International Journal of Advanced Science and Technology
Vol. 29, No. 12s, (2020), pp. 2599-2604

Machine Learning-Based IoT-Enabled Perspective Model for Prediction of COVID-19 Test in Early Stage

Rajesh N. Phursule[1], Parikshit N. Mahalle [2] , Prakash K. Ukhalkar[3], Satish R. Todmal[4]
[1] *Associate Professor, Department of Information Technology, Pimpri Chinchwad College of Engineering, Pune, Maharashtra, India*

[2] *Senior Member IEEE, Professor and Head, Department of Computer Engineering, STES'S Smt. Kashibai Navale College of Engineering, Pune, Maharashtra, India*

[3] *Assistant Professor, Department of Master of Computer Application, Pimpri Chinchwad College of Engineering, Pune, Maharashtra, India*

[4] *Associate Professor and Head, Department of Computer Engineering,, Imperial College of Engineering and Research, Pune, Maharashtra, India*

[1]*rphursule@gmail.com*, [2]*aalborg.pnm@gmail.com*, [3]*prakshu.mca@gmail.com*, [4]*srtodmal@gmail.com*

Abstract

COVID-19 is one of pandemic community disease. Cycle of COVID-19 is 1 to 14 days. The most widely recognised symptoms of COVID-19 are fever, dry Cough, difficulty in breathing and diarrhea. The safety of the health workers is an important issue. Equipments required for COVID-19 test and certified laboratories where tests are conducted are limited. This paper proposed a perspective model for the prediction of COVID-19 test at early stage based on symptoms. In this perspective model, input sounds of cough are recorded by Smart Voice Recorder and temperature by Bluetooth Thermometer. Classifiers such as random forest tree classifier, decision tree classifier can be used to predict the susceptible population for COVID -19 test on the basis of symptoms. Proposed perspective model will be helpful to reduce the human intervention and use of testing equipments for most infected population.

Keywords*-COVID-19, Decision Tree Classification, Machine Learning, Early Stage Detection*

1. Introduction

Machine learning algorithms are extensively utilized in the clinical field for grouping information into various classes according to feature similarity. The first pneumonia case is identified in Wuhan city of China on 8th December 2019. The pathogen has identified it as Severe Acute Respiratory Syndrome. Novel corona virus is identified by the Chinese authorities in January 2020[1]. This virus is responsible for an outbreak of respiratory disease known as COVID -19. World Health Organization mentioned Corona Virus Disease 2019 (COVID - 19) a general well being crisis of universal concern [2]. The most widely recognised symptoms of COVID-19 are fever, dry Cough, difficulty in breathing and diarrhea. According to the survey in China till 29th January 2020, COVID-19 infected persons have cough (67.8%), fever (43.8%) and diarrhea (3.8%) [1]. As per the above result, cough and fever are important symptoms of COVId-19. For diagnostic test of COVID-19,standard molecular methods are used to detect the presence of SARS-CoV-2 in respiratory samples are time specific RT-PCR [3] which target RNA dependant RNA polymerase and E- genes [3][4]. These tests are time consuming. The COVID-19 infected population is very low who have been tested and equipments required in certified laboratories for such test are costly [5]. So these tests cannot be

performed to all susceptible persons. According to India today survey on 13th April only 5% of 2 lacs persons in India are positive [6].

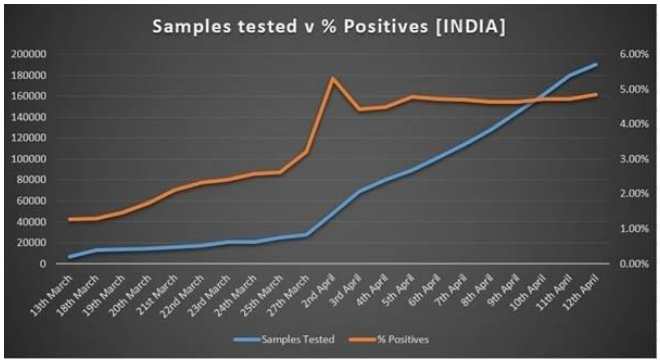

Figure1: Tests vs Positive % in India

Hence a decision making classifier is needed to classify the susceptible population into two classes as Probable and Threatened. These are explained in detail in Mathematical Model section - IV.

