|  |  |
| --- | --- |
|  | An Automated Approach to detect Covid-19 from CT-Scan Images. |
|  | Supervisor: |
|  | Submitted By:  Project Proposal  7th December 2022 |

# Table of Contents

[Table of Contents 1](#_Toc117360601)

[1) Abstract: 2](#_Toc117360602)

[2) Aim & Objectives: 3](#_Toc117360603)

[a) Background: 3](#_Toc117360604)

[3) Introduction: 3](#_Toc117360605)

[a) Objectives: 4](#_Toc117360606)

[b) Ethical considerations: 4](#_Toc117360607)

[4) Research questions 4](#_Toc117360608)

[a) Research Contributions 4](#_Toc117360610)

[b) Dataset Used 5](#_Toc117360611)

[5) Brief Literature of Review: 6](#_Toc117360612)

[a) DL for the diagnosis of COVID-19 7](#_Toc117360615)

[6) Research Methodology: 8](#_Toc117360616)

[7) Project Schedule: 10](#_Toc117360617)

[8) Conclusion: 13](#_Toc117360618)

9) Bibliography & References: 13

1. Abstract:

The COVID-19 pandemic has hit humanity hard in 233 different countries and so far 6,583,217 people have died and this death toll is still going on (World meter, 2022). The escalating prevalence of COVID-19 and the associated serious casualties are placing severe pressure on the system's woefully inadequate health resources. The inefficiency and lack of testing is a major obstacle to controlling the spread of this disease. Most current tests are based on reverse transcription polymerase chain reaction (RT-PCR). Results take 4-6 hours, which is a long time compared to the rapid spread of COVID-19. In addition to inefficiency, RT-PCR test kits are in short supply. As a result, many infected people are not identified in time and continue to unknowingly infect others.

Effective advances are needed to improve clinical productivity in healthcare and to simplify the diagnosis, care and monitoring of COVID-19. Since it can scale more successfully, process information faster, and actually bypass humans in some health-related jobs, later research has shown that artificial intelligence may be a promising breakthrough. Using machine learning and deep learning approaches, researchers have been encouraged to identify and classify new strains of COVID-19. This thesis will present a comprehensive study and analysis of the methodology of detection of coronavirus (COVID-19) from CT-Scan images using various deep learning and machine learning techniques.

The proposed method will help the radiologist in deciding on the detection and diagnosis of cases infected with COVID-19 with high accuracy and efficiency in a shorter period of time. We expect that the use of machine learning to predict COVID-19 will reduce the time delay in the diagnosis of medical tests and empower health professionals to provide appropriate medical care to patients. The proposed work will be used to evaluate the most optimal detection and classification technique for COVID-19 in terms of accuracy and real-time detection. The performance of our model will be evaluated based on the confusion matrix and accuracy, sensitivity, and specificity metrics. This project will highlight the limitations of the COVID-19 dataset along with the challenges this work faces and attempt to come up with future research directions.

Key Terms: CNN (Convolutional Neural Networks), CT Scan Image Classification, Covid-19, Chest X-ray, Unsupervised Learning, Transfer Learning

# Aim & Objectives:

This dissertation aims to develop a computer-aided machine learning model that can be used for early diagnosis of COVID-19 from CT scans and X-ray samples, which may be useful for the treatment of COVID-19.

We will be able to use deep learning and machine learning techniques like CNN to effectively classify images in the project, which will be particularly useful in pushing the use of machine learning to solve many of the world's most pressing problems, which is quite significant. This work will serve as a reference for future research work as well as facilitate the design of a structure for a very deep learning model that will facilitate the development of future research.

## Background:

In the current situation, COVID-19 has become a major issue affecting the population of the earth, shutting down various economic classes and numerous problems and challenges affecting everyone. For those who complain of its symptoms, accurate detection and diagnosis of the disease Covid-19 is essential. The researcher for this work found that there is a significant error rate in the diagnosis of the disease. In order to reduce the above percentage of errors, this circumstance must be optimized for the application of machine learning algorithms, especially deep learning.

Much effort has been devoted to finding alternative testing methods to mitigate the inefficiencies and shortages of existing tests for COVID-19.

