Ministry of Higher Education and
Scientific Research
University of Diyala
College of Science
Department of Computer Science

Diagnosis of COVID-19 Virus Using Convolutional Neural Network Algorithm

A thesis

Submitted to the Department of Computer Science\ College of Sciences\ University of Diyala in a Partial Fulfillment of the Requirements for the Degree of Master in Computer Science

By

Ahmed Muneam Hussein

Supervised By

Prof. Dr. Ziyad Tariq Mustafa Al-Ta'i

2022A.C ***1443 A.H***

بسم الله الرحمن الرحيم

(يَرْفَعِ اللَّهُ الَّذِينَ آمَنُوا
مِنْكُمْ وَالَّذِينَ أُوتُوا الْعِلْمَ
دَرَجَاتٍ)

صدق الله العظيم

(سورة المجادلة :الآية 11)

ACKNOWLEDGMENTS

Praise be to God, Lord of the Worlds, and prayers and peace be upon the Master of the Messengers, our Prophet Muhammad and all his family and companions. I would like to express my thanks and appreciation to my supervisor, Prof. Dr. Ziad Tariq Mustafa Al-Taei, for his valuable guidance, instructions and constructive comments that made this work possible. I would also like to express my special thanks to the Head of the Computer Department, Prof. Bashar Taleb Hamed, Asst. Prof. Ghassan Sabeeh Mahmood for their help and support. Special thanks to the members of the committee for discussing my thesis, and I thank all my friends who have given me all the care and assistance. My last words go to my family. I would like to thank them for their love and endurance of the difficult circumstances that accompanied the completion of this research and for helping me achieve my goals.

Ahmed M Hussein

Dedication

For the one who brings people out from the darkness of ignorance to the light of guidance, the beloved of our hearts, the Messenger of God.

My father's soul.

My dear mother, whom I ask God to protect her.

My grandfather and my grandmother, I ask God to protect them.

My brothers and sisters, whom I cherish so much.

My teachers…

My master classmates …

My friends…

Everyone who helped me…

Ahmed M Hussein

(Supervisor Certification)

I certify that this research entitled "***Diagnosis of COVID-19 Virus Using Convolutional Neural Network Algorithm***" was prepared by **"*Ahmed Muneam Hussein*"** Under my supervision at the Department of Computer Science\ College of Sciences\ University of Diyala in a Partial Fulfillment of the Requirements for the Degree of Master in Computer Science.

(Supervisor)

Name: ***Prof. Dr. Ziyad Tariq Mustafa Al-Ta'i***

Signature:

Date: / /2022

Approved by Department of Computer Science\ College of Sciences\ University of Diyala

Signature:

Name: **Assist. Prof. Dr.Bashar Talib Hamed**

Date: / /2022

Head of Department of Computer Science\ College of Sciences\ University of Diyala

(Linguistic Certification)

I certify that the thesis entitled "***Diagnosis of COVID-19 Virus Using Convolutional Neural Network Algorithm***" was prepared by "***Ahmed Muneam Hussein***" and was reviewed linguistically. Its language was amended to meet the style of the English language.

Signature:

Name:

Date: / /2022

(Scientific Amendment)

I certify that the thesis entitled "***Diagnosis of COVID-19 Virus Using Convolutional Neural Network Algorithm***" was prepared by "***Ahmed Muneam Hussein***" and has been evaluated scientifically ;therefore, it is suitable for debate by the examining committee.

Signature:

Name:

Date: / / 2022

(Scientific Amendment)

I certify that the thesis entitled "***Diagnosis of COVID-19 Virus Using Convolutional Neural Network Algorithm***" was prepared by "***Ahmed Muneam Hussein***" and has been evaluated scientifically ;therefore, it is suitable for debate by the examining committee.