Most of the doctors as well as other health care workers throughout the world are getting infected while giving services to the early stage COVID-19 susceptible population. Since patients are came to clinics with normal cold and fever symptoms. Later they may get detected as COVID -19 positive. Doctors should not come in physical contact with the person have symptoms cough and fever at early stage of symptoms. It is possible through IoT and Machine Learning.

This paper proposes a perspective model which can be used to predict the susceptible population for COVID-19 test at early stage on the base of symptoms. It will also helpful to reduce the human intervention with the COVID -19 patients at early stage with the help of IoT and Machine Learning.

2. Problem Statement

The safety of the health workers is an important issue. Equipments required for COVID-19 test and certified laboratories where tests are conducted are limited.

This paper proposes the perspective model to predict susceptible population of COVID-19 for COVID-19 test and reduce the human intervention with the COVID -19 patients at early stage.

3. Related Work

A SIDARTHE model of COVID-19 epidemic in Italy is proposed to discriminate between detected and undetected cases whether they are symptomatic or asymptomatic. It also discriminates between COVID-19 patients who have required a normal treatment and the patients who have required ICU [3]. The average age of the patient is 47 years where 41.9 % patients are female. The most widely recognised symptoms are fever and dry cough in COVID- 19 patients. 67.8% patients have coughed and 43.8% patients have fever [1]. The epidemic model of COVID-19 of china suggested government restriction on pubic to reduce the severity of epidemic. The announced and unreported cases are important in interpreting the damage. Asymptomatic infection cases are very much important in disease transformation [8]. Ensemble base classifier classifies the data more accurately than the individual classifier [9].

4. Mathematical Model

Proposed method is summarized as below in the form of state model diagram.

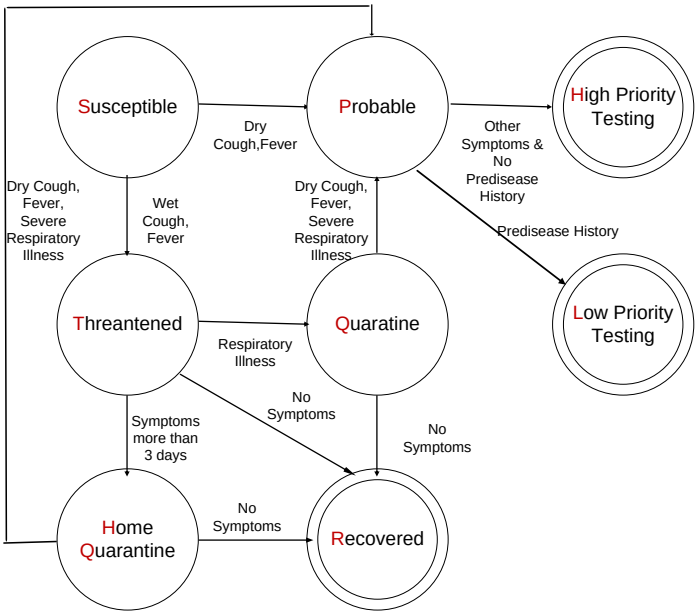

Figure 2: State Model Diagram

The states are as follows:

- S (Susceptible) - Susceptible Population having fever and cough/shortness of breath or having above symptoms and in contact with confirmed COVID-19 case.
- P (Probable) – Population considered for COVID-19 test having symptoms of Dry Cough and Fever.
- T (Threatened) – Population having wet cough and fever are threatened due to said symptoms.
- H (High Priority Testing) – Probable population will go for COVID-19 test with high priority since having other symptoms like diarrhea, strong shortness of breath, head ache etc and no pre-disease history.
- L (Low Priority Testing) - Probable population will go for COVID-19 test with low priority having pre-disease history such as asthma, diabetics or heart disease.
- Q (Quarantine in Hospital or Reserved Place) – Population from threatened state having increased shortness in breathing.
- HQ (Home Quarantine) – Population from threatened having symptoms of wet cough, fever more than 3 days but no respiratory illness.
- R (Recovered) - Population having no any symptoms.

Susceptible population either go for High Priority Testing, Low Priority Testing or Recovered state.

5. Proposed Methodology

Dry cough and Fever are the main symptoms of COVID-19. Dry cough doesn't bring up mucus like wet cough. It occurs due to irritation in respiratory tract, but no excess mucus to cough up. Wet

cough sounds wet because body is pushing mucus out of the respiratory tract. It sounds soupy and may come with wheezing or rattling sound.

