HERIOT-WATT UNIVERSITY

FINAL YEAR DISSERTATION

Automated Diagnosis of COVID-19 using Medical Imagery

Author:
Alister George Luiz

Supervisor:
Dr. Hani Ragab

A thesis submitted in fulfillment of the requirements for the honours degree of Bachelor of Science

in the

School of Mathematical and Computer Sciences

Declaration of Authorship

I, **Alister George Luiz**, confirm that this work submitted for assessment is my own and is expressed in my own words. Any uses made within it of the works of other authors in any form (e.g., ideas, equations, figures, text, tables, programs) are properly acknowledged at any point of their use. A list of the references employed is included.

Signed: **Alister George Luiz**

Date: **November 26, 2020**

"The only limit to our realization of tomorrow will be our doubts of today."

Franklin D. Roosevelt

Abstract

The Coronavirus Disease (COVID-19), ever since its inception in late 2019 has spread all across the world and as a result, led to an increased burden on healthcare professionals due to the urgent need for rapid disease diagnosis and effectuating quarantine protocols.

Currently, the Real-Time Reverse Transcription Polymerase Chain Reaction (RT-PCR) test recommended by the World Health Organization (WHO) remains the front runner in terms of COVID-19 diagnosis when compared to other testing mechanisms. But there involve serious downsides to using this test as a primary diagnosis tool, a few of them include the shortage in RT-PCR test kits, delays in receiving test results (up to 2 days), but most importantly the low accuracy rate of COVID-19 detection.

The primary objective of this project is to develop a fully automated framework minimizing human to human interaction to rapidly diagnose patients with COVID-19 using medical imagery such as Chest X-ray's or CT scan's applying deep learning techniques and therefore, achieve a much higher accuracy rate compared to the traditional RT-PCR test ultimately reducing the workload on healthcare professionals.

Keywords: COVID-19, SARS-CoV-2, Chest X-ray, Chest CT, Medical Image Classification, Deep Learning, Convolutional Neural Networks, Transfer Learning

Acknowledgements

I would like to thank Dr Hani Ragab for his constant guidance and support throughout the tenure of this project. I would also like to thank my parents whose prayers and encouragement made me who I am today …

Contents

List of Figures

List of Tables

List of Abbreviations

COVID-19	**C**oronavirus **D**isease 2019
SARS-CoV-2	**S**evere **A**cute **R**espiratory **S**yndrome **C**oronavirus **2**
RT-PCR	**R**everse **T**ranscription **P**olymerase **C**hain **R**eaction
CT	**C**omputed **T**omography
CNN	**C**onvolutional **N**eural **N**etwork
WHO	**W**orld **H**ealth **O**rganization
RNA	**R**Ribo**n**ucleic **A**cid
DNA	**D**eoxyribo**n**ucleic **A**cid

Chapter 1

Introduction

COVID-19, which was declared a global pandemic by the World Health Organization on March 11, 2020, has affected millions of lives worldwide in terms of both health and finances and have also had a severe global economic impact.

With over a million tests carried out on average daily across the world for diagnosing patients with COVID-19 (*Daily COVID-19 Tests* 2020), it is therefore the need of the hour to alleviate the burden on healthcare professionals who conducts these diagnoses on a day-to-day basis, and more importantly, minimize the exposure rate between patients and healthcare professionals.

1.1 Aim

This project aims to **automate the diagnoses of COVID-19 with medical imagery using Deep Learning**. The main focus being to achieve the highest diagnostic accuracy possible, minimizing the false-negative rate, which, if not accounted for may have adverse real-world implications. We also intent to compare the usage of X-rays vs CT scans in this project considering the limitations in terms of the equipment available on medical facilities, radiation exposure, and the results obtained in both cases.

The overall goal of this project is to develop a framework that enables highly accurate rapid diagnoses of COVID-19, all while providing a safer environment for healthcare professionals by minimizing the rate of exposure and following quarantine protocols therefore curbing the spread of COVID-19.

1.2 Objectives

The objective of this project is to develop a framework that rapidly diagnoses patients with COVID-19 with medical imagery and thereby provide assistance to healthcare professionals.

These are the primary objectives for this thesis:

- Analyse medical imaging features of COVID-19.
- Build a deep learning model which diagnoses patients with COVID-19 using chest X-rays.
- Test the accuracy of the proposed COVID-19 diagnoses model by comparing with other similar models implemented previously.
- Develop a deep learning API which can be used by healthcare professionals and medical facilities to diagnose COVID-19.
- Optionally, build a deep learning model which diagnoses patients with COVID-19 using chest CT scans and compare the accuracy and results obtained from both models respectively.