Signature:

Name:

Date: / / 2022

Examination Committee Certification

We certify that we have read the thesis entitled **"*Diagnosis of COVID-19 Virus Using Convolutional Neural Network Algorithm*"** *and as examination committee, examined the student* ***"Ahmed Muneam Hussein"*** *in the thesis content and that in our opinion, it is adequate as fulfill the requirement for the Degree of Master in Computer Science at the Computer Science Department, University of Diyala.*

Signature:

Name: **Assist. Prof. Dr. Jamal Mustafa Al-Tuwaijari** (Chairman)

Date: / / 2022

Signature:

Name**: Asst. Prof.Dr. Wathiq Najah Abdullah** (Member)

Date: / / 2022

Signature:

Name: **Asst. Prof. Ghassan Sabeeh Mahmood** (Member)

Date: / / 2022

Signature:

Name: **Prof. Dr. Ziyad Tariq Mustafa Al-T'ai** (Supervisor)

Date: / / 2022

Approved by the **Dean** of College of Science, University of Diyala (**The Dean)**

Signature:

Name: **Prof. Dr. Tahseen H. Mubarak**

Date: / /2022

Abstract

The COVID-19 virus has affected the world so severely that it is difficult for the top people in the world to escape it. Even nations with the best health care facilities struggle to keep up with the rising illnesses and fatalities. Humans who have this illness die as a result of significant damage to their lungs and respiratory systems.

Coronavirus virus 2019 (COVID-19) is a brand-new infection with no automatic model to correctly identify it from photos. It has spread around the world. With the help of chest computed tomography (CT), we hope to examine the ability of deep learning to diagnose COVID-19 infection. The suggested method could aid radiologists in making more accurate diagnoses and managing COVID-19 more effectively.

The suggested effort in this thesis involved the analysis of computed tomography (CT) images of the lung to categorize the infected and non-infected. To identify COVID-19 from regular CT scans, convolutional neural network (CNN) layers were pre-trained on the global COVID-19 dataset. Global accuracy has been tested for performance on test photos.

The non-COVID-19 classification has the highest accuracy of 99.1 percent when raw CT images are utilized as inputs to the assessed models. Our findings support CNN's very accurate prediction of COVID-19 on chest CT scans.

List of Contents

List of Figures

List of Tables

LIST OF *ABBREVIATIONS*

Abbreviations	Meaning
ANN	Artificial Neural Net□or□
AUC	Area under the □□C Curve
CNN	Convolutional Neural Net□or□
Colab	□oogle Colaboratory
C□□□D□□□	Corona □irus Disease 20□□ caused by SA□S□Co□2
C□	Computed □omography
DM	Data Mining
□N	□alse Negative
□□	□alse □ositive
□D□	□ntegrated Development □nvironment
□NN	□□Nearest Neighbor
M□□	Multi□□ayer □erceptron
NN	Neural Net□or□
NNs	Neural Net□or□s
□N□	□ortable Net□or□□raphic
□e□U	□ectified □inear Units
□□	□andom □orest Algorithm
□□C	□eceiver □perating Characteristic curve
□□□	□egion □f □nterest
S□DM	Stochastic □radient Descent □ith Momentum
S□M	Support □ector Machine
□N	□rue Negative
□□	□rue □ositive

Chapter One

Introduction

Chapter One

Introduction

1.1 Introduction

□eople currently suffer from a □ide range of diseases as a result of their lifestyle choices and their surroundings□As a result, detecting disease at an early stage is critical□□he most difficult step, ho□ever, is determining the right diagnosis of disease□Data mining plays a critical role in disease detection □□□□

A ne□ coronavirus disease, C□□□D□□□, appeared in December 20□□ in □uhan, China, and shortly after it had a big influence on the □orld□ Millions of cases have been verified thus far□and hundreds of fatalities □orld□ide□As a result, prompt and precise □dentification of C□□□D□□□ is crucial for managing the preventing the disease□s spread and lo□ering its fatality □2□□