In a proposed model, input sounds of cough are recorded by Smart Voice Recorder and temperature by Bluetooth Thermometer. By using internet, these are given as an input to any classifier like Random Forest decision tree classifier. It is an ensemble base classifier. Classification is a supervised machine learning technique. It uses a divide and conquers technique. Starting with root node, it splits the dataset into subsets. The core algorithm for decision tree classifier is ID3. It uses the entropy and information gain.

$$E(C,A) = \Sigma_{a \varepsilon A} = P(a)\, E(a) \quad (1)$$

Where,
C = Target Class, A = Attributes, P(a) = Probability of Attributes, E(a) = Entropy of Attributes and E(C,A) = Entropy of target class with attribute A.

Information Gain is

$$Gain(C,A) = Entropy(C) - Entropy(C,A) \quad (2)$$

The ID3 algorithm is run recursively on non-leaf nodes, until all data is not classified. Leaf nodes are target class with zero entropy. The Decision tree classifier is trained with training data set to classify Wet and Dry cough. Once wet and dry cough are identified, decision tree will be built as shown in Figure 3.

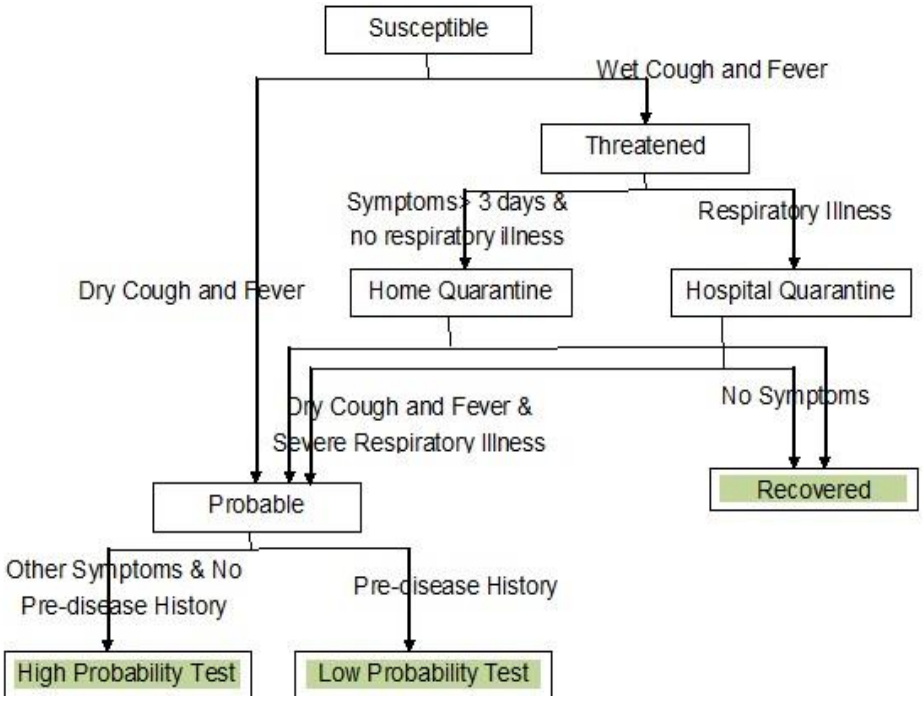

Figure 3: Decision Tree

6. Result Analysis

Proposed perspective model is helpful to distinguish the susceptible population of COVID-19 into Probable for Testing and Threatened. Probable population for testing of COVID-19 further divided into High Priority Testing and Low Priority Testing. If patients have pre-disease history like diabetes,

heart disease, then their priority of testing is low otherwise priority of testing is high. Threatened population gone for either Home Quarantine or Hospital Quarantine. If the cough and fever remain more than three days and respiratory illness is not observed, then they are quarantined in home. If respiratory illness is observed, then they are quarantined in hospitals or reserved place. The population form Home Quarantine or Hospital/Reserved Place Quarantine are either Recovered or Probable for testing. If symptoms are not detected then these population is Recovered. If symptoms of Cough and Fever remain as it is and respiratory illness is severe, then Quarantined population is gone for testing. Initial symptoms cough and fever are captured through smart IoT enabled devices. Random forest decision tree classifier and decision is used to decide whether it will go for testing of COVID-19. No official figure, but a lot of health workers are infected due to COVID-19. Chances of infection of COVID-19 to health care workers are reduced as there is no intervention of them throughout the whole process. The COVID-19 infected population is very low who have been tested in India. It is approximately 5%. But proposed method divides susceptible population into threatened and probable for testing. Since test equipments are costly and limited, equipments of certified laboratories can be used for highly probable patients.