1.3 Manuscript Organization

This manuscript contains 6 sections, starting with a comprehensive **Introduction** of the main objectives of this project. This is followed by the **Literature Review** section which aims to synthesize and sum up relevant research and implementations previously conducted in this same field. The next section **Requirements Analysis** conducts a detailed study on the use cases of this project and identifies user requirements and labels their priority. An additional section **Design** has also been included, which demonstrates the pipeline and the workflow of the proposed model. Following this is a section for **Evaluation Strategy** which specifies the analysis and assessment that needs to be administered for this project. The last section is dedicated to **Project Management** which provides a detailed schedule that must be strictly adhered to, in order to ensure the success of this project, as well as examining the risks involved and the ethical, legal and, social issues pertaining to this project.

Chapter 2

Literature Review

A comprehensive analysis of the existing research and methodologies pertaining to COVID-19 detection using medical imagery and deep learning models have been provided in this section.

2.1 The COVID-19 Pandemic Era

Coronavirus disease (COVID-19) is a highly contagious respiratory disease caused by the newly discovered coronavirus. The virus mainly spreads through the discharge of saliva droplets when an infected person coughs or sneezes (World Health Organization, 2020a). A brief description of the novel coronavirus and its spread across the world is discussed in the following section.

2.1.1 Rise of the Global Pandemic

COVID-19 which is now officially declared as a global pandemic by the WHO was initially discovered in late 2019 emerging from Wuhan, People's Republic of China. A media statement was released by the Wuhan Municipal Health Commission confirming multiple cases of "Viral Pneumonia from an unknown cause" on December 31^{st} 2019 (Wuhan Municipal Health Commission, 2020).

On January 7^{th} 2020, this unknown disease was identified as the novel coronavirus by the WHO. Three days later, the first known death caused by the coronavirus was reported (World Health Organization, 2020b). The spread of the virus continued rapidly within China and on January 20^{th} 2020, WHO reports the first confirmed cases outside China in

Thailand, Japan, and South Korea. The very next day The United States reports its first confirmed coronavirus case (Edwards, 2020a).

Following these set of events saw the introduction of quarantine protocols via lockdowns. Starting with Wuhan, many other cities across the world also adopted the same orders to suspend the spread of the coronavirus. This prompted the WHO to declare the outbreak a global public health emergency (Edwards, 2020b).

Within the span of a month, the death toll from COVID-19 surpassed that of SARS and the WHO gives the official name for the disease caused by the coronavirus "COVID-19" (Muccari, Chow, and Murphy, 2020). The adverse effects of COVID-19 on various industries and the stock market began to show.

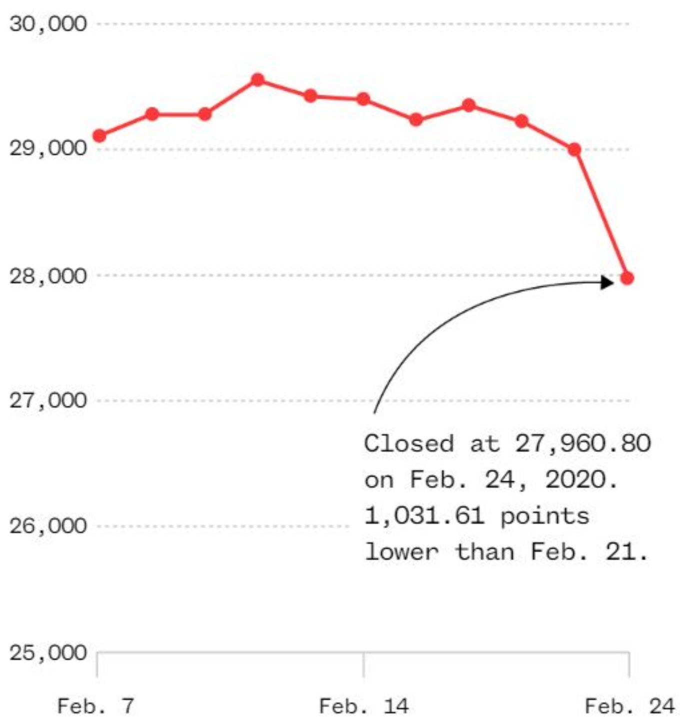

FIGURE 2.1: Dow Jones Industrial Average experienced the worst day in two years. (White and Bayly, 2020)

Travel bans, high-profile event cancellations and activation of emergency funds were implemented across different countries as the number of positive cases rose above 100,000 (Shabad, 2020).