Several studies have shown that computed tomography (CT) scans reveal clear radiological findings in patients with COVID-19 and hold promise as a more efficient and affordable testing method due to the widespread availability of CT equipment that can rapidly generate results. In addition, it reduces the burden on healthcare professionals when reading CT images, while several studies have developed deep learning methods to automatically interpret CT images and predict whether they are positive for COVID-19. Although these works have produced promising results, they are limited in two ways. First, for privacy reasons, the CT scan datasets used in these works are not publicly available. As a result, their findings cannot be replicated and the trained models cannot be used in other hospitals.

In addition, the lack of an open source annotated COVID-19 CT dataset severely limits the research and development of more advanced artificial intelligence methods for more accurate CT-based testing of COVID-19. Second, to achieve clinical-level performance, these works require a large CT collection during model training. In practice, such a requirement is strict and many hospitals may not meet it, especially considering that health professionals are overloaded with COVID-19 patients and are unlikely to have time to collect and annotate large numbers of COVID-19 CT scans. .

The coronavirus has spread rapidly around the world in the past two years. While the cause of this outbreak is unknown, research is still being done to find a cure. Due to the daily increase in cases, testing for the coronavirus is not possible for time and financial reasons. Machine learning has become increasingly reliable in the medical industry in recent years. Using machine learning to predict COVID-19 in patients will speed up the processing time of test results and direct medical staff to treat patients as needed. The success of this work will help us open a new horizon to use ML/DL tools to diagnose disease like epidemic in the future and will renovate how the world will look at any upcoming pandemic and how to treat it and how to generate AI-based systems to find and disease detection and subsequent cure.Introduction:

An introductory chapter describing the sufficient objectives for the project as well as the problem statement.

Millions of people have died from COVID-19, which is regarded to be the deadliest and most disruptive virus of the twenty-first century and is caused by the severe acute respiratory syndrome coronavirus 2, in less than two years (SARS-CoV-2). Early diagnosis is essential in the aforementioned scenario to ensure that patients receive the proper care while easing the burden on the healthcare system. COVID-19 is still a fatal disease because there are no early detection methods available anywhere. Computer scientists are called upon to lend a hand in various situations.

Machine learning techniques have demonstrated the effectiveness and superior performance of artificial intelligence in automated image classification problems, and it is currently being utilised to automate the diagnosis of a number of diseases.

Deep learning is a group of machine learning methods that largely focus on automatically extracting and classifying image features. Deep learning has shown great promise in a number of applications, particularly in the field of health care. For the detection and classification of Covid-19 syndrome, several AI techniques have so far been applied to find the solution for radiology and medical image processing. Among these methods, two well-known deep-learning-based networks are convolutional neural networks (CNNs) and recurrent neural networks (RNNs). Our work will help the community in finding the following objectives defined

## Objectives:

• Development of an automatic machine learning algorithm that could use CT scan images to accurately predict COVID-19 in patients.

• Determining the best predictive machine learning technique to assist radiologists and other medical professionals in their research and decision making.

• Identification of difficulties associated with the use of ML and DL algorithms to detect epidemic viruses

• Suggest future actions and recommendations for radiologists and data analysts to ensure effective analysis.

• Analyzing and comparing the results of machine learning and deep learning techniques for the detection of COVID-19.

The article begins with an overview of recent work done in ML/DL toward real-world applications that could help the community. The next goal will be to develop a methodology using DL that will use different stages of preprocessing and later we will discuss the result of my model by comparing it with current models.

Our goal is to build efficient on-sample deep learning methods using a publicly available dataset source that contains hundreds of positive COVID-19 CT scans, allowing us to obtain excellent diagnostic accuracy of COVID-19 from CT scans even with a limited training supply. CT scans. . To obtain a robust and objective representation of features while reducing the danger of overfitting, we propose a Self-Trans technique that combines self-supervised contrastive learning and transfer learning. Numerous tests show that our proposed Self-Trans method performs better than a number of high-end baselines. Our technique achieves an F1 of 0.85 and an AUC score of 0.91 despite having only a small number of training CTs. This is explained in the next document.