7. Conclusion

The proposed perspective model can be used to predict susceptible population for COVID-19 test by using machine learning techniques. Infections of health care workers can be reduced, as no participation of health care workers in the proposed model. Recording of symptoms are done by using smart devices and IoT.

8. Future Scope

This paper proposed an IoT enabled, machine learning based perspective model. It has a scope to predict susceptible population for COVID-19 premature detection. This perspective model can be extended for COVID-19 susceptible population real-time dataset and stated machine learning algorithms in future. Different classifiers are suggested for the prediction of susceptible population for COVID-19 test. Other machine learning techniques, for example clustering, prediction analysis can be utilized in future.

References

1. W. Guan, Z. Ni, Yu Hu, W. Liang, C. Ou, J. He, L. Liu, H. Shan, C. Lei, D. S. C. Hui, B. Du, L. Li, G. Zeng, K. Yuen, R. Chen, C. Tang, T. Wang, P. Chen, J. Xiang, S. Li, J. Wang, Z. Liang, Y. Peng, L. Wei, Y. Liu, Y. Hu, P. Peng, Jian-ming Wang, J. Liu, Z. Chen, G. Li, Z. Zheng, S. Qiu, J. Luo, C. Ye, S. Zhu, N. Zhong, “Clinical Characteristics of Coronavirus Disease 2019 in China", N. Engl.J.Med.,2020,doi:10.1056/NEJMoa2002032.
2. World Health Organization. Coronavirus disease(COVID-19) outbreak (https://www .who .int).
3. Giulia Giordano, Franco Blanchini, Raffaele Bruno, Patrizio Colaneri, A. Di Filippo, A. Di Matteo, Marta Colaneri, ”A SIDARTHE Model of COVID-19 Epidemic in Italy”,preprint arXiv:2003.09861(2020).
4. Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DKW. “Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR", Euro Surveill., 2020, 10.2807/1560-917.ES.2020.25.3.2000045.
5. J. Gallagher, \Are coronavirus tests awed?", BBC NEWS, 13 February 2020.
6. https://www.indiatoday.in/news-analysis/story/decoding-india-s-covid-19-testing-state-by-state-1666587-2020-04-13.

7. WHO. Coronavirus disease 2019 (COVID-19): Situation Report - 85. Covid-19 Situational Reports, 2020.
8. Z. Liu, P. Mangal, O. Seydi, G. Webb, " Predicting the cumulative number of cases for the COVID-19 epidemic in China from early data", preprint arXiv:2002:12298v1(2020).
9. Rajesh Phursule, " Performance Behaviour of Intrusion Detection Systems (Ids) Based on Ensemble Base Classifier (EBC)" International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume-8 Issue-11, September 2019.
10. Chuanyi Wang, Zhe Cheng, Xiao –GuangYue , Michael McAleer," Risk Management of COVID-19 bu Universities in China" Journal of Risk and Financial Management, February 2020
11. Kimberlyn Roosa, Yiseul Lee, Ruiyan Luo, Alexander Kirpich, Richard Rothenberg, James M Hyman, Ping Yan, Gerardo Chowell, " Short-term Forecasts of the COVID-19 Epidemic in Guangdong and Zhejiang, China: February 13-23, 2020" Journal of Clinical Medicine, February 2020.

Authors

Dr Rajesh N Phursule
Ph D (Computer Science and Engineering)
Area of Interest – Data Mining, Machine Learning, Data Science, Internet of Things

Dr Parikshit N Mahalle
Ph D (Computer Science and Engineering)
Senior member IEEE, Member ACM, Member IAENG (H.K.)
Area of Interest – Internet of Things, Block Chain Technology, Machine Learning

Mr Prakash K Ukhalkar
MCA
Area of Interest – Cloud Computing, Machine Learning, Data Science, Business Intelligence

Dr Satish R Todmal
Ph D (Computer Science and Engineering)
Area of Interest – Image Processing,
Cloud Computing, Software Engineering