Logistic Regression Model of Study of Identifying Population for Covid-19 Medical Test – Machine Learning Approach

John T Mesia Dhas[1*] T. S. Shiny Angel[2] J. Sheeba[3]

[1]Associate Professor, Department of Computer Science and Engineering,   T John Institute of Technology, Bangalore, India
[2]Associate Professor, Department of Software Engineering, SRM Institute of Science and Technology, Chennai, India
[3]Associate Professor, Mohamed Sathak A.J Academy of Architecture, Chennai, India

Abstract

Background / Objectives: COVID-19: Corona is a virus which affects the human lungs and causes death in the worst cases. It is a pandemic, which is killing millions of human beings in the first quarter of twenty first century. The existing vaccines, or medicines are not giving full healings from the corona virus. The prevention from corona is the better than cure from it.
Methods / Statistical Analysis: In this article, using logistic regression model for classifying the targeted population to undergo the COVID-19 test.
Findings: This study gives the expected number of COVID-19 cases at a particular time.
Applications / Improvements: To predict the population may be affected by COVID-19, it will be necessary to take precautions to stop the spread of the pandemic.

Keywords: COVID-19, Coronavirus, Machine Learning, Logistic Regression, Pandemic

1. Introduction

Coronaviruses affect both humans and animals worldwide. The first coronavirus was reported in 1949 [1], but it has been actively studied since the 1970's. The human coronaviruses cause only the common cold. But, in 2003, a new human coronavirus was responsible for severe acute respiratory syndrome (SARS). The origin of the SARS coronavirus (SARS-CoV) has increased the research related to the coronaviruses.

The SARS-CoV and the Middle East respiratory syndrome (MERS) in 2012 were both coronaviruses that were transmitted from animals to humans.

*John T Mesia Dhas, Department of CSE, T John Institute of Technology, Bangalore, India.
E-mail: jtmdhasres@gmail.com

Taxonomy

The term "coronavirus," coined in 1968, refers to the "corona"-like or crown-like morphology observed in electron microscopy for these viruses [1]. Many coronaviruses cause acute and chronic respiratory, enteric, and central nervous system (CNS) diseases in many species of animals, including humans.

Coronaviruses, Hosts and Diseases

The different types of coronaviruses [1] and their hosts, along with the diseases caused by them, are listed in Table 1.

Table 1. Coronaviruses, Hosts, Diseases

S. No.	Virus	Host	Disease(s)
1	229E	Human	Respiratory infection
2	TGEV	Pig	Respiratory and enteric infection
3	PRCoV	Pig	Respiratory infection
4	FIPV	Cat	Respiratory, enteric, and neurologic infection, and hepatitis
5	NL-63	Human	Respiratory infection, croup
6	OC43	Human	Respiratory infection and possibly enteric infection
7	MHV	Mouse	Enteric and neurologic infection and hepatitis

S. No.	Virus	Host	Disease(s)
6	OC43	Human	Respiratory infection and possibly enteric infection
7	MHV	Mouse	Enteric and neurologic infection and hepatitis
8	Sialodacryoadenitis coronavirus	Rat	Neurologic infection
9	Hemagglutinating encephalomyocarditis virus	Pig	Respiratory, enteric, and neurologic infection
10	BCoV	Cow	Enteric infection
11	HKU1	Human	Respiratory infection
12	SARS-CoV	Human	Severe acute respiratory syndrome
13	IBV	Chicken	Respiratory infection, hepatitis
14	Turkey coronavirus	Turkey	Respiratory and enteric infection
15	MERS	Human	Middle East respiratory syndrome
16	COVID-19	Human	Fever, fatigue, and dry cough

2. Machine Learning

The intelligent system is learning from its existing knowledge and environment. This is known as machine learning. It has three different types of learning. They are: supervised learning, unsupervised learning, and reinforcement learning. Logistic regression is one of the supervised learning algorithms for classification. In this research, a logistic regression model is used for identifying the COVID-19 patients from the total population using various syndromes affected by the people.

