A Training Report

On

**DETECTION OF COVID-19 BY X-RAYS**

Submitted in partial fulfilment of requirements for the award of the

Degree of

**Bachelor of Technology**

In

**Computer Science & Engineering**

**Submitted By**

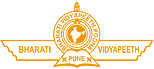
**ACHYUT KRISHNA, YASH VARSHNEY & PIYUSH ANAND**

**(41011502818, 41611502818 & 42011502818)**

**Under the guidance of**

**DR. RACHNA JAIN, NITIKA SHARMA, PREETI NAGRATH & ASHISH GUPTA**

**(PROFESSOR)**



**Department of Computer Science & Engineering**

**Bharati Vidyapeeths College of Engineering**

**A-4, Paschim Vihar, New Delhi-110063**

**June, 2020**

**Certificate**

I hereby certify that the work which is being submitted in this report titled **“Detection Of COVID-19 By X-rays”**, in partial fulfilment of the requirement for the award of certification of “In-House Summer Training in Machine Learning and Deep Learning” submitted in Bharati Vidyapeeth’s College of Engineering, New Delhi, is an authentic record of my own work carried out under the supervision of “Dr. Rachna Jain” and refers to other researchers work which are duly listed in the reference section.

The matter presented in this report has not been submitted for the award of any other certificate of this or any other institution.

**(Achyut Krishna, Yash Varshney & Piyush Anand)**

Roll No. 41011502818, 41611502818 & 42011502818

Registration. No.

This is to certify that the statements made above by the candidate are correct and true to the best of our knowledge.

|  |  |
| --- | --- |
| **DR. RACHNA JAIN**  ASSISTANCE PROFESSOR  Computer Science & Engineering  BVCOE  New Delhi - 110063 |  |

The Viva-Voce Examination of \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ has been held on \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

Internal Examiner External Examiner

**CANDIDATES DECLARATION**

I hereby declare that the work presented in this report entitled “**Detection COVID-19 by X-rays**”, in partial fulfilment of the requirement for the award of the degree **Bachelor of Technology** and submitted in **Department** **of** **Computer Science & Engineering,** **Bharati Vidyapeeth College of Engineering**, , **New Delhi** (**Affiliated to Guru Gobind Singh Indraprastha University**) is an authentic record of my own work carried out during the period from June July 2019 under the guidance of **Dr. Rachna Jain, Nitika Sharma, Preeti Nagrath & Ashish Gupta, Professor.**

The work reported in this has not been submitted by me for award of any other degree of this or any other institute.

**(Achyut Krishna, Yash Varshney & Piyush Anand)**

**(En. No: 41011502818, 41611502818 & 42011502818)**

**ACKNOWLEDGEMENT**

I express my deep gratitude to **Dr. Rachna Jain, Nitika Sharma, Preeti Nagrath & Ashish Gupta, Professor**, Bharti Vidyapeeth College of Engineering, for his/her valuable guidance and suggestion throughout my training**.** We are thankful to **Rubeena mam & Shifali mam** (ECE, 2020, II) for their valuable guidance.

**Sign**

(**Achyut Krishna, Yash Varshney & Piyush Anand)**

**Enrolment No:41011502818, 41611502818 & 42011502818**

**ORGANIZATION INTRODUCTION**

Bharati Vidyapeeth College of Engineering (BVCOE) is Private Institute located in New Delhi and was established in the year 1999. The institute is affiliated to Guru Gobind Singh Indraprastha University (GGSIPU), New Delhi and is approved by All India Council for Technical Education (AICTE). BVCOE offers four-year full time B.Tech. course in five specializations namely: Computer Science & Engineering, Electrical & Electronics Engineering, Electronics & Communication Engineering, Instrumentation & Control Engineering and Information Technology.

**Table of Content**

CERTIFICATE…………………………………………..……………....ii

CANDIDATE DECLARATION…..…………………………………….iii

ACKNOWLEDGEMENT…………………………………………….....iv

ORGANIZATION INTRODUCTION…………………………………...v

TABLE OF CONTENTS………………………………………….……..vi

PREFACE………………………………………………………………..vii

LIST OF FIGURES……………………………………………………..viii

LIST OF TABLES……………………………………………………….ix

Chapter 1: Introduction

1.1 CORONA Virus……………………………………..….x

1.2 COVID-19 Outbreak & Detection Process…………..…x

1.3 Machine Learning & Deep Learning………………..…xv

Chapter 2: Project

2.1 Project Objective………………………………………xix

2.2 Project Methodology….……………………………….xix

2.3 Project Results…………………………………………xx

Chapter 3: Literature Review

3.1 Review……………………………………………….. xxii

Chapter 4: Final remarks

3.1 Conclusion……………………………………………..xxiv

References…………………………………………………………….xxv

**Preface**

This report is prepared to fulfil the requirement of the B.Tech program of **Bharti Vidyapeeth College of Engineering** on “Detection of COVID-19 By X-Rays”. I have chosen this topic because its is a very important solution for Corona Virus detection using X-rays. It is affordable so that it could be available for everyone. It I can be widely used as X ray machines are available in large numbers comparatively to other means of testing.

The prime focus of this study is to highlight and analyse the infection among patients of corona virus in large numbers. It can detect with better results and the outcome can be useful to determine the intensity of infection inside the body too. All this information was basically secondary data.

This report is divided into several sections. In the first section, it is the introductory part. It is about the Corona Virus, the outbreak and the detection process. Then how can we use this machine learning & deep learning concepts to build a model for better detection of the virus. In the second section, it is discussed the objective and importance of this project, Research methodology, the project structure and the results obtained from the project. In the last section, the report discusses the conclusion and summary of the project. At the last the references are given for more information for information and research purposes.

—————————————  
Achyut Krishna

Yash Varshney

Piyush Anand

Enrolment no: 41011502818, 41611502818 & 42011502818

**LIST OF FIGURES**

1. Fig 1: [Transmission electron micrograph](https://en.wikipedia.org/wiki/Transmission_electron_micrograph) of [avian infectious bronchitis virus](https://en.wikipedia.org/wiki/Avian_infectious_bronchitis_virus) ..x