Due to the rapid spread of the virus worldwide, on March 11th 2020, the WHO declares the coronavirus outbreak as a global pandemic (Jason, 2020). Over the next few months, nationwide lockdowns were enforced in countries such as The United Kingdom, India, South Africa, Italy, Belgium, and so on.

At the end of March 2020, The United States coronavirus cases officially surpassed China, with the former reporting 82,474 cases and the latter 81,961 cases.

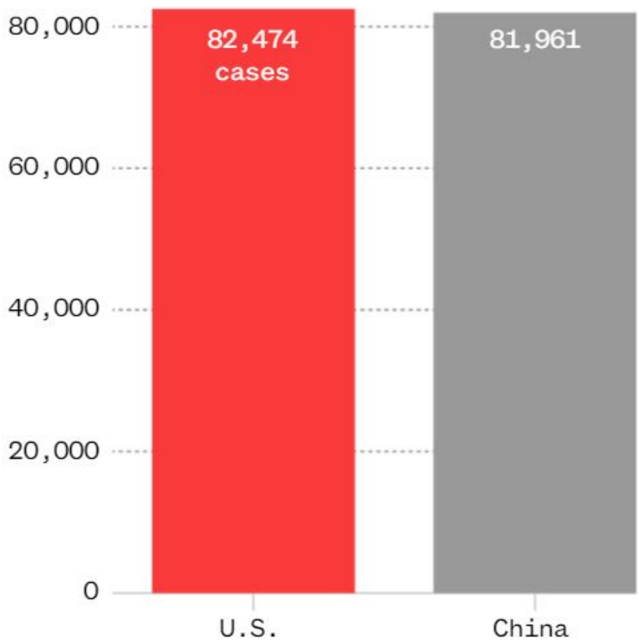

FIGURE 2.2: COVID-19 cases comparison between The United States and China (Givetash, 2020)

On April 2nd 2020, the number of coronavirus cases worldwide surpassed 1 million with the number of deaths exceeding 51,000 (Phillips, 2020). The number of positive cases doubled in April and this trend continued to follow for the next few months were as of October 2nd 2020, the total number of worldwide COVID-19 cases stand at 34,312,510, and the death toll surpassing over a million, to be precise 1,023,243 (*Coronavirus (COVID-19)*).

Among the countries affected by COVID-19, The United States, India and Brazil share the highest percentages of positive cases.

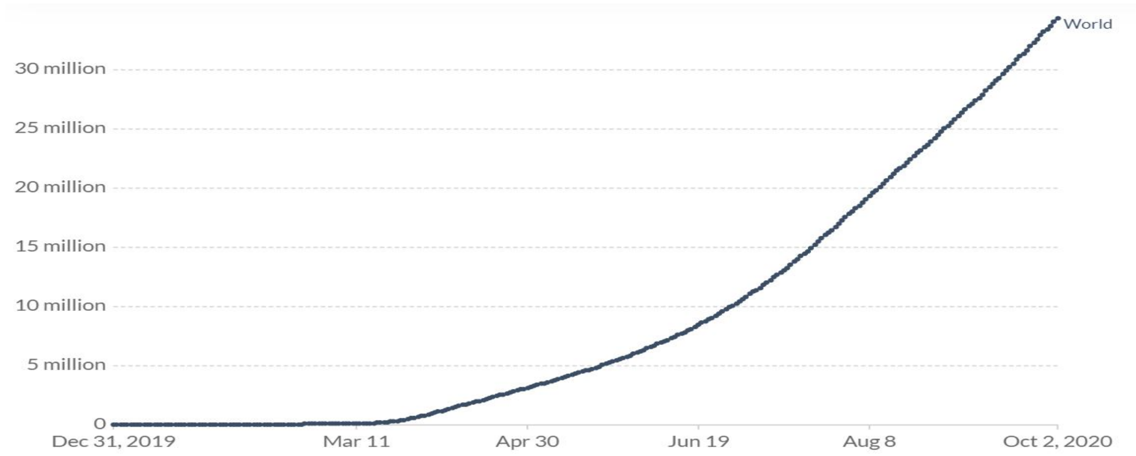

FIGURE 2.3: Cumulative confirmed COVID-19 cases (Disease Prevention and Control, 2020)