Machine Learning Model for COVID-19

COVID-19 is a pandemic that affected about 3 million people and killed around 170,000 people worldwide. It affects the social, economic, and cultural life of the people without any topographical or ethnic differences. According to World Health Organization (WHO) research, COVID-19 may affect approximately 30 million people worldwide and claim the lives of approximately 500 thousand people. The international community is fighting against this deadly virus without any specific medicine.

The Transmission Routes of COVID-19

The WHO identified the various transmission routes of COVID-19, they are,
1. The close contact with the symptomatic people is one of the main reasons for the spread of COVID-19.
2. The pre-symptomatic period also possible to spreads COVID-19. Around 14 days as incubation period for COVID-19.
3. The asymptomatic transmission or social transmission is another way of spread, which is spreading without knowing the origin or source of affection.

The Symptoms of COVID-19

The common symptoms of COVID-19 are, fever, tiredness, dry cough, etc. Some other symptoms from the clinical study are, aches and pains, nasal congestion, runny nose, sore throat and diarrhoea

Most Targeted Population for COVID-19

Around eighty percentage of the people recover from this decease without any special treatment. Around 1/6 of the COVID-19 affected patients only required a special care. It is mostly affecting seriously the older people with underlying medical problems like high blood pressure, heart problems or diabetes.
The countries in the world doesn't have sufficient medical facilities for diagnosing novel COVID-19. The respiratory samples testing kits like reverse transcription polymerase chain reaction (rRT-PCR) is not available sufficiently for the proposition of the population of the world. So that, a random sampling is taking place in the COVID-19 diagenesis process. This process must be performed only the targeted

population, that is the person having all the syndromes of COVID-19, then only, we can reduce the death toll by the human eater COVID-19 with limited available resources.

Treatments for COVID-19

No any particular treatments or medicines are available for novel COVID-19, so that, the combination of malarial drug (hydroxychloroquine), fever and cough drugs and immunity drugs and highly nutritious food are providing to the COVID-19 patients.
To avoid the spread of COVID-19: to frequently clean the hands, cover your cough with tissue, and maintain a distance of at least one meter from people who are coughing and sneezing.

3. Logistic Regression

The logistic regression is one of the classification algorithms of supervised learning. The main principle of logistic regression is learning the problem with many different sources and to make classifications based on the features of it. In this research, the various social, communal and medical parameters are analysed and to create the targeted population for doing the COVID-19 laboratory test. This model classifying the total population into two classes, they are positive and negative. The limited positive class only undergo the COVID-19 test cases.

Algorithmic Model for COVID-19 Test Cases

The classification method using logistic regression is used in the algorithm COVID-19TEST.

```
Algorithm COVID-19TEST
Inputs: Population (name, age, location, cough, fever, tiredness, pain, nosalCongestion,
        runnyNose, soreThroat, bloodPressure, heartProblem, diabetes, travel, party,
        PCRTest = FALSE, IGG, IGM)
        COVID-19Positives (name, age, location)
Output: PCRTestTRUE(name, age, location)
        PCRTestFALSE(name, age, location)

For all x in Population
        For all y in COVID-19Positives
                If x.location == y.location Then
                x.PCRTest = TRUE
                For all x in Population
                If x.age >= 50 Then
                        x.PCRTest = TRUE
                        Elseif x.travel == TRUE Then
                        x.PCRTest = TRUE
                Elseif x.party == TRUE Then
                        x.PCRTest = TRUE
                Elseif x. bloodPressure == TRUE OR x.heartProblem == TRUE OR
```

```
                    x.diabetes == TRUE Then
                    x.PCRTest = TRUE
            Elseif x. cough == TRUE OR x.fever == TRUE OR x.tiredness == TRUE OR
                    x.pain == TRUE Then
                    x.PCRTest = TRUE
            Elseif x.nosalCongestion == TRUE OR x.runnyNose == TRUE OR
                    x.soreThroat == TRUE Then x.PCRTest = TRUE
            Elseif x.IGG == TRUE OR x.IGM == TRUE Then
                    x.PCRTest = TRUE
            Endif
For all x, y, z in Population, PCRTestTRUE, PCRTestFALSE
        If x.PCRTest == TRUE Then
                y.name = x.name
                y.age = x.age
                y.location = x.location
        Else
                z.name = x.name
                z.age = x.age
                z.location = x.location
        Endif
```