2. Fig 2: Test kit…………………………………………………………………... xi

3. Fig 3: Testing kit………………………………………………………………xii

4. Fig 4: Neural network basic structure…………………………………………xvii

5. Fig 5: Different types of neural networks…………………………………….... xviii

6. Fig 6: Survival Graph…………………………………………………………... xx

7. Fig 7: IEEE model……………………………………………………………..xxi

8. Fig 8: Kaggle model……………………………………………………………xxi

9. Fig 9: IEEE Heatmap…………………………………………………………... xxi

10. Fig 9: Kaggle Heatmap………………………………………………………. xxi

**LIST OF TABLES**

1. Table 1: Weekly Progress…………………………... xxiii

2. Table 2: Comparison with other methods ……………… xxiv

**CHAPTER 1: INTRODUCTION**

**1.1 CORONA Virus**

**Coronaviruses** are a group of related [RNA viruses](https://en.wikipedia.org/wiki/RNA_virus) that cause diseases in [mammals](https://en.wikipedia.org/wiki/Mammals) and [birds](https://en.wikipedia.org/wiki/Birds). In humans, these [viruses](https://en.wikipedia.org/wiki/Virus) cause [respiratory tract infections](https://en.wikipedia.org/wiki/Respiratory_tract_infection) that can range from mild to lethal. Mild illnesses include some cases of the [common cold](https://en.wikipedia.org/wiki/Common_cold) (which is also caused by other viruses, predominantly [rhinoviruses](https://en.wikipedia.org/wiki/Rhinovirus)), while more lethal varieties can cause [SARS](https://en.wikipedia.org/wiki/Severe_acute_respiratory_syndrome), [MERS](https://en.wikipedia.org/wiki/Middle_East_respiratory_syndrome), and [COVID-19](https://en.wikipedia.org/wiki/Coronavirus_disease_2019). Symptoms in other species vary: in chickens, they cause an [upper respiratory tract disease](https://en.wikipedia.org/wiki/Upper_respiratory_tract_infection), while in cows and pigs they cause [diarrhea](https://en.wikipedia.org/wiki/Diarrhea" \o "Diarrhea). There are as yet no [vaccines](https://en.wikipedia.org/wiki/Vaccine) or [antiviral drugs](https://en.wikipedia.org/wiki/Antiviral_drug) to prevent or treat human coronavirus infections. Coronaviruses constitute the [subfamily](https://en.wikipedia.org/wiki/Subfamily) ***Orthocoronavirinae***, in the family *[Coronaviridae](https://en.wikipedia.org/wiki/Coronaviridae" \o "Coronaviridae)*, order *[Nidovirales](https://en.wikipedia.org/wiki/Nidovirales" \o "Nidovirales)*, and realm *[Riboviria](https://en.wikipedia.org/wiki/Riboviria" \o "Riboviria)*.

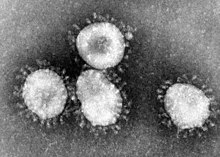


Fig 1: [Transmission electron micrograph](https://en.wikipedia.org/wiki/Transmission_electron_micrograph) of [avian infectious bronchitis virus](https://en.wikipedia.org/wiki/Avian_infectious_bronchitis_virus) [<https://en.wikipedia.org/wiki/Coronavirus>]

They are [enveloped viruses](https://en.wikipedia.org/wiki/Enveloped_virus) with a [positive-sense single-stranded](https://en.wikipedia.org/wiki/Positive-sense_single-stranded_RNA_virus) [RNA](https://en.wikipedia.org/wiki/RNA) [genome](https://en.wikipedia.org/wiki/Genome) and a [nucleocapsid](https://en.wikipedia.org/wiki/Nucleocapsid) of helical symmetry. The [genome size](https://en.wikipedia.org/wiki/Genome_size) of coronaviruses ranges from approximately 26 to 32 [kilobases](https://en.wikipedia.org/wiki/Kilobase#Length_measurements), one of the largest among [RNA viruses](https://en.wikipedia.org/wiki/RNA_virus). They have characteristic club-shaped [spikes](https://en.wikipedia.org/wiki/Peplomer) that project from their surface, which in [electron micrographs](https://en.wikipedia.org/wiki/Micrograph) create an image reminiscent of the [solar corona](https://en.wikipedia.org/wiki/Stellar_corona), from which their name derives.

**1.2 COVID-19 Outbreak & Detection Process**

**Testing**

Experts say ramping up the use of these three different types of coronavirus tests -- the PCR test, the antibody test and the antigen test, respectively -- is [essential to reopening America's shuttered economy safely](https://www.cnn.com/2020/04/21/health/millions-of-tests-report/index.html). Despite the need, the availability of testing remains in frustratingly short supply and is [dependent on individual states and local health care providers](https://www.cnn.com/2020/04/27/politics/white-house-testing-blueprint/index.html) rather than the federal government. And increasing that availability of testing is difficult because of significant scientific and logistical issues.

* + 1. **PCR TEST (SWAB TEST)**

**What are they?**

[Polymerase chain reaction tests, known as PCR, are the most common and most accurate tests](https://www.cnn.com/2020/03/04/health/coronavirus-test-what-happens-explainer/index.html) for determining whether someone is currently infected with the novel coronavirus.

**How do they work?**

Every virus has a unique genetic code. The novel coronavirus PCR test works by starting with a throat swab or sputum sample from a patient, amplifying that specimen in a machine, and then looking for unique coronavirus genetic material.

[[](https://edition.cnn.com/2020/03/04/health/coronavirus-test-what-happens-explainer/index.html)](https://edition.cnn.com/2020/03/04/health/coronavirus-test-what-happens-explainer/index.html)

Fig2: Test kit [<https://edition.cnn.com/2020/04/28/us/coronavirus-testing-pcr-antigen-antibody/index.html>]

[What actually happens during a coronavirus test?](https://edition.cnn.com/2020/03/04/health/coronavirus-test-what-happens-explainer/index.html)

First, a health care worker takes a swab or sputum sample from the back of the patient's throat or nose, generally using a long thin nasopharyngeal swab. That swab is stored in a sterile tube or vial and is then sent to a lab that has access to a test kit.

There, trained lab technicians extract any genetic information from the specimen. The purified genetic material is then mixed with materials, including some derived from the coronavirus itself, that are known as reagents.

The entire solution is then placed into a testing machine about the size of a toaster. If a patient's specimen contains coronavirus, then the virus's genetic material will be amplified, and the machine will return a positive result. And if the specimen has no coronavirus, there will be a negative result.

The whole process generally can be measured in days, though the time to get a result is shorter if hospitals or clinics have test kits in-house.

**Why is this test important?**

The PCR test is the most widespread and most accurate diagnostic test for determining whether someone is currently infected with coronavirus. In the United States, nearly 1 million people have tested positive for coronavirus using this PCR test.

Health officials have consistently said that testing is needed to understand where the virus is spreading -- and how to stop it.

People who test positive for the virus can quarantine and isolate themselves and keep it from spreading to others. This is particularly important with Covid-19 because [some infected people have no symptoms](https://www.cnn.com/2020/03/14/health/coronavirus-asymptomatic-spread/index.html) and, without a positive test, do not even know they are spreading the virus.

**What are the holdups?**

First, the Centres for Disease Control and Prevention's [initial rollout of the PCR test](https://www.cnn.com/2020/04/20/politics/coronavirus-testing-trump-administration-response-invs/index.html) was fraught with problems that delayed the US's early response. That issue has since been fixed but led to a delay in testing early on.

Now, the primary holdup is simply one of supply and demand.

