A knowledge-based self-pre-diagnosis system to predict Covid-19 in smartphone users using personal data and observed symptoms

Duygu Çelik Ertuğrul | Demet Çelik Ulusoy

Abstract
Covid-19 is an acute respiratory infection and presents various clinical features ranging from no symptoms to severe pneumonia and death. Medical expert systems, especially in diagnosis and monitoring stages, can give positive consequences in the struggle against Covid-19. In this study, a rule-based expert system is designed as a predictive tool in self-pre-diagnosis of Covid-19. The potential users are smartphone users, healthcare experts and government health authorities. The system does not only share the data gathered from the users with experts, but also analyzes the symptom data as a diagnostic assistant to predict possible Covid-19 risk. To do this, a user needs to fill out a patient examination card that conducts an online Covid-19 diagnostic test, to receive an unconfirmed online test prediction result and a set of precautionary and supportive action suggestions. The system was tested for 169 positive cases. The results produced by the system were compared with the real PCR test results for the same cases. For patients with certain symptomatic findings, there was no significant difference found between the results of the system and the confirmed test results with PCR test. Furthermore, a set of suitable suggestions produced by the system were compared with the written suggestions of a collaborated health expert. The suggestions deduced and the written suggestions of the health expert were similar and the system suggestions in line with suggestions of the expert. The system can be suitable for diagnosing and monitoring of positive cases in the areas other than clinics and hospitals during the Covid-19 pandemic. The results of the case studies are promising, and it demonstrates the applicability, effectiveness, and efficiency of the proposed approach in all communities.

KEYWORDS
Covid-19, inferencing, knowledge-based medical expert systems, mobile diagnosing and monitoring, ontology, upper respiratory infection diseases

1 INTRODUCTION

A new public health crisis has threatened the world with the emergence and spread of the 2019 new coronavirus (2019-nCoV) called Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2). Covid-19 is the name given to the disease related with this virus which is an acute respiratory infection disease. The virus spreads rapidly from Wuhan City of Hubei Province of China to the rest of the world. According to recently published research studies, Covid-19 is observed mild in most people; in some (usually the elderly), it may evolve to pneumonia, Acute
Respiratory Distress Syndrome (ARDS) and multi organ dysfunction (Singhal, 2020). Since December 31, 2019 and as of December 09, 2020, total 67,965,261 Covid-19 cases and including 1,557,616 deaths have been reported (European Centre for Disease Prevention and Control, 2020).

According to health experts, it is a highly contagious virus that is transmitted by the infected people via droplets (e.g., sneezing and coughing), by air (e.g., much smaller than droplets) and even remains in the air. Therefore, health experts argue that hand washing, social distance, and self-isolation are crucial to slowing coronavirus transmission (WHO, 2020a).

Covid-19 clinical spectrum is quite broad, from no symptoms (asymptomatic) to severe pneumonia and death. Clinically, Covid-19 initially appears as a flu-like syndrome. Covid-19 patients usually show signs and symptoms, such as mild respiratory illness and persistent fever, an average of 5–6 days after infection (range, 1–14 days). The fever in Covid-19 patients is persistent, unlike with the progressive decline observed in cases of influenza (Lima, 2020; WHO, 2020b). Fever may not be observed in some cases such as those occurring in patients that are very young, elderly, or immunocompromised, as well as in those that have used antipyretic medication (Lima, 2020).

As of 20 February 2020 and based on 55,924 laboratory confirmed cases, typical signs and symptoms include: fever (87.9%), dry cough (67.7%), fatigue (38.1%), sputum production (33.4%), shortness of breath (18.6%), sore throat (13.9%), headache (13.6%), myalgia or arthralgia (14.8%), chills (11.4%), nausea or vomiting (5.0%), nasal congestion (4.8%), diarrhoea (3.7%), and hemoptysis (0.9%), and conjunctival congestion (0.8%) (WHO, 2020a). In addition, the symptoms of Covid-19 cases are very similar the observed symptoms in influenza, allergies, and other respiratory tract diseases. Influenza and Covid-19 are both contagious respiratory infections, but they are caused by different viruses (WHO, 2020c). It spreads through respiratory droplets, which are thrown when coughing, sneezing or talking loudly. Fever, cough and muscle aches can be seen in both diseases.

