Proceedings of the Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV 2021).
IEEE Xplore Part Number: CFP21ONG-ART; 978-0-7381-1183-4

Naïve Bayes Classifier for Predicting the Novel Coronavirus

2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV) | 978-1-6654-1960-4/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICICV50876.2021.9388410

Sugandh Bhatia
Assistant Professor in Computer Science
Punjab School of Economics
Guru Nanak Dev University, Amritsar - India
sugandhcs.rsh@gndu.ac.in

Jyoteesh Malhotra
Associate Dean
Academics and Student's Welfare
GNDU Regional Campus
Jalandhar - India

***Abstract*— These days, the healthcare enterprises procure huge amount of healthcare data that most of the times is not processed to find out the hidden facts and patterns. Data mining along with machine learning performs a prominent role in predicting the diseases. Nowadays, COVID – 19 has become a pandemic for the mankind. It is a communicable disease and it takes 12 – 24 hours in receiving the reports of diagnose. In various remote and high altitude areas and due to the exponential growth of COVID – 19 in various parts of the world, it is not feasible to perform the test on mass population. This research article describes a novel technique to diagnose coronavirus using naïve bayes classifier and we hope that this technique would be useful and fruitful for the humanity and will be a great step to predict the COVID – 19.**

Keywords—coronavirus; naïve bayes classifier; data mining

I. INTRODUCTION

"Coronaviruses (COV) are a family of viruses that cause illness ranging from the common cold to more severe diseases such as Middle East Respiratory Syndrome (MERS – CoV) [1] and Severe Acute Respiratory Syndrome (SARS – CoV)" [2]. It is considered as a novel virus that has not been detected earlier in human beings. It is covered under the category of zoonotic as it can be transmitted between people and animals. Common signs of infection are shortness of breath, cough, fever and respiratory problems. In case, the patients who are of age 60 and above, infection can cause problem of pneumonia, kidney failure and even death. As per the report published in South China Morning Post [3], this novel virus was identified in 266 persons in November 2019. The details procured by the Post revealed that a 55 year old person from Hubei province could have been the first patient to contract COVID – 19. The virus which is now announced as pandemic by WHO [5] has infected more than 73,973,280 persons around the world and killed more than 1645260 human beings. Till now, no vaccine has been given for the COVID - 19. In light of this fact, there are only the precautionary measures such as social distancing, hands sanitizers, face masks and PPE kits that one should take to control the spread of coronavirus. The primary objective of this article is to design a mechanism for predicting the coronavirus disease. Therefore, the present paper highlights various data mining techniques to predict the positive cases of COVID – 19.

II. COMMON SYMPTOMS OF CORONAVIRUS

The incubation period of virus COVID – 19 range from 2 – 14 days and normally around 5 days. Most common symptoms [4] of coronavirus are:

- Fever
- Tiredness
- Cough
- Nasal congestion
- Runny nose
- Sore throat
- Diarrhoea
- Conjunctivitis
- Loss of smell and taste
- Rashes on skin
- Body ache and pain
- Bluish lips or face
- Loss of speech
- Respiratory Problem

Persons above 60 and other with underlying ailments such as diabetes, hypertension and respiratory diseases are more likely to develop severe problems.

III. STATISTICS OF COVID – 19

Coronavirus or COVID – 19 are a large family of viruses that are common in animals. Rarely, people get infected with these viruses which may then affect other persons. For example, SARS – CoV is related with civet cats and MERS – CoV is associated with dromedary camels. The sources of coronaviruses have not yet been confirmed as scientists are working on this area. It is declared as pandemic by World Health Organization [5] on March 11, 2020 and on March 13, 2020, The President of USA declared coronavirus outbreak a national emergency [6] and there are fears that it can moving towards the situation of catastrophic global economic crisis. In table I, data of confirmed cases of COVID – 19 along with deaths is given.

On January 20, 2020 total 282 cases of coronavirus were reported to WHO and on December 16, 2020 total number of cases in the world are 73,973,280. Table I clearly depicts the

total number of active cases and deaths in 10 worst affected countries in the world.