The PCR test requires specialized supplies, expensive instruments, and the expertise of [trained lab technicians](https://www.rasmussen.edu/degrees/health-sciences/blog/what-does-a-medical-lab-tech-do/). Because of the sudden mass demand for these tests and corresponding issues in the supply chain, there have been shortages at almost every step of the process, including of [swabs](https://www.cnn.com/2020/03/28/politics/coronavirus-swabs-supplies-shortage-states/index.html) and [reagents](https://www.cnn.com/2020/04/01/politics/testing-backlog-coronavirus-quest-invs/index.html).

"We all want all these things," said Dr. April Abbott, the director of microbiology for Deaconess Health System in Indiana. "Just managing that has been an extreme struggle."

Abbott said that as time has gone on, it has been easier for her hospital system to get swabs and reagents. Overall, though, the PCR test requires expertise and specialized supplies in a way that makes it difficult to ramp up production so immediately, especially as the [federal government has pushed responsibility onto individual states](https://www.cnn.com/2020/04/27/politics/white-house-testing-blueprint/index.html).

"This comes down to a large logistical problem," said Dr. Donald Thea, professor of global health at Boston University.

* + 1. **ANTIBODY TEST (BLOOD TEST)**

**What are they?**

[Antibody tests, also known as serology tests](https://www.cnn.com/2020/04/14/health/antibody-test-explainer/index.html), do not detect the virus itself.

Instead, they detect whether someone has the antibodies in their immune system to fight off the novel coronavirus. People develop antibodies when their body fights off the virus.

**How do they work?**

[](https://edition.cnn.com/2020/04/14/health/antibody-test-explainer/index.html)

Fig 3: Testing kit [<https://en.wikipedia.org/wiki/COVID-19_testing>]

[What are antibody tests and what do they mean for the coronavirus pandemic?](https://edition.cnn.com/2020/04/14/health/antibody-test-explainer/index.html)

Antibody tests are performed on a blood specimen. This can be from a finger prick or taken from a vein. The majority of antibody tests then require a trained expert to examine that blood specimen for antibodies specifically targeted to coronavirus.

Our bloodstream has a library of antibodies against various viral, bacterial and fungal infections that we've had over our lifetimes, said CNN Medical Analyst Dr. Kent Sepkowitz, an infectious disease specialist. The goal of this test is to find the antibody that specifically applies to the novel coronavirus that causes Covid-19.

However, there are [seven different coronaviruses that cause illness in humans](https://www.niaid.nih.gov/diseases-conditions/coronaviruses), including several that cause mild symptoms like the common cold. Our antibodies for each coronavirus appear similar, so deciphering between the common cold antibody and the Covid-19 antibody is a real challenge scientifically, Sepkowitz said.

**Why is this test important?**

Antibody tests are generally not used for diagnosis but to give a sense of how widely the virus has spread in a population. Given that PCR testing was in short supply in the early weeks of the coronavirus outbreak, wider antibody testing would help us know how many people recovered from coronavirus, possibly without even knowing they were infected.

In addition, most experts believe that people who have recovered from the virus will have some sort of protection against reinfection, so antibody testing could be used to decide who is immune to the virus and who may be able to return to work without fear of infection.

However, because the virus is still so new, it's unclear how strong that immunity might be or how long it might last.

"We expect that most people who are infected with #COVID19 will develop an antibody response that will provide some level of protection," the World Health Organization wrote in a series of tweets Saturday.

"What we don't yet know is the level of protection or how long it will last. We are working with scientists around the world to better understand the body's response to #COVID19 infection. So far, no studies have answered these important questions."

**What are the holdups?**

The primary issue with antibody testing is the question of accuracy. It's not yet clear if these tests are consistently correct.

In March, the Food and Drug Administration issued a policy to allow developers of antibody tests to go to market [without prior FDA review](https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-serological-tests) if they met certain conditions. The FDA also has authorized eight antibody tests under an Emergency Use Authorization, which means the FDA has not validated the accuracy of the tests.

This has led to a glut of unreliable tests and uncertainty among health officials tasked with getting them.

"That rigorous level of study just hasn't happened with the majority of antibody tests that are available," Abbott said.

Inaccurate results could also have deadly consequences. A false positive on an antibody test could lead someone to think they are immune to the virus, which could be more dangerous than no result at all.

"We really don't want to purchase something that may give inaccurate information," Abbott said. "It would be worse to give results that would be inaccurate. We want to make sure that we're very confident in the results we're giving."

* + 1. **ANTIGEN TEST**

**What are they?**

[Antigen tests are a quick and easy way](http://www.cnn.com/2020/04/27/health/antigen-tests-coronavirus-breakthrough/index.html) to look for a specific antigen -- a term for any foreign substance, like a virus or bacteria. Antigen tests for flu or strep throat, for example, can be done in a doctor's office without expensive equipment and give results in minutes.

There is not yet a reliable antigen test for the novel coronavirus, but it is theoretically possible and companies are working to make one.

**How do they work?**

Think of an at-home pregnancy test. In a coronavirus antigen test, you put a bodily sample such as a throat swab onto a specially treated strip, which then uses a color or marker to say whether the novel coronavirus is present.

The antigen test works by looking for a unique part of the novel coronavirus, such as a specific protein on one of [its namesake "corona" spikes](https://www.cnn.com/2020/03/31/health/what-is-coronavirus-covid-19-wellness/index.html). If that specific protein is present in detectable quantities, then the test returns a positive result.

**Why is this test important?**

Antigen testing could provide immediate answers to whether someone is infected with coronavirus and could in theory be mass produced for home use.

These likely would not replace PCR tests, which are the most reliable, but they would be a good first step. The WHO said these "could potentially be used as triage tests to rapidly identify patients who are very likely to have COVID-19, reducing or eliminating the need for expensive molecular confirmatory testing."

Dr. Deborah Birx, the White House coronavirus response coordinator, said the [US will need a breakthrough in antigen tests](https://www.cnn.com/2020/04/26/politics/coronavirus-testing-deborah-birx/index.html) to increase testing to sufficient levels.

"There will never be the ability on a nucleic acid test to do 300 million tests a day or to test everybody before they go to work or to school," Birx said earlier this month. "But there might be with the antigen test."

**What are the holdups?**

So far, antigen tests for coronavirus don't work consistently. Based on similar antigen tests for respiratory diseases like the flu, their sensitivity -- the ability to correctly identify patients with Covid-19 -- ranges between 34% and 80%, according to WHO.

"Based on this information, half or more of COVID-19 infected patients might be missed by such tests, depending on the group of patients tested," the agency said.

Designing the tests can be challenging, according to Gigi Gronvall, a senior scholar at the Johns Hopkins Center for Health Security.

"Sometimes viruses may have special folds, or protein modifications on their surface, which can interfere with the process," Gronvall said. In addition, while PCR tests amplify the virus's genetic material to detect even small amounts, antigen tests may need a larger amount of virus.

