Identification of COVID-19 can be quicker through artificial intelligence framework using a mobile phone–based survey when cities and towns are under quarantine

Arni S. R. Srinivasa Rao PhD1-2,3 and Jose A. Vazquez MD4

1Division of Health Economics and Modeling, Department of Population Health Sciences, Medical College of Georgia, Augusta University, Augusta, Georgia, 2Laboratory for Theory and Mathematical Modeling, Division of Infectious Diseases, Department of Medicine, Medical College of Georgia, Augusta, Georgia, 3Department of Mathematics, Augusta University, Augusta, Georgia and 4Division of Infectious Diseases, Department of Medicine, Medical College of Georgia, Augusta University, Augusta, Georgia

Abstract

We propose the use of a machine learning algorithm to improve possible COVID-19 case identification more quickly using a mobile phone–based web survey. This method could reduce the spread of the virus in susceptible populations under quarantine.

(Received 21 February 2020; accepted 25 February 2020; electronically published 3 March 2020)

Emerging and novel pathogens are a significant problem for global public health. This is especially true for viral diseases that are easily and readily transmissible and have asymptomatic infectivity periods. The novel coronavirus (SARS-CoV-2) described in December 2019 (COVID-19) has resulted in major quarantines to prevent further spread, including major cities, villages, and public areas throughout China and across the globe.1–3

As of February 25, 2020, the World Health Organization’s situational data indicate ~77,780 confirmed cases in 25 countries, including 2,666 deaths due to COVID-19.4 Most deaths reported so far have been in China.5 The Centers for Disease Control and Prevention (CDC) and the World Health Organization have issued interim guidelines to protect the population and to attempt to prevent the further spread of the SARS-CoV-2 virus from infected individuals.6 Cities and villages throughout China are unable to accommodate such large numbers of infected individuals while maintaining the quarantine, and several new hospitals have been built to manage the infected individuals.7 It is imperative that we evaluate novel models to attempt to control the rapidly spreading SARS-CoV-2.8 Technology can assist in faster identification of possible cases to yield more timely interventions.

To reduce the time needed to identify a person under investigation (PUI) for COVID-19 and their rapid isolation, we propose to collect a basic travel history along with the more common signs and symptoms using a mobile phone–based online survey. Such data can be used in the preliminary screening and early identification of possible COVID-19 cases. Thousands of data points can be processed through an artificial intelligence (AI) framework that can evaluate individuals and stratify them into no risk, minimal risk, moderate risk, and high risk groups. The high-risk cases identified can then be quarantined earlier, thus decreasing the chance of spreading the virus (Table 1).

The signs and symptoms data recorded in step 5 of the algorithm are collected prior to Health Check Recommended for Coronavirus (HCRC) alerts or Health Check Recommended for Coronavirus (MHCRC) alerts (for possible identification and assessment) and No Health Check Recommended for Coronavirus (NCR) alerts (for nonidentified respondents). These procedures are explained in steps 3 and 4 in Appendix 2. The extended analysis we propose can help determine any association among sociodemographic variables and the signs and symptoms, such as fever and lower respiratory infection including cough and shortness of breath, in individuals with and without possible infection. A 2 x 2 table of number of COVID-19 cases identified through AI and the number of people responded to a mobile survey is described in Figure 2.
Applications of AI and deep learning can be useful tools in assisting diagnoses and decision making in treatment.\textsuperscript{10,11} Several studies have promoted disease detection through AI models.\textsuperscript{12–15} The use of mobile phones\textsuperscript{16–19} and web-based portals\textsuperscript{20,21} have been tested successfully in health-related data collection. In addition, our proposed algorithm can be easily extended to identify individuals who might have any mild symptoms and signs. However, such techniques must be applied in a timely way for relevant and rapid results. Apart from cost-effectiveness, our proposed modeling method could greatly assist in identifying and controlling COVID-19 in populations under quarantine due to the spread of SARS-CoV-2.