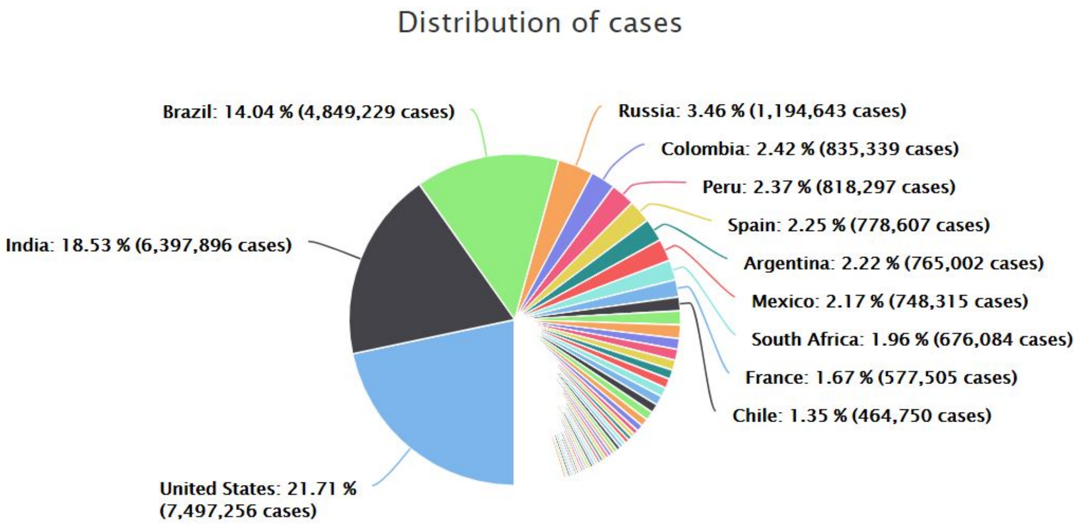

FIGURE 2.4: Country wise distribution of COVID-19 cases (*Coronavirus Cases Distribution*)

As we have now seen the major trends and events that took place in the rapid spread of COVID-19, a more detailed discussion about the imaging features of the novel coronavirus is presented in the following section.

2.2 Diagnosing COVID-19

The following section discusses in detail the traditional procedure used in detecting COVID-19 using the real-time RT-PCR test but more importantly gives an insight on how medical imagery can be used to achieve the same by showcasing the specific patterns and lesions observed from the lung scans of patients diagnosed with COVID-19.

2.2.1 The RT-PCR Test

The real-time RT-PCR test is a nuclear-derived method that detects the presence of specific genetic material in any pathogen which includes a virus. Scientists and medical practitioners would be able to see the results even when the process is ongoing using the real-time RT-PCR test whereas the conventional RT-PCR provides the result only after the process is complete.

The real-time RT-PCR test has been widely used to detect other viruses such as Ebola and Zika, and therefore is extensively used to detect COVID-19 in laboratories throughout the world (Jawerth, 2020). The real-time RT-PCR test has been recommended by WHO for COVID-19 diagnosis.

'Reverse Transcription' involves converting Ribonucleic Acid (RNA) to Deoxyribonucleic acid (DNA), the main reason for this being the ability to amplify specific parts of DNA which allows the scientists to spot strands of the virus among genetic information. Samples are collected from the patient's throat or nose, typically where the COVID-19 virus tend to gather (Jawerth, 2020).

Scientists add short DNA fragments, which complements the viral DNA. Therefore the virus, if present, leads to these added fragments to be attached to the target sections of the viral DNA. When marker labels attach to these DNA strands, a fluorescent dye is released, which is measured by the RT-PCR machine. When a certain threshold of fluorescence have passed the scientists could then diagnose the patient with COVID-19 (Jawerth, 2020).

Despite the high sensitivity and reliable diagnosis by the real-time RT-PCR test, there exists certain limitations which have led researchers to identify alternate methods of COVID-19 diagnosis, such as using medical imagery which includes X-Ray and CT scans.

During these trying times where the numbers keep rising rapidly, medical facilities are running short of RT-PCR test kits and therefore, are in dire need of alternate sources of diagnosis. Furthermore, the high false-negative rate of the real-time RT-PCR, which is as high as 100% before the time of symptom onset and decreases only up to 64% on the day of symptom onset (Kucirka LM and J, 2020) could lead to severe consequences in the real world where the infected person could spread the virus as they are not under quarantine.