"And then, on top of the design part, you need to have enough of the virus in you to test positive," she added. Because of these ongoing scientific issues, the antigen test is promising but remains several steps away, according to Sepkowitz.

**But, These tests have some limitations**

1. One is widespread but has key shortages across its supply chain
2. One remains questionably reliable and is only helpful after the fact.
3. And one is still early in development -- but could speed up and simplify results.
   1. **Machine Learning & Deep Learning**

**1.3.1 MACHINE LEARNING & ITS ROLE IN DETECTION**

Machine learning (ML) is the study of computer algorithms that improve automatically through experience. It is seen as a subset of artificial intelligence. Machine learning algorithms build a mathematical model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to do so.

**1.3.1.1 NUMPY**

NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays. The ancestor of NumPy, Numeric, was originally created by Jim Hugunin with contributions from several other developers. In 2005, Travis Oliphant created NumPy by incorporating features of the competing Numarray into Numeric, with extensive modifications. NumPy is open-source software and has many contributors.

**1.3.1.2 MATPLOTLIB**

Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK+. There is also a procedural "pylab" interface based on a state machine, designed to closely resemble that of MATLAB, though its use is discouraged. SciPy makes use of Matplotlib.

**1.3.2 How to use machine learning for detection purpose?**

Evidence has shown that COVID-19 can have a negative impact on the cardiovascular system, leaving patients at risk for adverse events such as heart failure, sustained abnormal heartbeats, heart attacks, and death. Because of the increased risk for these complications, there is a significant need to identify COVID-19 patients at high risk for heart problems, but these predictive capabilities don’t currently exist.

With this grant, researchers will aim to develop these capabilities using [**machine learning**](https://healthitanalytics.com/news/machine-learning-models-estimate-seasonal-impact-of-covid-19).

“This project will provide clinicians with early warning signs and ensure that resources are allocated to patients with the greatest need,” said Natalia Trayanova, the Murray B. Sachs Professor in the Department of Biomedical Engineering at The Johns Hopkins University Schools of Engineering and Medicine and the project’s principal investigator.

The first phase of the one-year project just received IRB approval for Suburban Hospital and Sibley Memorial Hospital within the Johns Hopkins Health System (JHHS).

In this first phase, researchers will collect data from more than 300 COVID-19 patients admitted to JHHS, including cardiac-specific laboratory tests, continuously-obtained vital signs, and imaging data like CT scans echocardiography. The team will use this data to train the machine learning algorithm.

Researchers will then test the algorithm using data from COVID-19 patients with heart injury at JHHS, other hospitals nearby, and maybe some in New York City. The overarching goal is to create a predictive risk score that can determine which patients are at high risk of developing adverse cardiac events up to 24 hours ahead of time. For new patients, the model will perform a baseline prediction that is updated each time new health data becomes available.

According to the research team, this will be the first approach to predict COVID-19-related cardiovascular outcomes. While similar studies exist, previous research has focused on predictions of [**general COVID-19 mortality**](https://healthitanalytics.com/news/data-driven-model-predicts-number-of-covid-19-deaths-in-us) or a patient’s need for ICU care.

This new machine learning approach will analyze [**multiple sources of data**](https://healthitanalytics.com/news/how-will-big-data-analytics-factor-into-the-next-phase-of-covid-19) to produce a risk score that is continually updated as researchers acquire new data.

The project will also help providers understand how COVID-19-related heart injury could lead to heart dysfunction and sudden cardiac death. The study will also help clinicians determine which biomarkers are most predictive of adverse clinical outcomes. After creating and testing the algorithm, researchers will make the tool widely available for healthcare institutions to implement.

“As a clinician, major knowledge gaps exist in the ideal approach to risk stratify COVID-19 patients for new heart problems that are common and may be life-threatening. These patients have varying clinical presentations and a very unpredictable hospital course,” said Allison G. Hays, Associate Professor of Medicine in the Johns Hopkins University School of Medicine’s Division of Cardiology and the project’s clinical collaborator.

“This project aims to help clinicians quickly risk stratify patients using real time clinical data, with the goal of widely disseminating this knowledge to help medical practitioners around the world in their approach to treating and monitoring patients suffering from COVID-19.”

This project will help researchers obtain information critical to fighting COVID-19.

“By predicting who’s at risk for developing the worst outcomes, healthcare professionals will be able to undertake the best routes of therapy or primary prevention and save lives,” said Trayanova.

**1.3.2 DEEP LEARNING & ITS ROLE IN THE DETECTION**

**Deep learning** (also known as **deep structured learning**) is part of a broader family of [machine learning](https://en.wikipedia.org/wiki/Machine_learning) methods based on [artificial neural networks](https://en.wikipedia.org/wiki/Artificial_neural_networks) with [representation learning](https://en.wikipedia.org/wiki/Representation_learning). Learning can be [supervised](https://en.wikipedia.org/wiki/Supervised_learning), [semi-supervised](https://en.wikipedia.org/wiki/Semi-supervised_learning) or [unsupervised](https://en.wikipedia.org/wiki/Unsupervised_learning).

Deep learning architectures such as [deep neural networks](https://en.wikipedia.org/wiki/Deep_learning#Deep_neural_networks), [deep belief networks](https://en.wikipedia.org/wiki/Deep_belief_network), [recurrent neural networks](https://en.wikipedia.org/wiki/Recurrent_neural_networks) and [convolutional neural networks](https://en.wikipedia.org/wiki/Convolutional_neural_networks) have been applied to fields including [computer vision](https://en.wikipedia.org/wiki/Computer_vision), [speech recognition](https://en.wikipedia.org/wiki/Automatic_speech_recognition), [natural language processing](https://en.wikipedia.org/wiki/Natural_language_processing), [audio recognition](https://en.wikipedia.org/wiki/Audio_recognition), social network filtering, [machine translation](https://en.wikipedia.org/wiki/Machine_translation), [bioinformatics](https://en.wikipedia.org/wiki/Bioinformatics), [drug design](https://en.wikipedia.org/wiki/Drug_design), medical image analysis, material inspection and [board game](https://en.wikipedia.org/wiki/Board_game) programs, where they have produced results comparable to and in some cases surpassing human expert performance.

**1.3.2.1 KERAS**

Keras is an open-source neural-network library written in Python. It is capable of running on top of TensorFlow, Microsoft Cognitive Toolkit, R, Theano, or PlaidML. Designed to enable fast experimentation with deep neural networks, it focuses on being user-friendly, modular, and extensible. It was developed as part of the research effort of project ONEIROS, and its primary author and maintainer is François Chollet, a Google engineer. Chollet also is the author of the XCeption deep neural network model.

**1.3.2.2 ARTIFICIAL NEURAL NETWORKS**

An artificial neural network is a system of hardware or software that is patterned after the working of neurons in the human brain and nervous system. Artificial neural networks are a variety of deep learning technology which comes under the broad domain of Artificial Intelligence.

