

Covid-19 Current Symptoms, Comorbidities, Treatments Study and Infection Prediction Based on Symptoms

PrithiviRaj K

Dept of Information Technology
University College Of Engineering, BIT Campus, Trichy, Tamilnadu

Abstract- COVID-19 is shaking the world ferociously. It is a virus contagion as pointed by the WHO. This infection attacks a person so instantly via droplets from sneezing or saliva, and touching any infected surface then Entering into the body via touching face. Diagnosis and treatment for this infection is the most challenging task in the medical field and it takes more time and resources. Looks like Social distancing is the only way to break the chain. Symptoms of this infection are not similar for every infected person and current information suggests that the incubation period ranges from 2 to 14 days. In this work to assess the symptoms, treatment and comorbidities by mining the available data and predict whether the person is infected by covid or not, based on symptoms and other clinical characteristics.

Keywords- covid, symptoms, diagnosis, treatment, mining.

I. INTRODUCTION

COVID stands for CoronaVirus Disease. COVID-19 is an infectious disease caused by a virus organism called coronavirus. This was referred to as the '2019 novel coronavirus'. The COVID virus a new virus linked to the same family of viruses as Severe Acute Respiratory Syndrome (SARS). The virus Transmitted through the droplets from the sneezing or saliva of the infected person and touching any infected surface then touching their nose, mouth or eye. The Corona virus strains are divided into seven different categories¹.

- i) 229E (alpha corona virus)
- ii) NL63 (alpha corona virus)
- iii) HKU1 (beta corona virus)
- iv) OC43 (beta corona virus)
- v) MERS-CoV (beta corona virus ; causes Middle East respiratory syndrome or MERS)
- vi) SARS-CoV (beta corona virus; causes severe acute respiratory syndrome or SARS)
- vii) SARS-CoV- 2 (novel corona virus ; causes corona disease in 2019 or COVID-19).

The Outbreak of coronavirus has been declared a Public Emergency of International concern and the virus has now spread to many countries and territories. While a lot is still unknown about the virus that causes COVID-19. The COVID does not differentiate between borders, ethnicities, disability status, age or gender⁷.

Generally All viruses are intracellular parasites. Their only mission in life, if you can call it that, is to use the mechanisms of human cells to make copies of themselves. The general life cycle of the virus is given in the Figure 1.1.

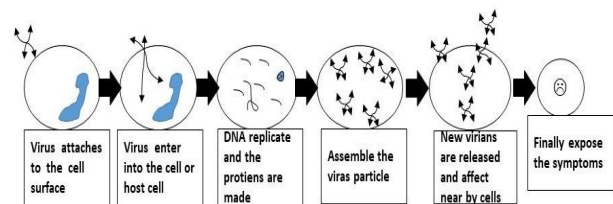


Figure 1.1

The origin of SARS-cov is thought to be bats. It's believed that the virus passed from the bats to an intermediate animal host, the civet cat before jumping to the human.

The tricky task is finding a way into the cells where they can replicate. The novel coronavirus, SARS-CoV-2, isn't particularly good at that. Research so far suggests that it sneaks in through a single cellular door: a membrane protein called ACE2. SARS-CoV-2 the coronavirus that causes the disease COVID-19 is completely new and attacks cells in a novel way. Every virus is different and so are the drugs used to treat them. That's why there wasn't a drug ready to tackle the new coronavirus that only emerged a few months ago.

In this time the digital data about the covid are rapidly emerging. Data mining has already been established as a novel field for exploring hidden patterns in the huge datasets. Medical science is another field where a large amount of data is generated using different clinical reports and other patient symptoms. These explored hidden patterns in

medical datasets can be used for clinical diagnosis. However, medical datasets are widely dispersed, heterogeneous, and huge in nature. These datasets can be used for the analytics and predictions about covid-19.

II. RELATED WORKS

1. Narmada Talasila et. al¹

In this paper the author presents an informative study about the Risk Factors, Death Rate And Recovery Rate of COVID-19 Affected Patients In India. The data was collected from the date of 30/01/2020 to 27/03/20. Sample size of the study is 730 COVID-19 patients. By mining the databases, the author found that The age groups of 21-40 and 41-60 years, male gender are more prone to COVID-19 infection. Lack of social distancing is the major risk factor of COVID-19 infection.