**Table 1.** Steps involved in the collection of data through a mobile phone-based survey

| Step | Description |
|------|-------------|
| 1    | Record the location details of the house/apartment from where the respondent uses a phone-based web survey/or the respondent’s usual place of stay. |
| 2    | Record demographic information like gender (G) (1-male, 2-female, 3-others), age (A), race (R) (1-white, 2-black, 3-Hispanics, 4-Others) |
| 3    | Have you traveled to (or living in) any of the COVID-19 affected areas/countries in the last 14 days? (Yes=1/No=0) |
| 4    | Have you had any close contact with a person who is known to have COVID-19 during the last 14 days? (Yes=1/No=0) |
| 5    | Record the presence or absence of signs and symptoms listed below and the duration of each of the signs and symptoms if yes to any of the signs and symptoms. |
|      | A) fever (Yes=1/No=0), if yes, then the duration in days —— |
|      | B) cough (Yes=1/No=0), if yes, then the duration in days —— |
|      | C) shortness of breath (Yes=1/No=0), if yes, then the duration in days —— |
|      | D) myalgia or fatigue (Yes=1/No=0), if yes, then the duration in days —— |
|      | E) sputum production (Yes=1/No=0), if yes, then the duration in days —— |
|      | F) headache (Yes=1/No=0), if yes, then the duration in days —— |
|      | G) diarrhea (Yes=1/No=0), if yes, then the duration in days —— |
|      | H) pneumonia in both lungs (Yes=1/No=0), if yes, then the duration in days —— |
| 6    | Enter the details of steps 1-5 above for any dependents or other individuals who live in the same location and do not have access to web-based survey. |

**Fig. 1.** Conceptual framework of data collection and possible COVID-19 identification. (a) A geographical region (eg, a city, county, town, or village) with households in it. (b) Respondents and nonrespondents of a mobile phone-based web survey. (c) Possible identified cases of COVID-19 among the survey respondents and possible cases of COVID-19 among nonrespondents of the survey.
Acknowledgments. We thank Professor N.V. Joshi, Indian Institute of Science, Bengaluru, and Mr P. Sashank, CEO Exaccto CompuSoft Global Solutions, Hyderabad, India, for their editorial comments.

Financial support. No financial support was provided relevant to this article.

Conflicts of interest. All authors report no conflicts of interest relevant to this article.

Authors contributions. ASRSR designed the study, developed the methods and wrote the first draft of the paper. JAV contributed in clinical verbiage editing, inputs and editing into the draft.

References

1. More Chinese cities shut down as novel coronavirus death toll rises. Channel News Asia website. https://www.channelnewsasia.com/news/asia/wuhan-coronavirus-more-china-cities-shut-hangzhou-zhejiang-hubei-12395706. Published February 5, 2020. Accessed February 10, 2020.

2. Weinland D, Yu S. Chinese villages build barricades to keep coronavirus at bay. The Financial Times website. https://www.ft.com/content/68792b9c-476e-11ea-aeb3-955839e06441 Published February 7, 2020. Accessed February 10, 2020.

3. Transit going in and out of Wuhan, China, is being shut down to contain coronavirus. Business Insider website. https://www.businessinsider.com/transit-wuhan-china-shut-down-coronavirus-2020-1. Published January 25, 2020. Accessed February 10, 2020.

4. The WHO COVID-19 situation report 36. World Health Organization website. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200223-sitrep-36-covid-19.pdf?sfvrsn=2791b4e0_2. Published February 25, 2020. Accessed February 26, 2020.

5. Coronavirus explained: All your questions about COVID-19 answered. C Net website. https://www.cnet.com/how-to/coronavirus-explained-all-your-questions-about-covid-19-answered/. Updated March 24, 2020. Accessed March 27, 2020.

6. Preventing the spread of coronavirus disease 2019 in homes and residential communities. Centers for Disease Control and Prevention website. https://www.cdc.gov/coronavirus/2019-ncov/hcp/guidance-prevent-spread.html. Updated March 6, 2020. Accessed March 27, 2020.

7. Wang J, Zhu E, Umlauf T. How China built two coronavirus hospitals in just over a week. The Wall Street Journal website. https://www.wsj.com/articles/how-china-can-build-a-coronavirus-hospital-in-10-days-11580397751. Published February 6, 2020. Accessed February 10, 2020.

8. Expert: better models, algorithms could help predict and prevent virus spread. The Augusta Chronicle website. https://www.augustachronicle.com/news/20200128/expert-better-models-algorithms-could-help-predict-and-prevent-virus-spread. Published January 28, 2020. Accessed on February 11, 2020.

9. Flowchart to identify and assess 2019 novel coronavirus. Centers for Disease Control and Prevention website. https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-criteria.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fhcp%2Fidentify-assess-flowchart.html. Updated February 27, 2020. Accessed March 27, 2020.

10. Liang H, Tsui BY, Ni H. et al. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. Nat Med 2019;25:433–438.

11. Rao ASRS, Diamond MP. Deep learning of Markov model-based machines for determination of better treatment option decisions for infertile women. Reprod Sci 2020;27:763–770.