Table I: Coronavirus Cases and Deaths by Country

Serial No.	Country Name	Confirmed Cases	Deaths
1	USA	17151872	311151
2	India	9933906	144141
3	Brazil	6974258	182854
4	Russia	2734454	48664
5	France	2391447	59072
6	Turkey	1898447	16881
7	UK	1888116	64908
8	Italy	1870576	65857
9	Spain	1771488	48401
10	Argentina	1510203	41204
11	**World**	**73973280**	**1645260**

Source: worldometers.info - December 16, 2020

IV. NAÏVE BAYES CLASSIFIER FOR THE PREDICTION OF COVID-19

Naïve Bayes seems a simple classification algorithm but in case of predictive modeling [7], it performs an impressive role. It is covered under the family of probabilistic classifiers which is entirely based on Bayer's theorem. Basically, Naïve Bayes is a model of conditional probability. With this model, it is possible to classify an instance of given problem which is denoted by a vector X= (x_1, …, x_n) where n represents features of independent variables. A group of cases were taken in to consideration and programmed with data sets.

The probabilities were calculated on all the classes and with all the conditions. Results were accumulated and when the test data was provided, we receive the probabilities for various classes on the basis of provided symptom details. The details can be used to categorize the patient in to the class with the probability. On the basis of value of the probability, it can be decided that a person is suffering with COVID – 19 or not. Traditionally, viruses were classified and differentiated by culture, serological and electron microscopy [8]. Applying these phenotypic methods, coronaviruses were termed as enveloped viruses of 120-160 nm in size with crown shape [9]. Coronaviruses are classified in to three categories. Under group 1 and 2, there are mammalian coronaviruses and group 3 covers avian coronaviruses. Different diagnose mechanisms are available to detect the novel coronavirus [10] such as ORF1ab and N, RdRP, E,N, Three targets in N gene, Two targets in RdRP and RT-PCR. Normally, tests results are available after 12 hours. As coronavirus is showing exponential growth in various parts of the world, at various remote and high altitude areas, the facility of diagnosis is difficult to provide or it take time of 48-72 hours to complete the process of diagnosis of patient. Therefore, in this article with the help of Naïve Bayes classifier [11], a mechanism is designed to predict the positive case of coronavirus. All the patients will be diagnosed on the basis of attributes [12] taken in the table II. When patient notice symptoms discussed in table II, it can be predicted with the mechanism that the patient has the COVID – 19 or not.

Table II: Attributes for the Detection of Coronavirus in a Suspect

Sore Throat and Cough	Runny Nose	Fever	Difficulty in Breathing	COVID - 19?
Yes	No	Yes	Mild	Yes
Yes	Yes	No	No	No
Yes	No	Yes	Strong	Yes
No	Yes	Yes	Mild	Yes
No	No	No	No	No
No	Yes	Yes	Strong	Yes
No	Yes	No	No	No
Yes	Yes	yes	Strong	Yes

The working of this mechanism is discussed in much detail with the help of flowchart in figure I. The flowchart depicts the operations in Naïve Bayer classifier algorithm. By applying this flowchart, we can conveniently predict [13] that the patient has been suffered from novel coronavirus or not. Initially, we calculate all possible individual probabilities applied on the target attribute of coronavirus containing all probabilities of attribute of coronavirus. Probabilities are calculated in the following manner.