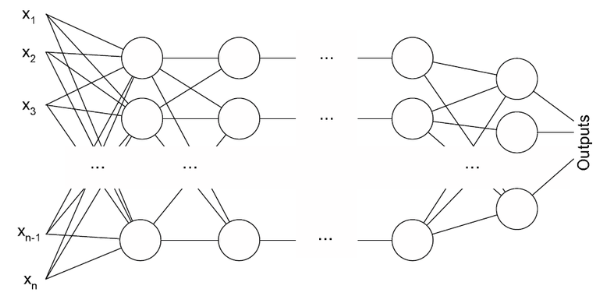


Fig 4: Neural network basic structure [<https://www.digitalvidya.com/blog/types-of-neural-networks/#:~:text=A%20Comprehensive%20Guide%20to%20Types,Neural%20Network.%207%20Sequence-To-Sequence%20Models.>]

**1.3.2.3 CNN**

In [deep learning](https://en.wikipedia.org/wiki/Deep_learning), a convolutional neural network (CNN, or ConvNet) is a class of [deep neural networks](https://en.wikipedia.org/wiki/Deep_neural_network), most commonly applied to analysing visual imagery. They are also known as shift invariant or space invariant artificial neural networks (SIANN), based on their shared-weights architecture and [translation invariance](https://en.wikipedia.org/wiki/Translation_invariance) characteristics. They have applications in [image and video recognition](https://en.wikipedia.org/wiki/Computer_vision), [recommender systems](https://en.wikipedia.org/wiki/Recommender_system), [image classification](https://en.wikipedia.org/wiki/Image_classification), [medical image analysis](https://en.wikipedia.org/wiki/Medical_image_computing), [natural language processing](https://en.wikipedia.org/wiki/Natural_language_processing), and financial [time series](https://en.wikipedia.org/wiki/Time_series).

**1.3.2.4 RNN**

A **recurrent neural network** (**RNN**) is a class of [artificial neural networks](https://en.wikipedia.org/wiki/Artificial_neural_network) where connections between nodes form a [directed graph](https://en.wikipedia.org/wiki/Directed_graph) along a temporal sequence. This allows it to exhibit temporal dynamic behavior. Derived from [feedforward neural networks](https://en.wikipedia.org/wiki/Feedforward_neural_networks), RNNs can use their internal state (memory) to process variable length sequences of inputs. This makes them applicable to tasks such as unsegmented, connected [handwriting recognition](https://en.wikipedia.org/wiki/Handwriting_recognition) or [speech recognition](https://en.wikipedia.org/wiki/Speech_recognition)**.**

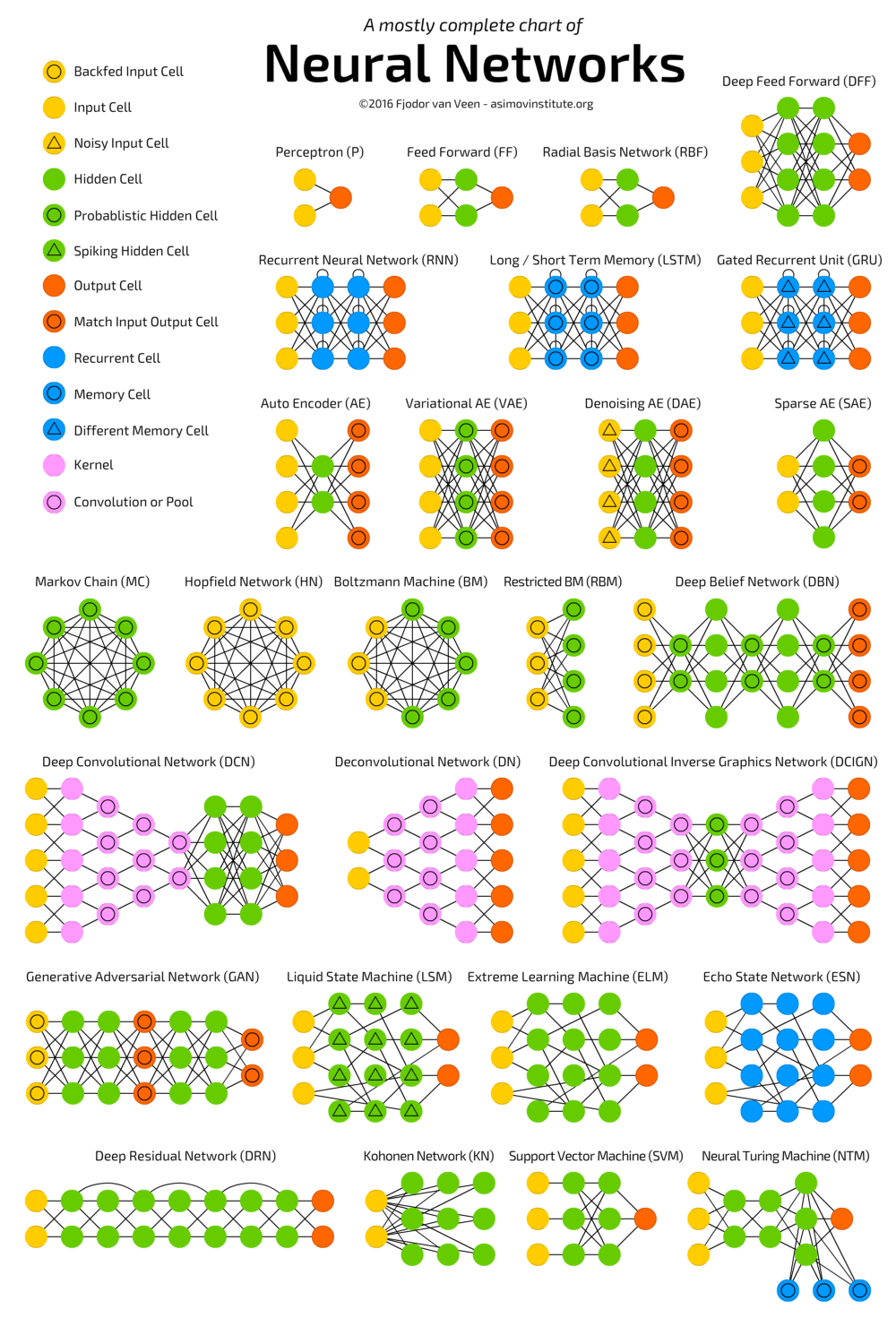


Fig 5: Different types of neural networks[<https://www.digitalvidya.com/blog/types-of-neural-networks/#:~:text=A%20Comprehensive%20Guide%20to%20Types,Neural%20Network.%207%20Sequence-To-Sequence%20Models.>]

**CHAPTER 2: PROJECT**

**2.1 PROJECT OBJECTIVE**

This project aims how we can detect COVID-19 by using X-rays by using Deep learning models for image processing like CNN (Convolutional Neural Network) with very high accuracy. This will also collect the initial COVID-19 open image data collection. It was created by assembling medical images from websites and publications. This project proposes a novel method for the detection of COVID-19 from chest X-ray using CNN(Convolutional Neural Network).