2. Ying Liu et. al²

This work is about “The reproductive number of COVID-19 is higher compared to SARS coronavirus”. Where the term ‘coronavirus & basic reproduction number’ was used. The time period covered was from 1 January 2020 to 7 February 2020.

Here they review the basic reproduction number (R_0) of the COVID-19 virus. R_0 is an indication of the transmissibility of a virus, representing the average number of new infections generated by an infectious person in a totally naïve population. For $R_0 > 1$, the number infected is likely to increase, and for $R_0 < 1$ transmission is likely to die out. The basic reproduction number is a central concept in infectious disease epidemiology, indicating the risk of an infectious agent with respect to epidemic spread. This review found that the estimated mean R_0 for COVID-19 is around 3.28, with a median of 2.79 and IQR of 1.16, which is considerably higher than the WHO estimate at 1.95. Based on these considerations, R_0 for COVID-19 is expected to be around 2–3, which is broadly consistent with the WHO estimate.

3. B. Cao et. al³

In this work the author did “A Trial of Lopinavir–Ritonavir in Adults Hospitalized with Severe Covid-19”. Test where conducted a randomized, controlled, open-label trial involving hospitalized adult patients with confirmed SARS-CoV-2 infection, which causes the respiratory illness Covid-19, and an oxygen saturation (SaO_2) of 94% or less while they were breathing ambient air or a ratio of the partial pressure of oxygen (PaO_2) to the fraction of inspired oxygen (FiO_2) of less

than 300 mm Hg. By this work they suggest that, in hospitalized adult patients with severe Covid-19, no benefit was observed with lopinavir–ritonavir treatment beyond standard care. Future trials in patients with severe illness may help to confirm or exclude the possibility of a treatment benefit.

4. Puja Mehta et. al⁴

In this work the author presents an “COVID-19: consider cytokine storm syndromes and immune suppression”. Where the various clinical characteristics of the COVID persons such as Temperature, Organomegaly, Number of cytopenias, Triglycerides, Fibrinogen, Ferritin, Hemophagocytosis were indexed. From this work they suggest that All patients with severe COVID-19 should be screened for hyperinflammation using laboratory trends (eg, increasing ferritin, decreasing platelet counts, or erythrocyte sedimentation rate) and the Score to identify the subgroup of patients for whom immune suppression could improve mortality. Therapeutic options include steroids, intravenous immunoglobulin, selective cytokine blockade (eg, anakinra or tocilizumab) and JAK inhibition.

5. Ying-Ying Zheng et. al⁵

This work presents the study about the SARS-CoV-2 and ACE2, Acute cardiac injury, Chronic cardiovascular damage, Patients with pre-existing CVD. This study makes an understanding about COVID-19, while also causing damage to the myocardium, although the specific mechanisms are uncertain. Patients with underlying CVD and SARS-CoV-2 infection have an adverse prognosis. Therefore, particular attention should be given to cardiovascular protection during treatment for COVID-19.

III. DATABASE DESCRIPTION

The Dataset was obtained from Kaggle, ISARIC Data Platform, press and other official websites. The dataset containing patient information from the date of 20/1/2020 to 8/3/2020. It contains covid data about 30 countries. The dataset contains 1,33,155 patient data. Some of the data of these COVID-19 patients were missing, so the data filled by using mining methods.

IV. METHODOLOGY

1. DATA MINING:

data mining is a process used to extract usable data from a larger set of any raw data. It implies

analysing data patterns in large batches of data using one or more software. Data mining is also known as Knowledge Discovery in Data. While considering this proposed work, it is used to extract the information such as symptoms, comorbidities, treatment, age group, and other epidemic characteristics of the patient.

2. MACHINE LEARNING :

Machine learning is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention. Machine learning algorithms divided into three types,

- i) Supervised Learning
- ii) Unsupervised Learning
- iii) Reinforced Learning

While considering the proposed work, the classifier is going to predict whether the person is infected or not by COVID based on symptoms and other relevant characteristics. Where the Decision Tree algorithm is used for the classification.