□he □orld □ealth □rgani□ation named the global illness coronavirus (C□□□D□□) a pandemic on March □□, 2020□Covid□□□ illness has so far been lin□ed to more than □0 million confirmed cases, more than □00 thousand fatalities □orld□ide (mortality rate: □□□□), and more than □ million persons □ho have recovered□A prompt diagnosis is essential for halting the disease□s progress, improving the efficacy of medical intervention, and subse□uently raising the li□elihood of survival □ithout the need for intensive and sub□intensive care□□ecause hospitals only have a limited supply of e□uipment for critical care, this is an important topic □□□□

Data mining is regarded as an emerging technology that has made a radical change in the information □orld□□he term Data Mining (often called □no□ledge discovery) refers to the method of analy□ing data from different

perspectives and summari□ing it into valuable information employing several analytical tools and techni□ues, □hich in turn may be useful to increase the performance of a system□□echnically, data mining is the method of finding correlations or patterns among do□ens of fields in large relational databases□Data mining techni□ues such as classification and prediction, clustering, association rule mining, and various mining methods can be useful to apply to medical data □□□□

Classification is a process that divides the dataset into specified sections and then classifies the data □hich is a t□o□phase process: □n the first phase, it develops a model based on educational datasets of databases and then creates an educational dataset including records, samples, e□amples, and things □ith a collection of attributes and aspects□□ach sample has a specific class label□□n the second phase, the developed model in the previous phase is used to classify ne□ samples □□□□

Classification algorithms can be either supervised or unsupervised based on the learning mechanism□Supervised learning is implemented by a set of labels defined before the training set□□he function is mapped for ne□ unseen data to predict the labels □□□□

Medical data mining is the process of finding a useful pattern that helps in medical diagnosis□□n the medical domain, the popularity of data mining is increasing constantly as it helps e□plore the un□no□n patterns and improves prediction models □hich help in medical decision ma□ing□Data classification is one of the tas□s in data mining □□□□

□eature selection is the process of identifying and removing as much of the irrelevant and redundant information as possible□□n the database, not all attributes available are useful □□□□

□ ithout prior □no□ledge, it is difficult to determine □hich features are useful□As a result, a large number of features are usually introduced to the

data set, □hich include relevant, irrelevant, and redundant features □o□ever, irrelevant and redundant features are not useful for classification, and they may even reduce the classification performance due to the large search space known as “the curse of dimensionality □□□□

□his thesis proposes a diagnosis of Covid□□□using data mining techni□ues□

1.2 Related Work

□his section revie□s some of the previous studies and e□plains the different techni□ues that are used to diagnose covid□□□disease□

□ Carvalho et□ al□ (2020) □□0□ a trainable resource e□tractor employing Convolutional Neural Net□or□s (CNN) and several classifiers □as presented as a solution for diagnosing C□□□D□□□□□irst, histogram e□uali□ation and Contrast □imited Adaptive □istogram □□uali□ation □ere used to enhance the photos□□uality (C□A□□)□After then, □0□C□s □ere used to harvest resources, □□2 of □hich had C□□□D□□□and □□□did not□ After the data has been retrieved, C□□□D□□□ and Non□C□□□D□□□ categori□ation is done using a variety of classifiers□□he findings reveal a □appa inde□of 0□□□□□, an accuracy of □□□□□□, a recall of □□□□□□, and a precision of □□□□□□

□ Carvalho e□t al□(2020) □□□□ Convolutional neural net□or□s (CNNs) for feature e□traction in computed tomography (C□) scans and □□□oost classification □ere suggested in a publication for the diagnosis of C□□□D□ □□ (□□□oost or the □□treme □radient boost is a machine learning algorithm that is used for the implementation of gradient boosting decision trees□)□A CNN is used in the approach to e□tract features from □0□C□s, □□2 of □hich include C□□□D□□□and □□□of □hich do not□□he outcomes revealed an accuracy of □□0□□, recall of □□0□□, precision of □□□□□□, □□score of □□□, AUC of □□□, and a □appa inde□of □0□□

 Alakus and Turkoglu (2020) [12], suggested a model to generate clinical prediction models that calculate which patients are most likely to get a COVID-19 disease using deep learning and laboratory data. Models were evaluated using 18 laboratory results from 600 patients, and 10 fold cross-validation and train-test split techniques were used to verify them. The experimental findings showed that COVID19 illness is detected by predictive models with 86.66 %, 91.89 % F1-score, 86.75 % precision, 99.42 % recall, and 62.50 % AUC.