Uses of X-ray method:

1. Blood Tests are Costly and Takes Time To conduct ~ 6-7 hours per patient.
2. Extent of Spread can be detected.
3. X-rays and scans are easier and cost-efficient also the spread for COVID-19 virus in the body can also be detected.

**2.2 PROJECT METHODOLOGY**

# **Step 1 - Visualize our image data and split it into training and test sets**

The dataset has been split into 'train' and 'test'' folders. Additionally, the images within these folders have been split into 'NORMAL' and 'PNEUMONIA' subfolders.

First, we will go through the folders of the dataset and append each image into an array. Both the training and test set images will be placed into the same matrices for now (later to be randomly split into our training and test sets). Note that each image will be resized to 64x64 to reduce computation time and to maintain consistency throughout our dataset. Because this is a binary classification problem, each image will have an associated classification vector of the form (1,0), meaning the patient has pneumonia, or (0,1), meaning the patient does not have pneumonia. After our images have been pulled and classified, they will be normalized. This is done to ensure that our gradients do not diverge during back propagation.

Finally, our data will be split into training and test sets. The training set will use 2/3 of the available X-ray images.

# **Step 2 - Normalizing and splitting our data**

We can see our data is biased to pneumonia positive results. It will be important to check the distribution of our training and test sets after we randomly shuffle and create them. In addition, to ensure our gradients do not diverge, we will normalize our image data.

# **Step 3 - Define our model**

Now that our variables, forward propagation and cost function have been defined, we can bring this all together to create our convolutional neural network model. We will start by setting up the structure of our network, and then coding a mini-batch optimization loop. We will log the costs as this loop progresses to ensure the algorithm is behaving as intended (these costs will be plotted against iterations) The following steps will be taken to set up the structure of our network:

* Call and assign out placeholder values with the appropriate parameters
* Initialize our weight parameters
* Assign our forward propagation to a variable
* Assign our cost function to a variable
* Define our optimization algorithm (we will be using an Adam Optimizer for this network)

# **Step 4 – Analysing our results**

Our results are looking good. In addition to our overall excellent training and test prediction accuracy, both our positive and negative predictions independently have beyond acceptable accuracy. Looking at our cost vs. iterations graph, we see a general downward trend which seems to flatten out. This suggests we have chosen an appropriate learning rate and number of iterations to balance weight optimization time and accuracy. We have created a convolutional neural network that can be used to predict pneumonia from X-ray images!

**2.3 PROJECT RESULTS**

**Plots Based on Model Accuracy**

The pyplot is a collection of command style functions that make matplotlib work like MATLAB. Each pyplot function makes some change to a figure: e.g., creates a figure, creates a plotting area in a figure, plots some lines in a plotting area, decorates the plot with labels, etc.

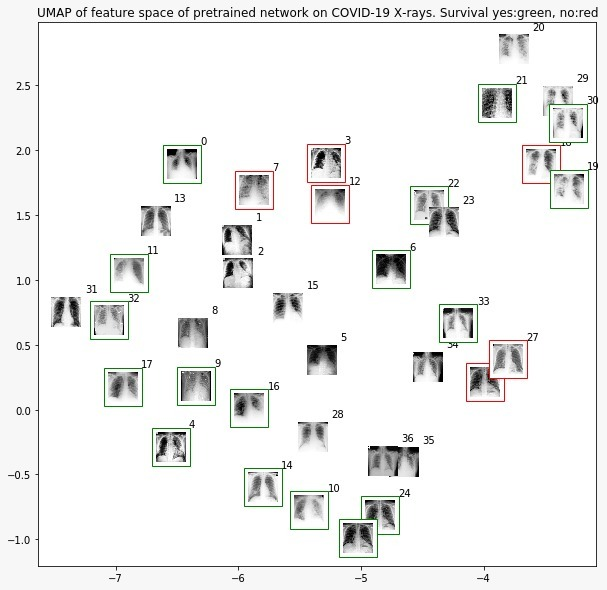


Fig 6: Survival Graph

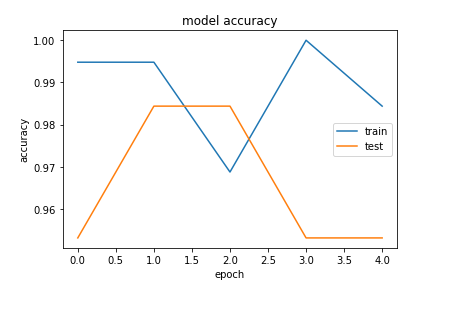
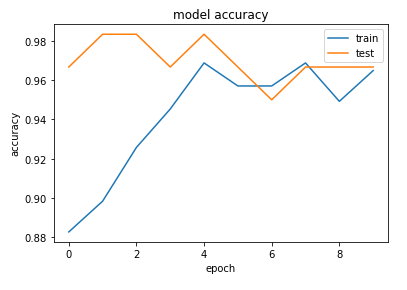
 

Fig 7: IEEE model (accuracy- 97.52 %) Fig 8:Kaggle Model (accuracy-96.66%)

Heat Maps

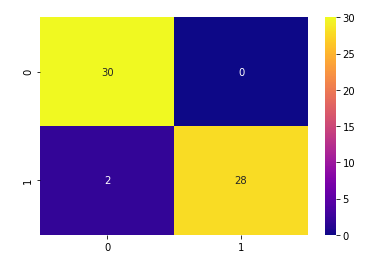
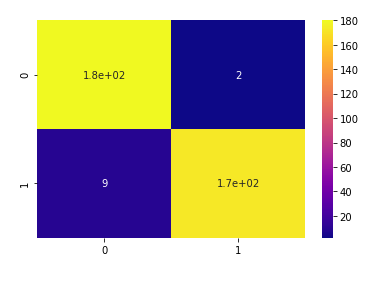


Fig 9: IEEE Heatmap Fig 10: Kaggle Heatmap

**Weekly progress**

|  |  |  |
| --- | --- | --- |
| Weeks | Topics | Subtopics |
| I | Basics of Python | Functions, Numpy, Pandas, Matplotlib |
| II | Machine Learning | Regression, NLP, SVM, OpenCV |
| III | Deep Learning | Decision Tree, K-means Clustering, Neural networks, Tensorflow, Keras, CNN, RNN |
| IV | Project |  |

Table 1: Weekly Progress

**CHAPTER 3: LITERATURE REVIEW**

1. **COVID-19 Image Data Collection**

By: Joseph Paul Cohen, Paul Morrison & Lan Dao

Summary:

This paper describes the initial COVID-19 open image data collection. It was created by assembling medical images from websites and publications and currently contains 123 frontal view X-rays. In this paper, they describe the public database of pneumonia cases with chest X-ray or CT images, specifically COVID19 cases as well as MERS, SARS, and ARDS. Data will be collected from public sources in order not to infringe patient confidentiality.