2.1 DECISION TREE ALGORITHM :

Decision Tree algorithm belongs to the family of supervised learning algorithms. The decision tree algorithm tries to solve the problem, by using a tree representation. Each internal node of the tree corresponds to an attribute, and each leaf node corresponds to a class label.

Algorithm :

1. Input training data and with features and class label
2. Compute Entropy of the whole training dataset

$$\text{Entropy}(E_x) = - \sum_{x \in X} p(x) \log p(x)$$
3. Compute the Information gain of the each attribute

$$\text{Information gain}(IG_x) = E_x - E_{x/A}$$
4. Place the best attribute (Which attribute contain maximum IG value) of the dataset at the **root** of the tree.
5. Split the training set into subsets. Subsets should be made in such a way that each subset contains data with the same value for an attribute.
6. Repeat step 4 and step 5 on each subset until you find **leaf nodes** in all the branches of the tree.

3. DATASET PREPARATION

Dataset contains more columns about the patient details. It needs to be organized for mining and classification. The following steps were followed to prepare the training dataset.

- Some columns have missing data. But there should not be a missing data for the classification. To solve this issue, an approach used forward filling and backward filling. But some data with more missing values are eliminated.
- Symptoms attribute contains the categorical data but it needs to be converted into the continuous data for the classification. Where 25 and above unique symptoms are experienced by the patient single or as a group of symptoms. To convert this into the continuous data, an approach used One-hot encoding.
- Feature selection is the process where the features automatically or manually select those features which contribute to the prediction variable to be classify. Having irrelevant features in the data can decrease the accuracy of the model. For the feature selection here is an approach using Logistic Regression Feature Elimination.
- After the feature elimination, the following attributes are selected as a features: age, whether the person is from or recently visited the infected area, and some symptoms such as cough, fever, pneumonia, chills, muscle pains, throat, difficulty in breathing and so on.

From the above steps, the training dataset was obtained for the data analysis and classification. The symptoms, comorbidities, treatment data were plotted against the number of Covid present(positive), Absent(negative), unknown(waiting for report) patients.

V. EXPERIMENTS AND RESULTS

From the date of 22/1/2020 to 8/3/2020 world wide covid confirmed, deaths, recovered cases plotted in the figure 5.1.

The dataset contains 67,800 confirmed cases and 5,476 death cases and 59,879 recovered cases.



Figure 5.1

The figure 5.2 reveals that the covid country wise distribution of the confirmed cases. This suggest that the china has a high impact of covid and france, germany, iran, italy, south korea, spain and U.S have a high confirmed covid rate. This study suggests that the above mentioned counties can get a high impact by covid as china suffered.

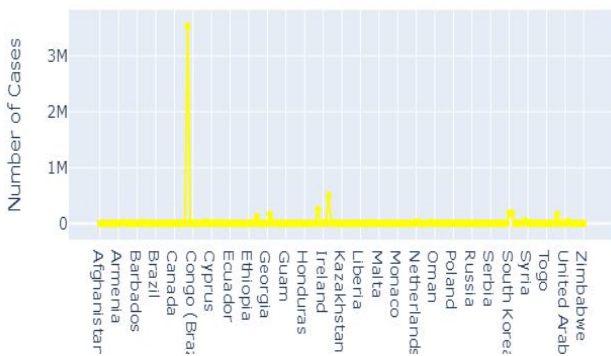


Figure 5.2

The figure 5.3 reflects that the treatment distribution of the Covid present, absent, and unknown patients. Oxygen therapy(50%), Invasive ventilation(17.1%), Non-Invasive ventilation (13.1%), Antibiotic (7.6%) has given the good benefit or the covid present patient and other treatments(12.2%) has given some approximate benefits.

