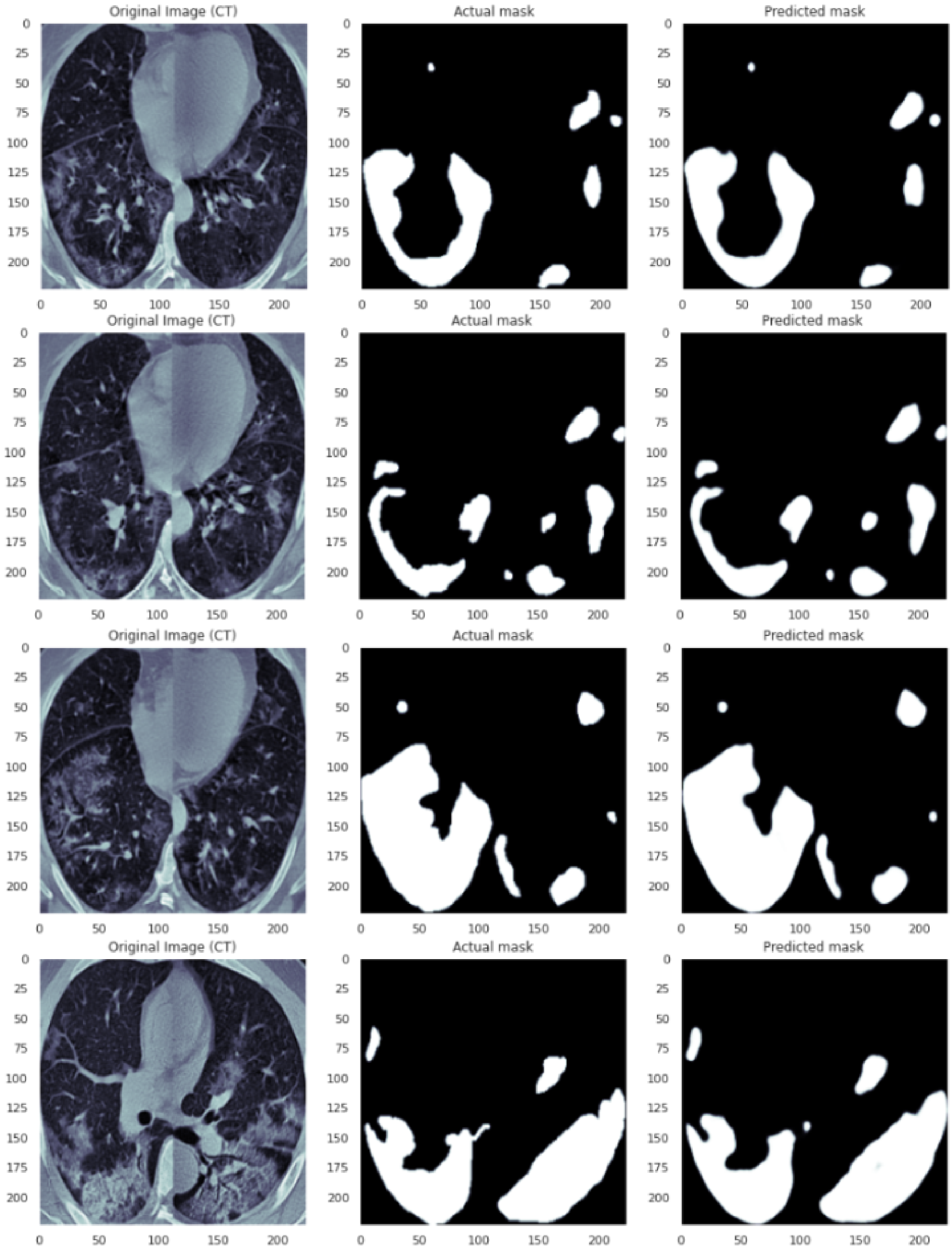

Figure 2.4: Examples of CT qualitative images infection masks[28]

The first image2.3 shows the result of the lung segmentation results between the ground truth and the researchers proposed four slice model which is compromised of four slices taken from different CT Scans. first column: original CT scan,second column: ground truth, third column: predicted lung masks
0.5mm In the second image2.4 we can observe the qualitative comparison between the researchers infection segmentation results which is made up of four slices from different CT scans and the ground truth. First column: original CT scan, second column: ground truth, third column: predicted infection masks
0.5mm Utilizing a 70% - 30% training set and validation set split the researchers demonstrated that the proposed system achieved a dice score of 0.98% and 0.91% for lung and infection segmentation tasks. Additionally the system accurately diagnosed patients afflicted with COVID-19 0.98% of the time. The development of this model suffered from lack of data as only 20 CT scans were used to train and the model, the limited data set used suggests the researchers model may have possibly been overfitting the training data. The researchers mention as much in the conclusion section of this paper "The main limitation of this study is the use of a small, but sufficient number, of model training data. Confidentiality restrictions and the high cost of labeling partly explain the absence of a large number of COVID-19 clinical CT images. Indeed, combining datasets collected under different labeling regimes is often problematic because, in general, the data collected is heavily influenced by the instructions provided to annotators.". As we can see from the quoted section of this paper there is a high potential for bias in the data set used by the researchers. This links back to what was termed "Frankenstein Datasets" which I mentioned in the introduction section to this literature review.