P (Coronavirus = Y) = 5/8 = 0.625

P (Sore throat and cough = Y | Coronavirus = Y) = 3/5 = 0.6

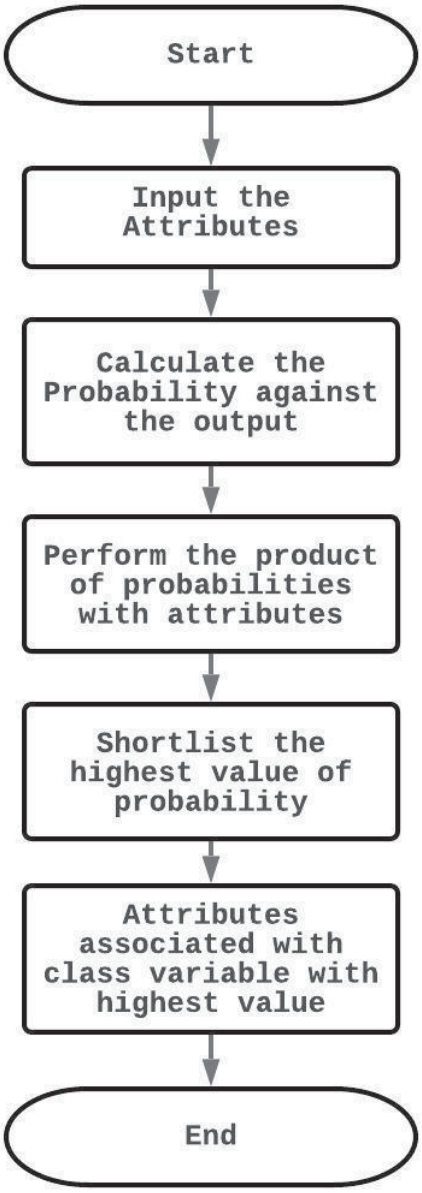
P (Coronavirus = N) = 3/8 = 0.375

P (Sore throat and cough = Y | Coronavirus = N) = 1/3 = 0.333

P (Runny nose = Y | Coronavirus = Y) = 3/5 = 0.6

P (Runny nose = Y | Coronavirus = N) = 1/3 = 0.333

Figure I: Working of the mechanism



In the similar manner, the above criterion is used to calculate the possible probabilities [14] for all the discussed conditions. These probabilities are demonstrated in table III.

Table III: Probabilities Calculated for each symptom

P (Coronavirus = Y)	0.625	P (Coronavirus = N)	0.375
P (Sore throat and cough = Y | Coronavirus = Y)	0.666	P (Sore throat and cough = Y | Coronavirus = N)	0.333
P (Sore throat and cough = N | Coronavirus = Y)	0.4	P (Sore throat and cough = N | Coronavirus = N)	0.666
P (Runny nose = Y | Coronavirus = Y)	0.666	P (Runny nose = Y | Coronavirus = N)	0.666
P (Runny nose = N | Coronavirus = Y)	0.4	P (Runny nose = N | Coronavirus = N)	0.333
P (Fever = Y | Coronavirus = Y)	1	P (Fever = Y | Coronavirus = N)	-------
P (Fever = N | Coronavirus = Y)	------	P (Fever = N | Coronavirus = N)	1
P (Difficulty in Breath = Strong | Coronavirus = Y)	0.6	P (Difficulty in Breath = Strong | Coronavirus = N)	-------
P (Difficulty in Breath = Mild | Coronavirus = Y)	0.4	P (Difficulty in Breath = Mild | Coronavirus = N)	-------
P (Difficulty in Breath = N | Coronavirus = Y)	------	P (Difficulty in Breath = N | Coronavirus = N)	1

In the above table, we disclosed that the probability P has classified in to two classes Yes and No.

P1:

Argmax P (Coronavirus = Y) * P (Sore throat and cough = Y | Coronavirus = Y) * P (Runny nose = N | Coronavirus = Y) * P (Fever = Y | Coronavirus = Y) * P (Difficulty in Breath = Mild | Coronavirus = Y)
0.625*0.666*0.4*1*0.4
= 0.0666

P2:
Argmax P (Coronavirus = N) * P (Sore throat and cough = Y | Coronavirus = N) * P (Runny nose = N | Coronavirus = N) * P (Fever = Y | Coronavirus = N) * P (Difficulty in Breath = Mild | Coronavirus = N)
0.375*0.333*0.333
= 0.0416

The argument of probability of P2 seems less than P1. Hence, the patient is not suffering with the COVID – 19. Different types of combinations can be used as per the situation of the symptoms [15] found in the patient and on the basis of result of probability, a decision can be taken that a person is suffering with coronavirus or not.

V. CONCLUSIONS

The data mining techniques can be implemented in association with Naïve Bayes classifier algorithm. Clearly the collaboration of these tools performs a prominent role in diagnosing the coronavirus infection or COVID – 19 which is declared as pandemic by WHO. The suggested mechanism showed impressive results which may lead to further refinements to utilize data mining, machine learning, artificial intelligence and information technology for diagnosing the patients for coronavirus. There is a dire need to find out both the cases of coronavirus, which are with the symptoms or asymptotic. It is only the way to break the chain and to control the spread of coronavirus. In future, the similar mechanism can be designed with more attributes and using other data mining tools.