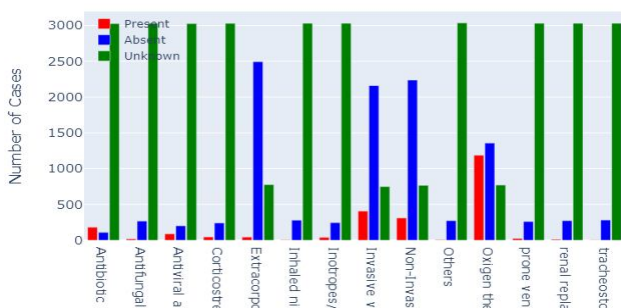


Figure 5.3

The figure 5.4 describe about the comorbidities of the covid present, absent and unknown patients.

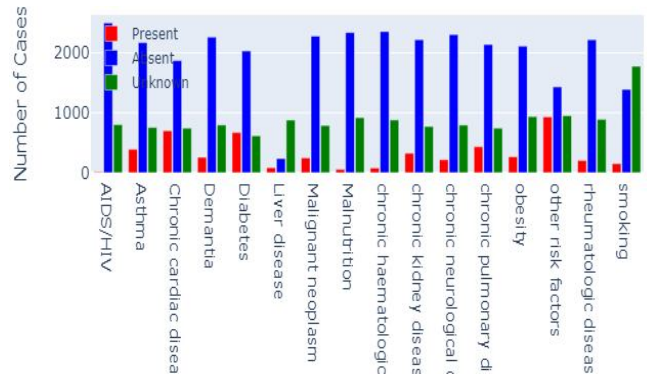


Figure 5.4

While considering all patient data, high exposure symptoms such as fever(21%), shortness of breath (18.8 %), cough(17%), fatigue/malaise(11.3%), Myalgia(5.4%) and others(26.5%) are experienced by the patient.

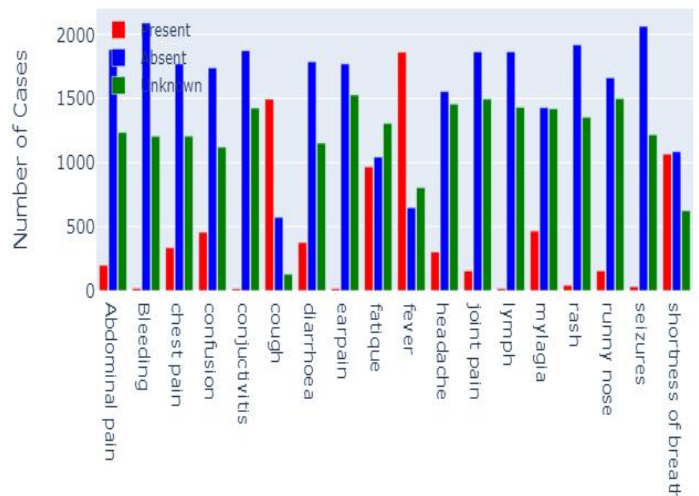


Figure 5.5

Covid present (positive) patients also experienced some low exposure symptoms such as abdominal pain, bleeding, conjunctivitis, ear pain, lymph and seizures .

Table 5.1 reflects the minimum- maximum age group infected and symptoms experienced by the covid positive patients with respect to the countries available in the dataset.

S.NO	Country	Min -Max age	Unique Symptoms Experienced
1	China	3-89	fever, cough, difficult in breathing
2	France	9-80	Fever, cough
3	Japan	5-85	Fever,cough, chills, joint pain,abdominal pain, pneumonia, ,vomiting, diarrhea, fatigue, muscle pain, spuntum, malaise, breathlessness, sore throat, runny nose
4	Malaysia	2-83	fever, cough, headache, flu, sore throat, mild cough
5	Nepal	32-	Fever, sore throat ,cough
6	Singapore	0.35-72	sore throat, fever, cough, chills
7	SouthKorea	11-82	high fever, chills, muscle pain,cough, sore throat, muscle pain, sore body
8	Taiwan	15-85	fever, cough, shortness of breath, chills, sore throat, runny nose, chest pain, shortness of breath, reflux, throat discomfort
9	Thailand	6-74	fever, chills, sore throat, headache, cough, muscle aches
10	USA	25-85	fever, cough
11	Vietnam	0.25-86	headache, runny nose, tired, fever
12	Australia	8-85	Tired , fever, cough
13	Canada	25-85	fever, cough
14	Cambodia	60-	fever, runny nose
15	UAE	8-73	fever, cough
16	HongKong	16-96	Fever, cough, myalgia, shortness of breath, runny nose, diarrhea, sore throat, cough with spuntum, vomiting
17	Italy	35-89	Fever, cough, shortness of breath, chills, sore throat
18	Russia	35-	Fever, cough
19	UK	35-55	Fever, cough
20	India	43-85	Cough, fever, cough, sore throat
21	Philippines	38-80	Cough, fever, cough, sore throat
22	Finland	35-	Cough, Finland
23	Spain	18-81	sore throat, fever
24	Sweden	25-	Cough, fever, sore throat.