12. Nell DB. Using artificial intelligence to improve hospital inpatient care. IEEE Intell Syst 2013;28:92–95.

13. Rajalakshmi R, Subashini R, Anjana RM, et al. Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence. Eye 2018;32:1138–1144.

14. Zeinab A, Roohallah A, Mohamad R, Hossein M, Ali AY. Computer-aided decision making for heart disease detection using hybrid neural-network genetic algorithm. Comput Methods Programs Biomed 2017;141:19–26.

15. Kumar VR, Kumar SS, Saboo V. Dermatological disease detection using image processing and machine learning. Third International Conference on Artificial Intelligence and Pattern Recognition (AIPR), Lodz; 2016:1–6.

16. Tomlinson M, Solomon W, Singh Y. et al. The use of mobile phones as a data collection tool: a report from a household survey in South Africa. BMC Med Inform Decision Making 2009;9:51.

17. Ballivian A, Azvedo JP, Durbin W. 2015. Using mobile phones for high-frequency data collection. In: Toninelli D, Pinter R, de Pedraza P, eds. Mobile Research Methods: Opportunities and Challenges of Mobile Research Methodologies. London: Ubiquity Press; 2015:21–39.

18. Braun R, Catalani C, Wimbush J, Israelski D. Community health workers and mobile technology: a systematic review of the literature. PLoS One 2013;8(6):e67772.

19. Bastawrous A, Armstrong MJ. Mobile health use in low- and high-income countries: an overview of the peer-reviewed literature. J Roy Soc Med 2013;106:130–142.

20. Paolotti D, Carnahan A, Colizza V, et al. Web-based participatory surveillance of infectious diseases: the Influenzanet participatory surveillance experiment. Clin Microbiol Infect 2014;20:17–21.

21. Fabic MS, Choi YJ, Bird S. A systematic review of demographic and health surveys: data availability and utilization for research. Bull World Health Org 2012;90:604–612.

Appendix 1. Steps Involved in Data Collection Through Mobile Phones

We have developed our data collection criteria based on the CDC’s Flowchart to Identify and Assess 2019 Novel Coronavirus, and we have added additional variables for the extended utility of our efforts in identifying infected and controlling the spread (see Table 1 in the text).

![Flowchart to Identify and Assess 2019 Novel Coronavirus](https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-criteria.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fhcp%2Fidentify-assess-flowchart.html. Updated February 27, 2020. Accessed March 27, 2020.)
Appendix 2. Algorithm

Let \( O_1, O_2, O_3, O_4, O_5 \) be the outputs recorded during the data collection steps 1 through 5 described in the Appendix 1. The 3 outputs within \( O_2 \) are given as

\[ O_2 = \{O_{2G}, O_{2A}, O_{2R}\}, \]

and 9 pairs of outputs within \( O_5 \) are given as

\[ O_5 = \left\{ \left( O_{5A}, D_{5A} \right), \left( O_{5B}, D_{5B} \right), \left( O_{5C}, D_{5C} \right), \left( O_{5D}, D_{5D} \right), \left( O_{5E}, D_{5E} \right), \left( O_{5F}, D_{5F} \right), \left( O_{5G}, D_{5G} \right), \left( O_{5H}, D_{5H} \right), \left( O_{5I}, D_{5I} \right) \right\}, \]

where the pair \( (O_{5i}, D_{5i}) \) for \( i = A, B, \ldots I \) represents the respondent’s response regarding the presence or absence of \( i^{th} \) sign and symptom \( O_{5i} \) and duration of corresponding sign and symptom \( D_{5i} \).

1. If the set of identifiers, \( I_1 \), for

\[ I_1 = \{O_3, O_{5A}, O_{5B}, O_{5C}\} \]

is equal to one of the elements of the set \( C_1 \), for

\[ C_1 = \left\{ \begin{array}{c} (1, 1, 1, 1) \\
(1, 1, 1, 0) \\
(1, 1, 0, 0) \\
(1, 0, 1, 1) \\
(1, 0, 0, 1) \\
(1, 0, 1, 0) \\
(1, 1, 0, 0) \\
(1, 1, 0, 1) \end{array} \right\}, \]

for a respondent, then, send HCRC or MHCRC. If \( I_1 \) is not equal to any of the elements of the set \( C_1 \), then proceed to test criteria (3).