 (Villavicencio et al., 2021) [13], utilizing the COVID-19 symptoms and presence dataset from Kaggle, built a model to assess and predict the presence of COVID-19 using a variety of supervised machine learning methods. Waikato Environment for Knowledge Analysis (WEKA) machine learning software was used to apply the Decision Tree, Random Forest, Support Vector Machine, K-Nearest Neighbors, and Naive Bayes algorithms. Using 10-fold cross-validation, the performance of each model was assessed and contrasted in accordance with the key accuracy metrics. The outcomes demonstrated that Support Vector Machine using Pearson VII (The Pearson VII function was a popular function during the 1980s and 1990s for describing peak shapes from conventional X-ray powder diffraction patterns, though it has since been surpassed in popularity by the pseudo-Voigt peak-shape function) universal kernel outperforms other algorithms by achieving 98.81% accuracy and a 98.81% precision rate.

 (Dutta et al., 2021) [14], offered three distinct supervised machine learning models for COVID-19 diagnosis. For categorizing the COVID-19 datasets, they have examined the classification outcomes of several strategies, including the bagging algorithm, k-nearest neighbor, and random forest. They used symptoms from an Indian Covid-19 tracker for categorization purposes even though India has advanced to the second

stage. Different performance metrics were used to assess each technique's performance. The classification findings demonstrate that the random forest, which used an accuracy of 85.71 % and an F1 score of 0.833%, had better results.

- (Kugunavar & Prabhakar, 2021) [15], suggested a straightforward CNN framework for binary classifying COVID-19 CT images. With an F1-score of 93%, they obtained an accuracy of 93%. Thus, it was clear that CNNs are highly valuable for the effective diagnosis and prognosis of COVID-19 with the availability of better medical picture datasets.

- (Wang et al., 2021)[16], suggested employing a convolutional neural network a deep learning technique (CNN), the specificity and sensitivity of the internal validation were 0.88 and 0.87, respectively, yielding a total accuracy of 89.5 percent. An overall accuracy of 79.3 percent, a specificity of 0.83, and a sensitivity of 0.67 were shown in the external testing dataset. Furthermore, 46 of 54 COVID-19 photos were correctly predicted by the algorithm to be COVID-19 positive, with an accuracy of 85.2%, even though the first two nucleic acid test results were negative.

1.3 Problem Statement

The emerging coronavirus disease (COVID-19) was and still is a dilemma, due to its different behavior according to the nature of the body of each infected person and his/her immunity. This is in addition to the overlap of the symptoms of COVID-19 disease with the symptoms of other diseases. Therefore, the correct diagnosis of this disease is one of the most important stages of its treatment, which is the problem of this thesis.

1.4 Aim of the Thesis

This work aims to design and implement a diagnostic model for the emerging coronavirus disease (COVID-19). This model can accurately classify if the patients have the COVID-19 disease or not based on the CT scan images dataset using CNN algorithm.

1.5 Outline of Thesis

The other chapters in this thesis are as follows:

Chapter Two: Theoretical Background

This chapter gives the background and review of diagnosis Covid-19, feature selection techniques, and classification techniques.

Chapter Three: The Proposed Model

This chapter describes the proposed diagnosis Covid-19 with its design and implementation.

Chapter Four: Results and Discussion

This chapter explains the results and evaluations that have been getting from the proposed diagnosis.

Chapter Five: Conclusions and Suggestions for Future work

This chapter presents the conclusions of this work. Furthermore, it provides suggestions for future work.