In fact, it is very difficult to separate these diseases without testing and absolutely possible to have both at the same time. Both diseases are more likely to cause serious illnesses in older individuals, especially, the ones with chronic medical problems, with those in pregnancy period, and individuals with impaired immunity. The most common and serious complication for both conditions is pneumonia. Therefore, influenza, allergies and other respiratory diseases can be confused by individuals since they are similar to Covid-19 symptoms, especially at the early stages. In such cases, individuals may experience serious suffer caused by anxiety and stress-related symptoms even if they are not Covid-19 (Hendricks, 2020). This situation brings people into the hospital environment and increases the risk of contamination and unnecessary stress and anxiety.

Mobile health (m-Health) applications are software tools that can be installed on mobile devices, usually smartphones. With the increasing use of technology, it has been proven that m-Health applications can spread health information and resources widely. The use of technological solutions in the health sector with the potential to make patients' lives easier has gradually increased with m-Health. M-Health systems help observe patients during treatment and provide services such as gathering medical data during the monitoring stage (Alepis & Lambrinidis, 2013; Baysari & Westbrook, 2015; Merrell & Doarn, 2014). Health experts can reach the data from their environment via internet and examine their patients’ medical history, the prompt symptoms observed, and medical responses to a treatment (Mohammadzadeh & Safdari, 2014). Thanks to smart m-Health systems with a strong knowledge base, instant inferencing of new relational medical data can be made based on medical data gathered for many purposes (Shridevi et al., 2018; Yin et al., 2018).

Semantic Web (SW) is a useful technology, which proposed in (Berners-Lee et al., 2001), to develop smart and rule-based medical expert systems. SW technology is known as an extension of the current Web and a subset of artificial intelligence in which information has a well-defined meaning. The semantic approach enables better cooperation between computers and people, and it is used to create “intelligent agents” that assist users answering their queries more precisely via ontologies.

Gruber (2008) described the ontology as a conceptual language of the SW that is a specification of a conceptualization of a knowledge domain. Indeed, ontology is a controlled vocabulary that formally describes concepts and their relationships. Recently, many ontology languages have been proposed and standardized, such as Resource Description Framework Schema (RDFS) (Lassila & Swick, 1999) and Web Ontology Language (OWL) (OWL Working Group, 2009). According to the World Wide Web Consortium (W3C), OWL is a family of knowledge representation languages for reasoning on ontologies. OWL expresses concepts in a machine-understandable form with specific spatial terms for a specific domain.

In this study, a knowledge-based self-pre-diagnosis system is designed to predict Covid-19 risk in smartphone users using their demographic and symptoms data. The system has its own an ontology knowledge-base and semantic rules and can be used as a predictive tool of Covid-19 risk in distant areas from clinics or hospitals.

The methodologies used and main technical contributions constitutes on three main aspects: (1) gathering demographic and medical data from smartphone users through the mobile application of the system, (2) sharing the data gathered with a health expert or government health authorities instantly, if consent is obtained (3) running its own semantic rule-based inference engine to predict a possible Covid-19 case. The inference engine is a Java bot that is responsible for all reasoning processes and consists of: (1) symptom-oriented Covid-19 disease ontology for self-pre-diagnosis—OntCoV19 and (2) semantic Web rules. In this article, the design, development, and usage of the system are discussed and the usage of the major functional services of the system by a system user via its mobile application is discussed. Finally, case test studies for system evaluations and the results and findings obtained are discussed in the article.
1.1 Contributions

In this study, the following contributions have been tried and aimed with the proposed system:

- To the best of our knowledge, this study is one of the first medical expert system attempt to the struggle against Covid-19 in the TRNC.
- In addition, this study is the first attempt to develop a knowledge-based self-pre-diagnosis system for predicting Covid-19 according to the requirements on law of personal data rights in TRNC via its mobile app.
- This study is first attempt that aimed to reveal the usability of such a self-pre-diagnosis system and test it with the demographic and symptomatic data of users in TRNC.
- The proposed system is one of the first attempt which is able to run a set of proper semantic rules on its ontology, to predict a probability score of having Covid-19 for its users via a mobile application, and to deduce a set of precautionary and supportive action suggestions using demographic and symptomatic data.
- The proposed system is one of the first attempts to legally examine how the personal and medical data of smartphone users should be used in national crisis situations such as epidemics.
- The proposed system is one of the first attempts to syndicate various statistical models, ontology knowledge base, and semantic Web rule technologies in the struggle against Covid-19.
- This study attempts to demonstrate the reliability and accuracy of the proposed system by comparing the results of the inference engine module with the results of a collaborated health expert. For now, 169 patient cases were used in experimental studies. However, these verification studies are our initial efforts to test the entire system using real cases’ data in the TRNC in near future.
- This study attempts to provide a system involving both patients and healthcare experts. The system has been designed as a platform where both patients and healthcare experts can easily interact when desired in the struggle against Covid-19.
- This study attempts to provide an alternative solution to prevent delay in the diagnosing of the Covid-19 risk in individuals and tries to make a significant contribution to today’s world by allowing remote monitoring of the disease.

The rest of the paper is organized as follows. Section 2 discusses similar research studies in literature. Section 3 discusses the proposed system architecture and user interactions. Section 4 presents the data types gathered from smartphone users according to the Law of Personal Data Protection in TRNC. Section 5 presents the usage of the mobile application of the proposed system through a case study. Section 6 presents the system ontology which named as “Symptom-oriented Covid-19 ontology for self-pre-diagnosis—OntCoV19.” Section 7 discusses the semantic rules of inference engine of the proposed system. Section 8 discusses the experimental studies performed, the dataset used, and the system verification preferences. Section 9 discusses evaluation results and limitations observed in this research study. Section 10 presents the summary and conclusions.

2 LITERATURE SURVEY

In this section, various scientific research studies based on medical solutions and patient monitoring, various m-Health applications offered in mobile stores, and Web applications, in the Covid-19 pandemic are discussed. The research studies are especially investigated to understand recent developments in the Covid-19 disease management and patient monitoring. The information obtained after the literature study reveals the basic key criteria for calculating the probability of having Covid-19, based on the symptoms and demographic data observed in individuals. With this literature survey effort, it has been experienced that which symptom data need to be collected with the mobile app and which key metrics necessary in calculating the probability value of having Covid-19. Therefore, the research effort is seperated in two primary research areas.

The first research area is about the recent research studies on mobile applications and Web-based systems about Covid-19 management in literature. Various research articles are investigated and some of them are discussed to indicate the required information based on Covid-19 management and patient monitoring. The investigations directed us to find out the information required about the proposed system modules, the system actors, the key metrics, the user interactions, how Covid-19 medical diagnosing is performed for potential cases etc. In other words, the research studies are investigated to gather information about the problem domain and to form the proposed system architecture and user functionalities.

In the Czech Republic, some researchers (Komenda et al., 2020) have developed an interactive Web-based application to provide complex reporting of the Covid-19 epidemic. It was developed as a tool that creates vital reports based only on valid data sources. In the Czech Republic, analytical processing and visualization of real data from standard sectoral processes providing verified information about the Covid-19 epidemic was implemented through data mining. The system is used as an online platform that provides a series of outputs in the form of web-based overview, public tables, graphs and maps of the current spread of Covid-19 in the Czech Republic.
Various researches have undertaken for Covid-19, such as improving the quality and shortening of the antibody test, availability of vaccines, and calculating the probability of reaching Covid-19 immunity. Due to the increased Covid-19 pandemic and the importance of time in various sectors, a Covid-19 “Immune Passport” has been considered by the researchers (Eisenstadt et al., 2020) as a solution to enable individuals to return to their jobs. The researchers consider a prototype mobile application and decentralized server architecture that eases instant verification of test results. Personally, identifiable information is only stored at the user’s choice, and the application lets the end-user selectively to share only the specific test result with no other personal information exposed. For the test certificate holder, issuer (e.g., healthcare staff, pharmacy) and verifier (e.g., employer), it is ‘just another app’ which takes only minutes to use. Many other uses of secure and private certification have also been made possible through the developed mobile phone application and decentralized servers. The infrastructure of the system is thought to be embedded in any other application or Web portal via APIs.