Graphical user interface, text

Description automatically generated

Figure: Result showing 0.91 AUC score obtained from our own trained model

The current version of my model is a very simple convolutional neural network. This model consists of only two convolutional layers, a maximum pooling layer, two more convolutional layers, a second maximum pooling layer, and a modest feedforward network.

We'll introduce some complexity later.

Finally, we will try to come up with some overview and future research possibilities that we learn from this work.

## Ethical considerations:

As the research would determine methods for classifying and detecting CT images from lung images, it will therefore not examine the areas of race, religion and ethnicity.

A few other ethical points to consider throughout the dissertation process are listed below:

1. Respecting privacy according to Birmingham City University (BCU) rules,

2. De-identify information that violates any privacy.

3. The data collected must reflect the exact purpose of the research.

# Pre-processing Overview

We begin by assembling the COVID-19-CT dataset, which contains 349 CT images with clinical findings from 216 cases of patients with COVID-19.

We are developing deep learning (DL) methods to perform CT-based diagnosis of COVID-19 using this dataset. Although the largest of its kind, the COVID-19-CT has a limited number of images. When trained on small datasets, DL models are data intensive and have a high risk of overfitting.

All projects should have direction, organization and also a method showing an educated approach to how this project should be completed. It is important that the methodologies constantly reflect the work to date and also describe the deadlines that are necessary for the project to be completed within the time frame specified in the terms of reference for this project.

Below is a summary of the contributions of this work:

• Development and improvement of machine learning algorithms for identification of COVID-19.

• Using deep learning and machine learning to select features to reduce false positives and false negatives.

• Improved binary and multiple classification accuracy.

## Research Contributions

We begin by assembling the COVID-19-CT dataset, which contains 349 CT images with clinical findings from 216 cases of patients with COVID-19.

We are developing deep learning (DL) methods to perform CT-based diagnosis of COVID-19 using this dataset. Although the largest of its kind, the COVID-19-CT has a limited number of images. When trained on small datasets, DL models are data intensive and have a high risk of overfitting.

All projects should have direction, organization and also a method showing an educated approach to how this project should be completed. It is important that the methodologies constantly reflect the work to date and also describe the deadlines that are necessary for the project to be completed within the time frame specified in the terms of reference for this project.

Below is a summary of the contributions of this work:

• Development and improvement of machine learning algorithms for identification of COVID-19.

• Using deep learning and machine learning to select features to reduce false positives and false negatives.

• Improved binary and multiple classification accuracy.

## Dataset Used

CT scans show promise in providing accurate, rapid and inexpensive covid-19 screening and testing.

These statistics come from surveys of real patients in hospitals in Sao Paulo, Brazil. This dataset aims to support the study and development of artificial intelligence techniques that can determine whether a person has SARS-CoV-2 by examining their CT scans.

<https://www.kaggle.com/datasets/plameneduardo/sarscov2-ctscan-dataset>

The utility of the following data set was confirmed by a senior radiologist at Tongji Hospital, Wuhan, China, who performed the diagnosis and treatment of a large number of patients with COVID-19 during the outbreak of the disease between January and April.

<https://github.com/UCSD-AI4H/COVID-CT>

# Brief Literature Review:

We begin by constructing the COVID-19-CT dataset, which contains 349 CT images with clinical findings from 216 cases of patients with COVID-19.

We are developing deep learning (DL) methods to perform CT-based diagnosis of COVID-19 using this dataset. Although the largest of its kind, the COVID-19-CT has a limited number of images. When trained on small data sets, DL models are data intensive and have a high risk of overfitting.

All projects should have direction, organization and also a method showing an educated approach to how the project is to be completed. It is important that the methodologies consistently reflect the work to date and also describe the deadlines that are necessary to complete the project within the time frame specified in the terms of reference for this project.

Below is a summary of the contributions of this work:

• Development and improvement of machine learning algorithms to identify COVID-19.

• Using deep learning and machine learning to select features to reduce false positives and false negatives.

• Improved binary and multiple classification accuracy.

## DL for the diagnosis of COVID-19

Since the outbreak of COVID-19, efforts have been increasing to develop deep learning methods to perform screening for COVID-19 based on medical images such as CT scans and chest X-rays. Wu et al. developed an early screening model based on multiple CNNs to classify CT scans of patients with COVID-19. A CT scan contains more information than corresponding X-rays. These methods take input image datasets, perform various preprocessing, segmentation, and optimization, and then perform the task of classification and detection. These methods like CNN, DCNN, ResNet, Alexnet, V-net, VB-net and U-net etc. are used for the task of binary, ternary and even for classification of four class problems.