The COVID-19TEST algorithm has two input dictionaries namely Population and COVID-19Positives. In the dictionary population: name and location are string variables, age is an integer variable, cough, fever, tiredness, pain, nosalCongestion, runnyNose, soreThroat, bloodPressure, heartProblem, diabetes, travel, IGG, IGM and party are Boolean variables, and the default value of Boolean variable PCRTest = FALSE. The values of all the variables are received from the person and some minor medical check-ups. The IGG and IGM are the immunoglobulin test for identifying the immunization of a person using rapid test kits. Another dictionary COVID-19Positives is an already maintaining database of existing COVID-19 patients.

The output dictionaries PCRTestTRUE and PCRTestFALSE are two classifications based on our algorithm for PCR test required persons and not required persons respectively.

The COVID-19 is easily spreading by the primary and secondary contacts of existing COVID-19 patients. So that, all the people in the locations of existing COVID-19 patients are must undergo rRT-PCR test. Therefore, the persons living at the locations of COVID-19 is getting PCRTest = TRUE.

The age factors, persons with repeated travel histories, persons are attending big gatherings, persons with blood pressure, heart problem and diabetes, persons affected with cough, fever, tiredness, pain, nosal congestion, runny nose and sore throat, and rapid test IGG and IGM positive are important factors and syndromes of COVID-19. Therefore, people with these syndromes must undergo PCR test. Hence, PCRTest = TRUE.

The TRUE and FALSE values of variable PCRTest of dictionary Population is separated in the dictionaries PCRTestTRUE and PCRTestFALSE respectively. The details present with dictionary PCRTestTRUE is population want to undergo rRT-PCR test. That is, the details present in PCRTestTRUE is possible persons can be affected with COVID-19.

4. Conclusion

The fast-spreading COVID-19-man eater virus not only killing human beings, it is destroying social, cultural and economic life of the people throughout the world. The international community is fighting against this cruel virus without any specific treatment or medicine. This research will help to find the targeted population and make them to undergo treatment for the deadly decease. It is making the resource utilization effectively. So that, we can expect better results in this COVID-19 biological world war.

Bibliography

1. Coronavirus Pathogenesis and the Emerging Pathogen Severe Acute Respiratory Syndrome Coronavirus Susan R. Weiss1* and Sonia Navas-Martin2*
2. Systematic Comparison of Two Animal-to-Human Transmitted Human Coronaviruses: SARS-CoV-2 and SARS-CoV Jiabao Xu 1,y, Shizhe Zhao 1,y, Tieshan Teng 1 , Abualgasim Elgaili Abdalla 2, Wan Zhu 3, Longxiang Xie 1,* , Yunlong Wang 4,* and Xiangqian Guo 1,*
3. Transmission routes of covid-19: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200402-sitrep-73-covid-19.pdf?sfvrsn=5ae25bc7_2
4. SYMPTOMS of COVID-19: https://www.who.int/news-room/q-a-detail/q-a-coronaviruses
5. Hydroxychloroquine: https://www.statnews.com/2020/04/20/novartis-study-hydroxychloroquine/
6. Covid-19 testing: https://en.wikipedia.org/wiki/COVID-19_testing
7. IGG IGM : https://www.elabscience.com/p-covid_19_igg_igm_rapid_test-375335.html