For infection prediction, supervised learning approach decision trees were used. The prediction accuracy is 85% and Precision is 64%.

V. CONCLUSION

Covid-19 is the most lethal one. There is no proper vaccine exist for covid still now. These kind of predictive study and data analysis will help to understanding the pandemic level of the covid. This study suggests that the impacts of the existing level symptoms, comorbidities, treatment and covid pandemic impacts in the world wide. The

predictive study recommends some key patterns to predict the covid infection in a patient.

VI. FUTURE WORKS

The covid data is rapidly growing with respect to the time. Lot of data are emerging in the field. The analysis and the model can be improved by collecting updated data of the pandemic level and additional clinical features.

REFERENCES

- [1] Narmada Talasila, Sai Amith Papatla “Risk Factors, Death Rate And Recovery Rate of COVID-19 Effected Patients In India: An Informative Study”, *ISSN [ONLINE]: 2395-1052, IJSART - Volume 6 Issue 4 – APRIL 2020*
- [2] Ying Liu¹, Albert A. Gayle, Annelies Wilder-Smith, and Joacim Rocklöv, “The reproductive number of COVID-19 is higher compared to SARS coronavirus”, *Journal of Travel Medicine*, 2020, 1–4 doi: 10.1093/jtm/taaa021 Advance Access
- [3] Drs. Cao, Y. Wang, Wen, W. Liu, Jingli Wang, Fan, L. Ruan, Song, Cai, and M. Wei and Drs. D. Zhang and C. Wang, “A Trial of Lopinavir–Ritonavir in Adults Hospitalized with Severe Covid-19”. Published on March 18, 2020, at NEJM.org. DOI: 10.1056/NEJMoa2001282
- [4] Puja Mehta, Daniel F McAuley, Michael Brown, Emilie Sanchez, Rachel S Tattersall, Jessica J Manson. “COVID-19: consider cytokine storm syndromes and immune suppression”. Published Online March 13, 2020 [https://doi.org/10.1016/S0140-6736\(20\)30628-0](https://doi.org/10.1016/S0140-6736(20)30628-0)
- [5] Ying-Ying Zheng, Yi-Tong Ma, Jin-Ying Zhang and Xiang Xie. “COVID-19 and the cardiovascular system”. <https://doi.org/10.1038/s41569-020-0360-5>
- [6] Severe Outcomes Among Patients with Coronavirus Disease 2019 (COVID-19) —United States, February 12–March 16, 2020. *Morbidity and Mortality Weekly Report Early Release / Vol. 69 March 18, 2020*
- [7] Key Messages and actions for COVID-19 Prevention and control in schools, Education in emergencies, UNICEF New York.
- [8] International Severe Acute Respiratory and Emerging Infections Consortium (ISARIC) A global federation of clinical research networks, providing a proficient, coordinated, and agile research response to outbreak-prone infectious diseases COVID-19 Report: 08 April 2020.
- [9] Said Nadeem , “Coronavirus COVID-19: Available Free Literature Provided by Various Companies, Journals and

Organizations around the World”.Ong Chem Res, 2020,
5(1): 7-13, Document ID: 2020JOCR37 doi:10.528
/zenodo.3722904

[10]Rajesh T. Gandhi, M.D., John B. Lynch, M.D, M.P.H,
Carlos del Rio, M.D. “Mild or Moderate Covid-19 ”.
DOI:10.1056/NEJMcp2009249.