Another important factor considering the daily rising numbers includes the delays in receiving the results for the RT-PCR test which places extra strain on the medical professionals in terms of workload and also makes it difficult for them to apply quarantine protocols on suspected patients.

Therefore, due to these aforementioned limitations of the real-time RT-PCR test, finding a safer, more accurate, and faster diagnosis mechanism is essential. Thus, making COVID-19 diagnosis using medical imagery an ideal alternative candidate.

Bibliography

Coronavirus Cases Distribution. URL: https://www.worldometers.info/coronavirus/.

Coronavirus (COVID-19). URL: https://news.google.com/covid19/map?hl=en-US&gl=US&ceid=US:en.

Daily COVID-19 Tests (2020). URL: https://ourworldindata.org/grapher/full-list-covid-19-tests-per-day?time=2020-02-20..latest (visited on 09/20/2020).

Disease Prevention, European Centre for and Control (2020). *The daily number of new reported cases of COVID-19 by country worldwide*. URL: https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide (visited on 10/02/2020).

Edwards, Erika (2020a). *Coronavirus U.S. case confirmation: 1st coronavirus case is in Seattle*. URL: https://www.nbcnews.com/health/health-news/1st-case-coronavirus-china-confirmed-u-s-n1119486 (visited on 09/28/2020).

– (2020b). *New coronavirus outbreak: WHO declares it a global public health emergency*. URL: https://www.nbcnews.com/health/health-news/new-coronavirus-outbreak-declared-global-public-health-emergency-who-says-n1126836.

Givetash, Linda (2020). *U.S. surpasses China with coronavirus cases as global total tops 500,000*. URL: https://www.nbcnews.com/news/world/u-s-surpasses-china-coronavirus-cases-global-total-tops-500-n1170136 (visited on 10/02/2020).

Jason (2020). *WHO declares COVID-19 disease to be a pandemic*. URL: https://www.nbcnews.com/health/health-news/live-blog/coronavirus-updates-live-u-s-cases-top-1-000-spread-n1155241/ncrd1155646#blogHeader (visited on 10/02/2020).

Jawerth, Nicole (2020). *How is the COVID-19 Virus Detected using Real Time RT-PCR?* URL: https://www.iaea.org/newscenter/news/how-is-the-covid-19-virus-detected-using-real-time-rt-pcr (visited on 10/02/2020).

Kucirka LM Lauer SA, Laeyendecker O Boon D and Lessler J (2020). "Variation in False-Negative Rate of Reverse Transcriptase Polymerase Chain Reaction–Based SARS-CoV-2 Tests by Time Since Exposure". In: *Annals of Internal Medicine* 173.4. PMID: 32422057, pp. 262–267. DOI: 10.7326/M20-1495. eprint: https://doi.org/10.7326/M20-1495. URL: https://doi.org/10.7326/M20-1495.

Muccari, Robin, Denise Chow, and Joe Murphy (2020). *Coronavirus timeline: Tracking the critical moments of COVID-19*. URL: https://www.nbcnews.com/health/health-news/coronavirus-timeline-tracking-critical-moments-covid-19-n1154341 (visited on 09/28/2020).

Phillips, Morgan (2020). *Coronavirus cases top 1 million worldwide, researchers say*. URL: https://www.foxnews.com/health/coronavirus-cases-top-1-million-worldwide-johns-hopkins (visited on 10/02/2020).

Shabad, Rebecca (2020). *Senate passes $8.3 billion emergency bill to combat coronavirus*. URL: https://www.nbcnews.com/politics/congress/senate-passes-8-3-billion-emergency-bill-combat-coronavirus-n1150521 (visited on 10/02/2020).

White, Martha C. and Lucy Bayly (2020). *Dow closes down 1,000 points as coronavirus fears slam Wall Street*. URL: https://www.nbcnews.com/business/markets/dow-plunges-950-points-fears-coronavirus-will-tank-global-economic-n1141546 (visited on 09/28/2020).

World Health Organization (2020a). *Coronavirus*. URL: https://www.who.int/health-topics/coronavirus (visited on 09/20/2020).

– (2020b). *Situation Report - 1*. Novel Coronavirus (2019-nCoV). URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200121-sitrep-1-2019-ncov.pdf?sfvrsn=20a99c10_4.

Wuhan Municipal Health Commission (2020). *Wellness Commission on the current pneumonia outbreak in our city*. URL: https://web.archive.org/web/20200109215413/http://wjw.wuhan.gov.cn/front/web/showDetail/2019123108989 (visited on 09/28/2020).