In March–May 2020, some researchers (Currie et al., 2020) evaluated the potential contribution of the CoviddSafe (Prime Minister of Australia, 2020) contact tracing application in Australia to the continuous control of the Covid-19 epidemic. The researchers thought that if a high level of community participation could be achieved, a new Covid-19 wave would be alleviated, along with CoviddSafe administration, testing and social distance, with improved and accelerated case detection. The username, mobile number, postcode and age information is collected from mobile users. The extent of viral testing, community participation in social distancing, and the level of understanding of the CoviddSafe implementation have been studied as factors influencing predicted trends in Australia.

CoronApp (Barbosa et al., 2020) is a mobile application developed by the Colombian government and created to monitor Covid-19 cases in the country. CoronApp was developed for both Android and Apple operating systems. Colombia has a population of 50 million and this application has been downloaded about 5 million times until May 2020. From the citizens’ point of view, the benefit of using the CoronApp is that individuals can make self-diagnosis assessments and access recommendations for the treatment of Covid-19. In addition, the main benefit of the Government is to obtain geolocalized health data created by citizens, and to use it to support real-time decision-making. User registration of the CoronApp is optional for those looking for information only, but it is mandatory for those wanting to do a health self-assessment test. Identity data required from CoronApp users during the registration are the user’s full name, national ID and a mobile number. Users can enter their symptoms, receive preventive advices and share locational information.

Another mobile application is Aarogya Setu (Gupta et al., 2020) was made by India government. It is a mobile application that must be downloaded by government and private sector employees. Using a mobile phone’s Bluetooth and location data, the application allows its users to scan a database of known infection cases to know if they are with a Covid-19 person. The data is then shared with the government. For example, “if you’ve met someone who has had a positive test in the past two weeks, the application will calculate your risk of infection and send out precautionary messages, depending on how close you are and how close you are.” Although username and number are not made public, the application collects the information of a user’s gender, travel history, and whether the user smoke or not. If a person wants to leave from the system, entire information of the person stored is deleted after 30 days.

According to our literature research in the first area, currently developed smartphone and Web systems for Covid-19 tracking, GPS location data obtained from Bluetooth signals of smart devices (e.g., electronic bracelets, phone) as well as the information of individuals’ profile and observed symptoms are collected.

The second area of the literature survey is about the observed symptoms and the severity level in the pre-diagnosing of Covid-19.

A smartphone application (Menni, Valdes, Freidin, et al., 2020) is developed and used that allows individuals to report Covid-19 symptoms. In the study, a total of 2,618,862 participants entered their symptoms on a daily basis to determine most significant factors in Covid-19 risk using the mobile application. For example, whether they have asymptomatic or symptomatic symptoms, whether they have been hospitalized, whether they have pre-existing medical conditions, and whether they have been tested for an active infection of SARS-CoV-2. The researchers analysed these self-reported symptoms data of 2,618,862 people living in the United States and the United Kingdom. Users recorded these data between March 24 and April 21. During this time, 18,401 participants entered the results of the SARS-CoV-2 test into the system. The proportion of participants who stated loss of smell and taste was higher in those with a positive test result (4668 of 7178 individuals; 65.03%) than in those with a negative test result (2436 of 11,223 participants; 21.71%). The researchers analysed all data from UK users to identify the most strongly associated independent symptoms with Covid-19 and adjusted the results for age, sex, and BMI. The results showed that the loss of taste and smell, excessive fatigue, cough, and loss of appetite are the most significant indicators of Covid-19 disease. Using this information, the researchers developed and proposed a statistical model to predict whether a user has the Covid-19. The researchers applied their predictive model to the 805,753 UK and US symptom-reporting participants who had not been tested for Covid-19. According to the formula and results retrieved, 140,312 (116400–164,224) of these 805,753 participants probably had Covid-19. It is put forwarded that the formula is nearly 80% accurate at predicting whether or not a user has Covid-19 based on a positive diagnostic test (Menni, Valdes, Freidin, et al., 2020).