References

[1] Warning issued on Middle East respiratory syndrome. *The Pharmaceutical Journal.* 2014 https://doi.org/10.1211/pj.2014.11138822

[2] Evans, M., & Bell, D. J. Severe Acute Respiratory Syndrome (SARS). *Oxford Medicine Online*. 2011 https://doi.org/10.1093/med/9780198570028.003.0046

[3] Ma, J. *China's first confirmed Covid-19 case traced back to November 17*. South China Morning Post. March 2020 https://www.scmp.com/news/china/society/article/3074991/coronaviruschina-first-confirmed-covid-19-case-traced-back

[4] Zhang, Yong, et al. "Gastrointestinal tract symptoms in coronavirus disease 2019: Analysis of clinical symptoms in adult patients." 2020.

[5] *Coronavirus confirmed as pandemic*. (11th March 2020). BBC News. https://www.bbc.com/news/world-51839944

[6] Dan Mangan, Christina Wilkie. *Trump declares national emergency over coronavirus.* (14th March 2020) CNBC. https://www.cnbc.com/2020/03/13/trump-will-hold-a-press-conference-at-3-pm-et-to-discuss-coronavirus-response.html

[7] Zakikhani, K., et al. "A Failure Prediction Model for Corrosion in Gas Transmission Pipelines." Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, 2020, p. 1748006X2097680, doi:10.1177/1748006x20976802.

[8] Fehr, Anthony R., and Stanley Perlman, "Coronaviruses: An Overview of Their Replication and Pathogenesis." Coronaviruses, 2015, pp. 1-23.

[9] Hu, Zhiwen, et al. "Nomenclature: Coronavirus and the 2019 Novel Coronavirus." 2020.

[10] A, Judy, and Ian Mackay. "Wuhan coronavirus (2019-nCoV) real-time RT-PCR ORF1ab 2020 (Wuhan-ORF1ab) v1 (protocols.io.bbsginbw)." *protocols.io*, 2020.

[11] Wood, Alexander, et al. "Private naive bayes classification of personal biomedical data: Application in cancer data analysis." Computers in Biology and Medicine, vol. 105, 2019, pp. 144-150.

[12] Salekin, Asif, and John Stankovic. "Detection of Chronic Kidney Disease and Selecting Important Predictive Attributes." 2016 IEEE International Conference on Healthcare Informatics (ICHI), 2016.

[13] S. Raj, Jennifer, and Vijitha Ananthi J. "Recurrent Neural Networks and Non linear Prediction in Support Vector Machines ." Journal of Soft Computing Paradigm, vol. 2019, no. 1, 2019, pp. 33-40.

[14] "New Artificial Intelligence Software to Help Determine Probability of Contracting Coronavirus." *WION*, 14 Dec. 2020, www.wionews.com/technology/new-artificial-intelligence-software-to-help-determine-probability-of-contracting-coronavirus-349728.

[15] *The Washington Post*, 2 Dec. 2020, https://www.washingtonpost.com/health/2020/12/02/covid-symptoms

Bibliography of Authors

Sugandh Bhatia is currently working as Assistant Professor of Computer Science in Punjab School of Economics, Guru Nanak Dev University, Amritsar – 143005, India and pursuing his PhD in the Faculty of Engineering and Technology, Guru Nanak Dev University, Amritsar. His field of research interests includes Cloud Computing, Cloud Forensics, Machine Learning, Security and Privacy.

Jyoteesh Malhotra is Associate Dean of Academics, Student Welfare and Professor and Head in the Department of Electronics and Communication Engineering and Department of Computer Science and Engineering in the Regional Campus Jalandhar of Guru Nanak Dev University, Amritsar. He received PhD, M.Tech (Gold Medalist) and has more than 20 years of experience of teaching and research. He has more than 175 publications in journals of International and National repute in his credit. His research interests include Wireless Networks and Optical Communication.