The use of artificial intelligence techniques has consistently produced reliable and trustworthy results from applications that use image data. Research recently used deep learning to detect COVID-19 by examining and analyzing chest X-rays.

• In their work (Das et al., 2020), the authors proposed a method based on Deep-CNN to identify patients positive for COVID-19 from chest X-rays. They used DenseNet201, Resnet50V2 and Inceptionv3 CNN models. They used publicly available datasets that contained 1,006 chest X-ray images, divided into 771 training images and 235 test images, because chest X-rays are a quick and cheap way to identify people with COVID-19. This dataset contained 538 images of infected patients and 468 images of uninfected individuals. According to the data, the test setup had a classification accuracy of 95.7% and a sensitivity of 98%. In addition, they developed graphical user interface (GUI) software for public use. Patients positive for COVID-19 can be detected using images from chest X-rays within seconds. X-ray changes in CT scans of patients with COVID-19 in China were identified by Shuai Wang et al. In this study, he developed COVID-19 as an alternative diagnostic method using deep learning techniques to extract graphical aspects from CT images. They collected CT scans of patients diagnosed with pneumonia as well as patients with confirmed COVID-19. The results of their research offer a proof of concept for the application of artificial intelligence to accurately predict COVID-19. Unlike our research, which makes predictions using clinical features and laboratory data, this study analyzes CT images.

• Epidemiological, demographic, clinical, laboratory, radiological, and therapeutic data from Zhongnan Hospital, Wuhan, China, were reported in this study by Dawei Wang et al. Information was evaluated and recorded to track infections. The author provides a clearer view of radiological and therapeutic data that we can use in our model to predict COVID-19.

• Ali Narin et al. created an automated detection system for diagnosing COVID-19. According to this paper, “three alternative models based on the convolutional neural network ResNet50, InceptionV3 and Inception-ResNetV2 were developed for the detection of patients infected with coronavirus pneumonia using chest X-ray images.

The research mentioned above has produced a number of prediction techniques using CT scan images and symptoms to predict the COVID-19 epidemic, mortality risks, and mortality risk in different countries. According to available information, there is not much evidence for prediction systems using clinical data. In this work, COVID-19 will be predicted using machine learning algorithms and clinical data from patients with COVID-19.

# Research Methodology:

During the investigations of the exploitation of machine learning methods, the methods I am interested will be to use the version of CNN or RNN to perform the task of ternary classification using the dataset available publicly either on GitHub or Kaggle.

After obtaining the dataset, the task is then to inquire into possible libraries and tools that may be required for the project. This is because, once the datasets have been obtained and the methods decided upon from the research undertaken, it is then required to implement the methods upon the explored dataset. This would then allow for the project to be viable as the datasets have been found and obtained, the algorithms identified from the research done of the most common and most effective of them. And then finally, to be able to apply it to the dataset, using a programming tool such as python.

Furthermore, once these key stages of the project have been achieved the tasks then would be to provide recommendations based on the results achieved.

# Project Schedule:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Start Date | End Date | Duration (Days) |
| Selecting and Searching Dataset from CT-scan Images | 03/10/2022 | 09/10/2022 | 6 |
| Research/Experiment with possible libraries could be used | 03/10/2022 | 09/10/2022 | 7 |
| Studying different research articles on the detection of COVID-19 using ML/DML | 03/10/2022 | 09/10/2022 | 7 |
| Make a prototype | 17/10/2022 |  | 81 |
| Training our own model |  |  | 9 |
| Evaluations and experimentation |  |  | 9 |
| Collate the results into tables | 17/11/2022 |  | 8 |
| Propose recommendations | 17/122022 |  | 5 |
| Write up the project Dissertation report | 17/12/2022 |  | 30 |
| Complete the project presentation | 13/12/2022 | 10/12/2022 | 4 |

This tables informs of the key milestones and tasks that would be undertaken in order to carry this project out. Further explanations of the millstones have been explained further below.