1. **Deep Neural Network Models for Paraphrased Text Classification in the Arabic Language**

By: Adnen MahmoudMounir Zrigui

Summary:

Paraphrase is the act of reusing original texts without proper citation of the source. Different obfuscation operations can be employed such as addition/deletion of words, synonym substitutions, lexical changes, active to passive switching, etc. This phenomenon dramatically increased because of the progressive advancement of the web and the automatic text editing tools. Recently, deep leaning methods have gained competitive results than traditional methods for Natural Language Processing (NLP). In this context, we consider the problem of Arabic paraphrase detection. We present different deep neural networks like Convolutional Neural Network (CNN) and Long Short term Memory (LSTM). Our aim is to study the effective of each one in extracting the proper features of sentences without the knowledge of semantic and syntactic structure of Arabic language.

1. **COVID-CT-Dataset: A CT Image Dataset about COVID-19**

By: Jinyu Zhao, Xuehai He, Xingyi yang, Yichen zhang & Shanghang Zhang

Summary:

Computed tomography (CT) is a useful manner for diagnosing COVID-19 patients. To mitigate the lack of publicly available COVID-19 CT images for developing CT-based diagnosis deep learning models of COVID-19, we build an open-sourced dataset COVID-CT, which contains 349 COVID-19 CT images from 216 patients and 463 non-COVID-19 CTs. The utility of this dataset is confirmed by a senior radiologist and via experimental studies. Using this dataset, we develop a joint classification and segmentation method that achieves an F1 of 0.85, an AUC of 0.95, and an accuracy of 0.83.

1. **COVID-DA: Deep Domain Adaptation from Typical Pneumonia to COVID-19**

By: Yifan Zhang, Shuaicheng Niu, Zhen Qiu, Ying Wei, Peilin Zhao, Jianhua Yao, Junzhou Huang, Qingyao Wu, and Mingkui Tan

Summary:

The outbreak of novel coronavirus disease 2019 (COVID-19) has already infected millions of people and is still rapidly spreading all over the globe. Most COVID-19 patients suffer from lung infection, so one important diagnostic method is to screen chest radiography images, e.g., X-Ray or CT images. However, such examinations are time-consuming and laborintensive, leading to limited diagnostic efficiency. To solve this issue, AI-based technologies, such as deep learning, have been used recently as effective computer-aided means to improve diagnostic efficiency. However, one practical and critical difficulty is the limited availability of annotated COVID-19 data, due to the prohibitive annotation costs and urgent work of doctors to fight against the pandemic. This makes the learning of deep diagnosis models very challenging. To address this, motivated by that typical pneumonia has similar characteristics with COVID19 and many pneumonia datasets are publicly available, we propose to conduct domain knowledge adaptation from typical pneumonia to COVID-19. There are two main challenges: 1) the discrepancy of data distributions between domains; 2) the task difference between the diagnosis of typical pneumonia and COVID-19. To address them, we propose a new deep domain adaptation method for COVID-19 diagnosis, namely COVIDDA. Specifically, we alleviate the domain discrepancy via feature adversarial adaptation and handle the task difference issue via a novel classifier separation scheme. In this way, COVID-DA is able to diagnose COVID-19 effectively with only a small number of COVID-19 annotations. Extensive experiments verify the effectiveness of COVID-DA and its great potential for realworld applications.

# **A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images**

By: Linda Wang Zhong Qiu Lin and Alexander Wong

Summary:

The COVID-19 pandemic continues to have a devastating effect on the health and well-being of the global population. A critical step in the fight against COVID-19 is effective screening of infected patients, with one of the key screening approaches being radiology examination using chest radiography. It was found in early studies that patients present abnormalities in chest radiography images that are characteristic of those infected with COVID-19. Motivated by this and inspired by the open source efforts of the research community, in this study we introduce COVID-Net, a deep convolutional neural network design tailored for the detection of COVID-19 cases from chest X-ray (CXR) images that is open source and available to the general public. To the best of the authors’ knowledge, COVID-Net is one of the first open source network designs for COVID-19 detection from CXR images at the time of initial release. We also introduce COVIDx, an open access benchmark dataset that we generated comprising of 13,975 CXR images across 13,870 patient patient cases, with the largest number of publicly available COVID-19 positive cases to the best of the authors’ knowledge. Furthermore, we investigate how COVID-Net makes predictions using an explainability method in an attempt to not only gain deeper insights into critical factors associated with COVID cases, which can aid clinicians in improved screening, but also audit COVID-Net in a responsible and transparent manner to validate that it is making decisions based on relevant information from the CXR images. By no means a production-ready solution, the hope is that the open access COVID-Net, along with the description on constructing the open source COVIDx dataset, will be leveraged and build upon by both researchers and citizen data scientists alike to accelerate the development of highly accurate yet practical deep learning solutions for detecting COVID-19 cases and accelerate treatment of those who need it the most.

**CHAPTER 4: CONCLUSION**

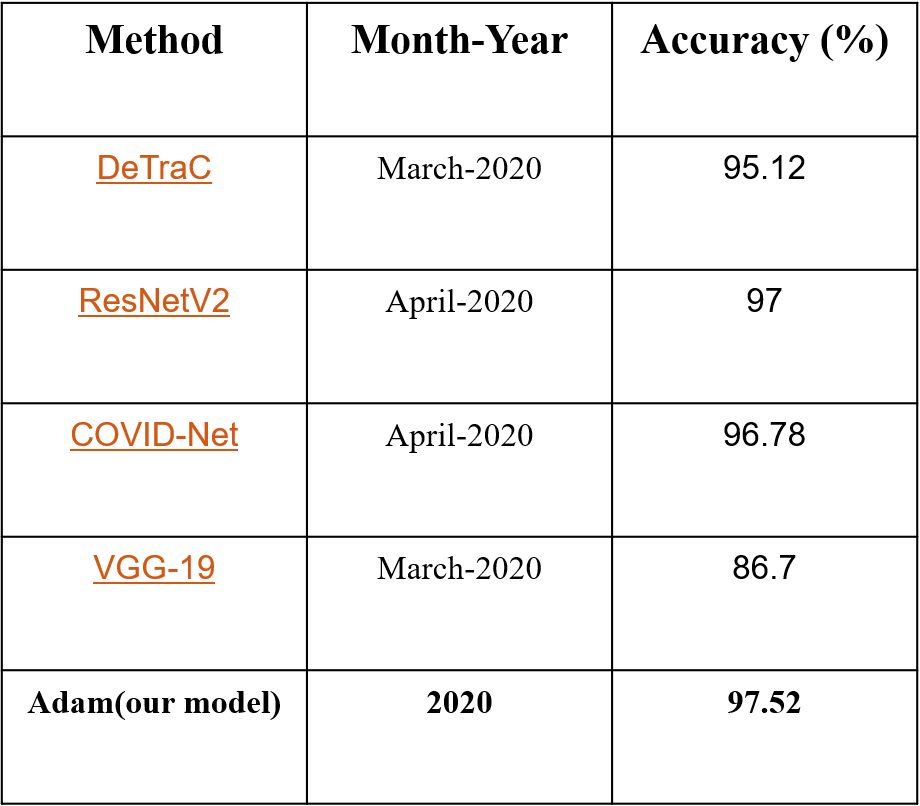


Table 2: Comparison with other methods

**Remarks:**

The Validation accuracy for Kaggle model is **96.66%** and IEEE model is **97.52%**.