Chapter 3

Implementation

3.1 Libraries Used

3.2 GAN Architectures

3.3 CNN Design

Chapter 4

Results of Research

Chapter 5

Future Work and Research

Bibliography

[1] A. Binkheder *et al.*, "COVID-19 Data Explorer," [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8701465/.

[2] S. Maior *et al.*, "Convolutional Neural Network Model Based on Radiological Images to Support COVID-19 Diagnosis: Evaluating Database Biases," [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0247839#sec025/.

[3] A. Goodfellow *et al.*, "Generative Adversarial Nets," 2014. [Online]. Available: https://arxiv.org/pdf/1406.2661.pdf.

[4] P. Zhu *et al.*, "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks," 2020. [Online]. Available: https://arxiv.org/pdf/1703.10593.pdf.

[5] A. Karras Laine, "A style-based generator architecture for generative adversarial networks," 2019. [Online]. Available: https://arxiv.org/pdf/1812.04948.pdf.

[6] K. Oord Kalchbrenner, "Pixel recurrent neural networks," 2016. [Online]. Available: https://arxiv.org/pdf/1601.06759.pdf.

[7] K. O'Shea and R. Nash, "An Introduction to Convolutional Neural Networks," [Online]. Available: https://arxiv.org/pdf/1511.08458.pdf.

[8] M. Hashemi, "Enlarging smaller images before inputting into convolutional neural network: zero-padding vs. interpolation," [Online]. Available: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0263-7.

[9] J. Delua, "Supervised vs. Unsupervised Learning: What's the Difference?," [Online]. Available: https://www.ibm.com/cloud/blog/supervised-vs-unsupervised-learning.

[10] Multiple, "Tensorflow," [Online]. Available: https://www.tensorflow.org/.

[11] Multiple, "Tensorflow Case Studies," [Online]. Available: https://www.tensorflow.org/about/case-studies/.

[12] Multiple, "Keras," [Online]. Available: https://keras.io/.

[13] Multiple, "Keras Examples," [Online]. Available: https://keras.io/examples/.

[14] Multiple, "WHO Coronavirus (COVID-19) Dashboard," [Online]. Available: https://covid19.who.int/.

[15] Multiple, "WHO Coronavirus (COVID-19) Dashboard Ireland," [Online]. Available: https://covid19.who.int/region/euro/country/ie/.

[16] Multiple, "COVID-19 Deaths and Cases Statistics," [Online]. Available: https://www.cso.ie/en/releasesandpublications/ep/p-covid19/covid-19informationhub/health/covid-19deathsandcasesstatistics/.

[17] Our World In Data(multiple), "COVID-19 Data Explorer," [Online]. Available: https://ourworldindata.org/explorers/coronavirus-data-explorer?tab=map&facet=none&Metric=Confirmed+cases&Interval=Cumulative&Relative+to+Population=true&Color+by+test+positivity=false&country=USA~ITA~CAN~DEU~GBR~FRA~JPN.

[18] a. o. Ahmed Wang, "Automated Detection of COVID-19 Through Convolutional Neural Network Using Chest X-ray Images," [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0262052/.

[19] C. E. Belman-López, "Detection of COVID-19 and Other Pneumonia Cases using Convolutional Neural Networks and X-ray Images," [Online]. Available: http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0120-56092022000100108/.

[20] Multiple, "CDC Museum COVID-19 Timeline," [Online]. Available: https://www.cdc.gov/museum/timeline/covid19.html#Late-2019.

[21] A. Maxmen, "Wuhan market was epicentre of pandemic's start, studies suggest," [Online]. Available: https://www.nature.com/articles/d41586-022-00584-8.

[22] S. M. Amy Maxmen, "The COVID lab-leak hypothesis: what scientists do and don't know," [Online]. Available: https://www.nature.com/articles/d41586-021-01529-3.

[23] E. Gamillo, "Covid-19 Surpasses 1918 Flu to Become Deadliest Pandemic in American History," [Online]. Available: https://www.smithsonianmag.com/smart-news/the-covid-19-pandemic-is-considered-the-deadliest-in-american-history-as-death-toll-surpasses-1918-estimates-180978748/.

[24] A. Mckeever, "COVID-19 surpasses 1918 flu as Deadliest Pandemic in U.S. History," [Online]. Available: https://www.nationalgeographic.com/history/article/covid-19-is-now-the-deadliest-pandemic-in-us-history/.

[25] R. Roberts and Jarić, "Dating First Cases of COVID-19," [Online]. Available: https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1009620/.

[26] W. D. Heaven, "Hundreds of AI tools have been built to catch covid. None of them helped.," [Online]. Available: https://www.technologyreview.com/2021/07/30/1030329/machine-learning-ai-failed-covid-hospital-diagnosis-pandemic/.

[27] A. Shrivastava, "Underfitting Vs Just right Vs Overfitting in Machine learning," [Online]. Available: https://www.kaggle.com/getting-started/166897.

[28] B. Mahmoudi *et al.*, "A Deep Learning-Based Diagnosis System for COVID-19 Detection and Pneumonia Screening Using CT Imaging," [Online]. Available: https://www.mdpi.com/2076-3417/12/10/4825/.