In earlier version of the study (Menni, Valdes, Freydin, et al., 2020) conducted by the same researchers, the loss of smell prevalence was found to be three times higher (59%) among individuals. In the study, the researchers have proposed another form of their statistical model that combines different symptoms with different coefficients to predict individuals having with Covid-19. This model has been applied to more than 400,000 people. Researchers have stated that roughly 13% of those who display symptoms have Covid-19 infection. They also stated that this
rate is slightly higher in women than in men. These studies (Menni, Valdes, Freidin, et al., 2020; Menni, Valdes, Freydin, et al., 2020) provided a functional basis for the development of the proposed system and an enlightenment for the solution.

In (Michelen et al., 2020; Eliezer et al., 2020), the researchers discussed the symptoms observed and its severity in Covid-19 disease. The researchers stated that the cough symptom observed in two-thirds of the cases was unreliable alone as a key diagnostic symptom. Therefore, these researchers mentioned that “anosmia” may be a stronger symptom for the diagnosis of Covid-19 than “fever” data. The researchers defined the strong predictor symptoms at onset of Covid-19 are as “fever,” “cough,” “loss of smell,” and “shortness of breath.” In addition, other symptoms observed in Covid-19 are defined as “dyspnea,” “headache,” “anosmia,” “diarrhoea,” “sore throat,” “fatigue,” “rhinorrhea,” “abdominal pain,” and “loss of appetite” with a specificity of 83% and sensitivity of 55% by the researchers. Lastly, the researchers identified two more symptoms that were not previously reported before; “ocular reaction” and “skin rash”. About the symptom severity observations, they specified that approximately 80% of patients are mild (no symptoms of pneumonia are seen) or asymptomatic, but still contagious. As the research shows (Michelen et al., 2020), if the virus does not cause serious symptoms, individuals in the community will not be able to recognize the virus, nor need preventive measures, or not require medical help. Therefore, the virus can continue to spread rapidly, and more cases may occur over time. However, in order to combat this disease, health experts need to know the total number of Covid-19 cases in community and accurately identify the most significant symptoms that indicate this disease. With this information, it will be easier for health experts to distinguish between mild and moderate patients.

In another study (Christiano, 2020), the researchers investigated the cases with Covid-19 according to the severity categories such as “mild,” “moderate,” and “severe” (as seen in Table 1). The research discusses that which symptoms make a case severe. The researchers discuss significant research studies which reports the vast majority of Covid-19 cases fall into the least severe category; 81% Mild to Moderate, 14% Severe, and 5% Critical. Age appears to be a significant factor in who gets the sickest. Due to elder age, immune system problems and any underlying health conditions, such as diabetes, lung condition, or heart disease, make individuals weaken in catching Covid-19. According to the WHO, the incubation period (1–14 days) is the time between when you are infected with the virus and when symptoms start with (5 days in average). It has been reported that mild cases of coronavirus recover within about 2 weeks. Recovery of severe cases can take 3–6 weeks (Christiano, 2020). The disease can start with a mild case, and then it can become severe. In addition, it has been reported that patients are again infected with Covid-19 after recovery. Therefore, if a person survives a coronavirus infection, it is recommended to take all precautions to prevent re-infection (Christiano, 2020).

Another study (Yu et al., 2020) discusses severity detection of Covid-19 paediatric cases. The study used the data of 105 children aged 1–16 (64 males and 41 females) in Wuhan who admitted to Wuhan Children's Hospital (it is the sole designated hospital in Wuhan for Covid-19 children patients) from Feb. 1 to Mar. 3, 2020. The youngest child was 1 day after birth, and the oldest was 15-year-old. The children infected with Covid-19, of which eight were critically ill. The epidemiological, clinical laboratory, and outcome data were extracted from the medical records of these patients. The researchers developed a supervised decision-tree classifier to analyse these data. They used all clinical measurements as features and defined ‘mild’ and ‘severe’ as labels for them. They used standard F1-score to evaluate the performance of the classifier. Male infection rate (60.95%) is higher than female. In addition, children over 6 years old have the highest infection rate (60.95%). In results, researchers have discovered a clinical pathway that can achieve 100% F1 score. The researchers successfully identified mild and severe paediatric patients over the clinical route due to two properties (namely, Direct Bilirubin (DBIL) and alanine transaminase (ALT)).