It was required to first find the dataset that would be needed to carry out the tasks. For this it was essential to look for datasets that would be related to the project in this case. By doing so the dataset would be there to apply the methods on. After this it was needed to find out which methods of ML and DL would be used in order to find ideal methods and research would be needed to design our own model to be used in COVID-19 detection and classification and then to consider them to be applied to the obtained datasets. We got our data set in form of CT scan images.  We aim for Accuracy of our model to classify CT scans into Covid Positive or Covid Negative. We want to create a neural network to classify CT scan images into Covid Positive or Covid Negative. To do so, we need to calculate a classification score for each image.

The first stage would be data gathering and investigation, such as gathering data from various sources and determining whether it is useable; we got our image data samples. the second step would be pre-processing the data to convert the data into a format that can be used to train/evaluate a model; the third step would be modelling experimentation; we'll try different neural networks (different DL approaches) and see what works and what doesn't; and the fourth step would be to use the model.

Projects have deadlines, and meeting deadlines is critical to success. The project's key deliverables and their scope of completion are clearly outlined in the project timeline. The main resource for this research project is time. (Minimize resource overload by effectively allocating time to different tasks. Avoiding resource overload reduces the chance of quality degradation

Gantt charts are powerful time management tools. Figure 1 shows the structure of the Gantt chart created for this project. On the left side of the screen is a chronological list of tasks. Estimated completion time is displayed along the x-axis. If the schedule is met, the project will be completed on time and with high quality.

# Conclusion and Future Work

The recent COVID-19 pandemic has caused enormous damage to the social economy and people's health. Major challenges in the fight against COVID-19 include early and accurate diagnosis and the lack of ventilators for critically ill patients.

Artificial intelligence is receiving a lot of attention in the medical sector, especially for improving therapeutic and healthcare capabilities. The use of virtual assistants has also become popular in the healthcare industry. A virtual assistant can be found in a variety of applications, including medical applications such as voice recognition and information retrieval for cognitively impaired patients. In a recent survey conducted by IBM Watson Health, 43 percent of respondents said AI will disrupt healthcare rather than save it.

This article provides a systematic and comprehensive overview of the applications of artificial intelligence in the form of ML and DL in the context of COVID-19.

The work will highlight the latest applications of AI-based techniques to classify, diagnose and even predict Covid-19. State-of-the-art pre-processing techniques will be applied to real-life CT scan images to denoise and remove distortion from these images. The paper will try to find a dataset that can be labeled as a benchmark for evaluating ML/DL techniques. There are many datasets where there are CT images containing images of different stages of covid and a few images are unlabeled. Labeling these images will also be a challenge to be solved logically using unsupervised labeling methods.

Computer trust is a contentious part of computer science, not specifically because its applicability varies in different contexts, but because it is necessary to secure and protect networks from potential threat actors. Therefore, computational trust is an incredibly active area of ​​research, and future work in this topic area is necessary to improve and adapt security policies to protect systems from potential threats.

# Bibliography & References:

1. Soares, Eduardo, Angelov, Plamen, Biaso, Sarah, Higa Froes, Michele, and Kanda Abe, Daniel. "SARS-CoV-2 CT-scan dataset: A large dataset of real patients CT scans for SARS-CoV-2 identification." medRxiv (2020). doi: <https://doi.org/10.1101/2020.04.24.20078584>.  
   Angelov, P., & Soares, E. (2020). Towards explainable deep neural networks (xDNN). Neural Networks, 130, 185-194.
2. @article{zhao2020COVID-CT-Dataset,

title={COVID-CT-Dataset: a CT scan dataset about COVID-19},

author={Zhao, Jinyu and Zhang, Yichen and He, Xuehai and Xie, Pengtao},

journal={arXiv preprint arXiv:2003.13865},

year={2020}

}

1. @Article{he2020sample,
2. author = {He, Xuehai and Yang, Xingyi and Zhang, Shanghang, and Zhao, Jinyu and Zhang, Yichen and Xing, Eric, and Xie, Pengtao},

title = {Sample-Efficient Deep Learning for COVID-19 Diagnosis Based on CT Scans}, journal = {medrxiv}, year = {2020},}