We can conclude that our Model for IEEE Dataset is much better than that of other models that other are using for KAGGLE.

**References**

[1] Wang, W.et al.Detection of SARS-CoV-2 in different types of clinical specimens.JAMA(2020). IEEE.​

 [2] Fang, Y.et al.Sensitivity of chest CT for covid-19: Comparison to RT-PCR.Radiology200432 (2020).​

[3] X. Du, Y. Cai, S. Wang and L. Zhang, "Overview of deep learning," 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC), Wuhan, 2016, pp. 159-164, doi: 10.1109/YAC.2016.7804882.​

[4] A. Shrestha and A. Mahmood, "Review of Deep Learning Algorithms and Architectures," in IEEE Access, vol. 7, pp. 53040-53065, 2019, doi: 10.1109/ACCESS.2019.2912200.​

[5] X. Zhou, W. Gong, W. Fu and F. Du, "Application of deep learning in object detection," 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), Wuhan, 2017, pp. 631-634, doi: 10.1109/ICIS.2017.7960069.​

[6] S. Albawi, T. A. Mohammed and S. Al-Zawi, "Understanding of a convolutional neural network," 2017 International Conference on Engineering and Technology (ICET), Antalya, 2017, pp. 1-6, doi: 10.1109/ICEngTechnol.2017.8308186. ​

[7] R. Chauhan, K. K. Ghanshala and R. C. Joshi, "Convolutional Neural Network (CNN) for Image Detection and Recognition," 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), Jalandhar, India, 2018, pp. 278-282, doi: 10.1109/ICSCCC.2018.8703316.​

[8] G. Zhu, B. Li, S. Hong and B. Mao, "Texture Recognition and Classification Based on Deep Learning," 2018 Sixth International Conference on Advanced Cloud and Big Data (CBD), Lanzhou, 2018, pp. 344-348, doi: 10.1109/CBD.2018.00068.​

[9] Z. Chen and Y. Zhou, "Research on Automatic Essay Scoring of Composition Based on CNN and OR," 2019 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD), Chengdu, China, 2019, pp. 13-18, doi: 10.1109/ICAIBD.2019.8837007.​

[10] A. Imran, I. Posokhova, H. N. Qureshi, U. Masood, S. Riaz, K. Ali,C. N. John, M. Nabeel, and I. Hussain, ‘‘AI4COVID-19: AI enabled pre-liminary diagnosis for COVID-19 from cough samples via an app,’’ 2020,arXiv:2004.01275.

[11] IEEE Covid Chest X-Ray Dataset. Accessed: Mar. 7, 2020. [Online].Available: <https://github.com/ieee8023/covid-chestxray-dataset> . ​

[12]  L. Wang and A. Wong, ‘‘COVID-Net: A tailored deep convolutional neuralnetwork design for detection of COVID-19 cases from chest radiographyimages,’’ 2020,arXiv:2003.09871. https://arxiv.org/abs/2003.09871.​

[13] L. Zhong, L. Mu, J. Li, J. Wang, Z. Yin, and D. Liu, ‘‘Early predictionof the 2019 novel coronavirus outbreak in the mainland China basedon simple mathematical model,’’IEEE Access, vol. 8, pp. 51761–51769,2020.​

[14]  A. Beers, J. Brown, K. Chang, J. P. Campbell, S. Ostmo, M. F. Chiang,and J. Kalpathy-Cramer, ‘‘High-resolution medical image synthesisusing progressively grown generative adversarial networks,’’ 2018,arXiv:1805.03144. [Online]. Available: <http://arxiv.org/abs/1805.03144>. ​

[15] W. Dai, J. Doyle, X. Liang, H. Zhang, N. Dong, Y. Li, and E. P. Xing,‘‘SCAN: Structure correcting adversarial network for chest X-raysorgan segmentation,’’ 2017,arXiv:1703.08770. [Online]. Available: <https://arxiv.org/abs/1703.08770> .​

[16] S. Mondal, K. Agarwal and M. Rashid, "Deep Learning Approach for Automatic Classification of X-Ray Images using Convolutional Neural Network," 2019 Fifth International Conference on Image Information Processing (ICIIP), Shimla, India, 2019, pp. 326-331, doi: 10.1109/ICIIP47207.2019.8985687 .​

[17] E. Ayan and H. M. Ünver, "Diagnosis of Pneumonia from Chest X-Ray Images Using Deep Learning," 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT), Istanbul, Turkey, 2019, pp. 1-5, doi: 10.1109/EBBT.2019.8741582.​

[18]  M. Bouchahma, S. Ben Hammouda, S. Kouki, M. Alshemaili and K. Samara, "An Automatic Dental Decay Treatment Prediction using a Deep Convolutional Neural Network on X-Ray Images," 2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA), Abu Dhabi, United Arab Emirates, 2019, pp. 1-4, doi: 10.1109/AICCSA47632.2019.9035278 .​

[19] A. Wibisono, J. Adibah, F. S. Priatmadji, N. Z. Viderisa, A. Husna and P. Mursanto, "Segmentation-based Knowledge Extraction from Chest X-ray Images," 2019 4th Asia-Pacific Conference on Intelligent Robot Systems (ACIRS), Nagoya, Japan, 2019, pp. 225-230, doi: 10.1109/ACIRS.2019.8935951.​

[20] X. Ren et al., "Regression Convolutional Neural Network for Automated Pediatric Bone Age Assessment From Hand Radiograph," in IEEE Journal of Biomedical and Health Informatics, vol. 23, no. 5, pp. 2030-2038, Sept. 2019, doi: 10.1109/JBHI.2018.2876916.​

[21] N. Bonettini, M. Paracchini, P. Bestagini, M. Marcon and S. Tubaro, "Hyperspectral X-ray Denoising: Model-Based and Data-Driven Solutions," 2019 27th European Signal Processing Conference (EUSIPCO), A Coruna, Spain, 2019, pp. 1-5, doi: 10.23919/EUSIPCO.2019.8903151.

**APPENDIX**

1. FILTERIZATION OF DATSETS

<https://colab.research.google.com/drive/1D9f1ccaSr2SO9HrfTfT1voRbqP36ej9m#scrollTo=5OY2M40Za9cL>

1. TRAINING & TEST OF DATA

<https://www.kaggle.com/creatorpiyush/covid19-compare>