### TABLE 1  Comparison of coronavirus symptoms obtained research studies (Christiano, 2020)

| Mild cases of Covid-19 | Moderate cases of Covid-19 | Severe cases of Covid-19 |
|-----------------------|---------------------------|-------------------------|
| Symptoms              |                           |                         |
| Low-grade fever, dry cough, nasal congestion, runny nose, fatigue, sore throat, headache, new loss of taste and smell. | Fever >38.3–38.9, chills with repeated shaking, deep cough, fatigue, body aches, muscle pain, general feeling of being unwell. | Fever >39, deep cough, fatigue, body aches, breathing difficulties, chest discomfort, confusion/ unresponsiveness, bluish face/lips (a sign you are not getting enough oxygen), possible gastrointestinal issues, like diarrhoea or nausea. |
| Prevalence            | 81% of Covid-19 cases.    | 14% of Covid-19 cases.  | 5% of Covid-19 cases.  |
| Incubation Period     | 1–14 days                 | 1–14 days               | 1–14 days               |
| Treatment             | Rest, fluids, over-the-counter pain and fever reducer. | Rest, fluids, over-the-counter pain and fever reducer. | May need hospitalization for IV fluids, oxygen, and help breathing. |
| Recovery              | 2 weeks                   | 2 weeks                 | 3–6 weeks               |
In summary, with the contribution of the relevant research studies discussed above, this study focuses on to develop an inference mechanism for the proposed system by considering the most important medical symptoms observed and prediction models obtained from various clinical studies. The research studies discussed above and their contributions to the proposed system are summarized in Table 2.

3 | SYSTEM ARCHITECTURE AND USER INTERACTIONS

This section clarifies the users and their roles of the self-pre-diagnosis system. System architectural design details and user interactions are demonstrated in Figure 1.

The user data required (Step one in Figure 1) to run the inferencing module is gathered through the mobile application of the system. The application includes a patient examination card screen to collect patient demographic and symptoms data. The data then sent to the system ontology to retrieve the inferencing results (Step two and three in Figure 1) via the semantic rules. As a result of the inference step, by running semantic rules in the system ontology, the probability score of having Covid-19 and a set of precautionary and supportive action suggestions are deduced (Step four).

As shown in Figure 1, the system includes five main modules which are: (1) Admin Module (AdM), (2) Health Expert Module (HeM), (3) Smartphone User Module (SuM), (4) Government Health Authority User Module (GHAM), and (5) Inference Engine Module (IeM). Table 3 shows certain main functionalities of the HeM, SuM, and GHAM. The system gathers several types of data from the patient users, such as: (1) SUs demographic details (e.g., name and surname, age, located city, etc.), (2) health-related measurements (e.g., fever, height, weight, etc.), (3) the presence of pre-existing medical conditions or any hospitalization, (4) travel details in the past 14 days, (5) symptom-based medical data (e.g., recent symptoms observed, breath sound data, images of the ear surface/throat/tonsil, etc.), (6) self-isolation or medical status, and (7) date of latest COVID-19 tests with results, if exists. Some of the data daily reported by users (i.e., breath sound data, images of the ear surface/throat/tonsil, and video records of user's general face views) are not used in the inferencing module of the system. However, the data is kept on the system database and then presented to the responsible health expert to perform a remote offline consultation or get an opinion and evaluation from the expert at any time (Step five in Figure 1).

Health Expert (HE): HEs provide professional support by evaluating the medical data uploaded by their registered patients. HEs are able to view the list of all patients assigned to them with their (1) demographic details, (2) health-related measurements (e.g., fever, height, weight, BMI), (3) the presence of pre-existing medical conditions or any hospitalization, (4) travel details in the past 14 days, (5) recent symptoms observed, and (6) other medical data gathered (e.g., breath sound records, images of throats etc.). A smartphone user opens a new “chatbot” page and can initiate a chat with his/her doctor. After the user enters the above-mentioned data in a blank patient examination card on the same page, he/she reaches the online Covid-19 test result. The patient examination card opened in order to receive support from HEs via the system mobile application is shared with the HEs by the patients at any time to be evaluated by the HEs. A patient examination card opened to receive support and evaluation from HEs is shared with their physicians at any time via the mobile application. HEs are able to view the estimated score calculated (it is the probability of the patient is having Covid-19 based on a prediction model that will be detailed in further sections) for each created patient examination card. HEs are also able to display recent statistical data on the number of Covid-19 cases seen in the country, in any selected country, in the world. The users are also able to display latest statistical data of current number of Covid-19 patients seen in the country, in any selected country, in the world.