2. If the set of identifiers, \( I_2 \), for

\[ I_2 = \{O_4, O_{5A}, O_{5B}, O_{5C}\} \]

is equal to one of the elements of the set \( C_2 \), then send HCRC or MHCRC to that respondent, else proceed to the test criteria (4).

3. If \( I_1 \) is equal to one of the elements of the set \( C_2 \), for

\[ C_2 = \left\{ \begin{array}{c} (0, 1, 1, 1) \\
(0, 1, 1, 0) \\
(0, 1, 0, 0) \\
(0, 0, 1, 1) \\
(0, 0, 0, 1) \\
(0, 0, 1, 0) \\
(0, 1, 0, 1) \end{array} \right\}, \]

then the respondent will be sent an NCRC alert.

4. If \( I_2 \) is equal to one of the elements of the set \( C_2 \), then the respondent will be sent an NCRC alert.

A comparison of test criteria results of (3) and (4) with their corresponding geographic and sociodemographic details will yield further investigations of signs and symptoms based on whether or not an individual in the survey has traveled to coronavirus-affected areas or has had contact with any person who is known to have COVID-19. Here, we focus only on the identification of cases; further analysis techniques are beyond our scope. However, our approach is flexible enough to capture various other associations within the populations.

Appendix 3. Further Computations on the Data Collected

Suppose \( n \) and \( m \) are individuals in a region who have responded and not responded, respectively, for a mobile phone–based online survey. Responses are randomly associated and not depended on the sickness due to the virus. The pair

\[ \left( \frac{n}{n+m}, \frac{m}{n+m} \right) \]

yields the proportions of those who have responded and not responded in that region. Notably, we can compute \( \frac{m}{n+m} \) because the value \( m \) is known to us in that region. Here, \( n_1 \) of \( n \) are possible cases identified through our algorithm, and \( m_1 \) of \( m \) are possible cases of the virus that were not identified by the algorithm because \( m \) individuals never responded to the survey. Because \( n \) and \( m \) are known to us, one of the following relations will hold:

\[ \{n > m, n = m, n < m\} \] (A2.1)

Thus, we will see which of the relations listed in (A2.1) is true. When \( n > m \), one of the following relations will hold:

\[ \left\{ \frac{n_1}{n} > \frac{m_1}{m}, \frac{n_1}{n} = \frac{m_1}{m}, \frac{n_1}{n} < \frac{m_1}{m} \right\} \] (A2.2)

However, we will never know which of the relations in (A2.1) is true because \( m_1 \) were never identified by the algorithm. For example, suppose 2,000 individuals respond to the survey, and of these, 500 individuals do not respond to the survey and 400 are identified as possible cases by the algorithm. If there are 100 possible cases of virus (which we do not have a mechanism to count) among the 500 who never responded, then the relation

\[ \frac{n_1}{n} = \frac{m_1}{m} \]

is true. Similarly, other relations of (A2.2) could arise when \( n > m \).

Using a similar argument, we can verify that when other relations of (A2.1) are true, we are still unsure which of the relations in (A2.1) are true. The \( 2 \times 2 \) contingency options are provided in Figure 2 (in the text) to visualize the data to be generated using the proposed method.

Theorem: Let there be \( N \) individuals in a region. The probability that \( n_1 \) cases identified through the AI framework given that there are \( n \) individuals responded to the survey is \( \frac{n_1 N}{n^2} \).

Proof: Let \( N = n + m \), and let

\[ U = \{u_1, u_2, \ldots, u_n\} \]

be the collection of \( n \) individuals who responded,

\[ V = \{v_1, v_2, \ldots, v_m\} \]
be the collection of \( m \) individuals who did not responded. Suppose

\[ U_1 = \{ u_{a_1}, u_{a_2}, \ldots, u_{a_n} \} \subset U \]

is the collection of respondents who are identified as possible cases. Here \( U \cup V \) can be considered the region shown in (a), \( U \) shown in (b) and \( U_1 \) in (c) shown in Figure 1 (in the text).

Suppose we define 2 events \( E_1 \) and \( E \) using the sets \( U, V \) and \( U_1 \) as follows:

\( E_1: n_1 \) of \( n \) responded cases are identified through the algorithm

\( E: n \) of \( N \) have responded to the survey.

The conditional probability of the event \( E_1 \) given the event \( E \), say, \( P(E_1/E) \), is computed as follows:

\[
P(E_1/E) = \frac{P(E_1 \cap E)}{P(E)} = \frac{n_1}{N} = \frac{n_1N}{N^2}.
\]