Smartphone Users (SUs): SUs can be all suspected Covid-19 patients, anyone who wants to predict Covid-19 disease with an online test and needs a remote physician support. SUs is responsible for providing the instant medical data needed by the system. In the SuM, for continuity, the medical data of the SUs are expected to be entered at shorter intervals (e.g., daily, weekly or twice a month). The user is able to choose their region and display all the hospitals in their region. The users are able to initiate a new patient examination card to run an online Covid-19 diagnostic test. To do this, the user needs to fill the blank “patient examination card” on the same screen. The user is able to ask for a consultation to a HE in their regions’ hospital. To do this, the user is able to initiate chat with his/her doctor on the same chatbot screen, if doctor answer to the chat request. Each chatbot screen automatically runs the online Covid-19 test and presents some precautionary and supportive action suggestions for the patient user through IeM. The users are able to view all their examination cards they have created so far, with Covid-19 estimated result values. Various precautionary and supportive action suggestions are deduced to the SUs by the IeM using the demographic details, some of the health-related measurements, and observed symptoms data, such as:

- SG_01: Continue online tests.
- SG_02: You need to wash hands with soap and water often and for at least 20 sec.
- SG_03: You need to practice good respiratory hygiene and avoid touching the eyes, nose or mouth with unwashed hands.
- SG_04: You need rest.
- SG_05: Stay home and self-isolate even with minor symptoms such as cough, headache, mild fever, etc. until you recover.
- SG_06: You need to drink plenty of fluids.
| Reference               | Research topic                                                                 | Technology/ methodology                      | Product / result                                                                 | Contributions to our study                                                                 |
|------------------------|---------------------------------------------------------------------------------|-----------------------------------------------|----------------------------------------------------------------------------------|------------------------------------------------------------------------------------------|
| Komenda et al. (2020)  | Developed a Web-based tool that creates complex reports based on real valid data of the Covid-19 epidemic in the Czech Republic. | Data mining techniques are used to analytical processing and visualization based on the data. | A Web-based tool is developed.                                                    | Necessary functional requirements about the problem area has been collected.             |
| Eisenstadt et al. (2020)| Covid-19 “Immune Passport” application has been developed to enable individuals to return to their jobs. | Web 2.0 and mobile technologies.              | A prototype mobile app and decentralized server architecture that facilitates instant verification of test results is introduced. | The information has been collected about necessary system modules, system actors, medical metrics used in the Covid-19 pre-diagnosis, user interactions, identification of potential cases, proper advices for the treatment of Covid-19. |
| Currie et al. (2020)   | Assessed the potential contribution of the Australian Government’s mobile smartphone tracing app (COVIDSafe) to the continuous control of the Covid-19 epidemic. | A system dynamics model of Covid-19 on a modified susceptible, exposed, infected, recovered (SEIR) compartmental model structure is developed and simulated. | A system dynamics model is developed.                                               | The information has been collected about necessary system modules, system actors, user interactions, social distancing details, the states of the Covid-19 disease. |
| Barbosa et al. (2020)  | Monitoring Covid-19 cases in Colombia via CoronApp.                             | Web 2.0 and mobile technologies.              | Android and Apple mobile applications are developed. Its users can add share locational information, make self-diagnosis assessments, and access advices for the treatment of Covid-19 after entering their symptoms. | The information has been collected about necessary system modules, system actors, medical metrics used in the Covid-19 pre-diagnosis, user interactions, identification of potential cases, proper advices for the treatment of Covid-19. |
| Gupta et al. (2020)    | Assessed the contact tracing mobile application “Aarogya Setu” of India to the continuous control of the disease. The “Self-Assessment Test” on Aarogya Setu provides some questions related to the health and symptoms observed to its users, and based on their answers, the application shows a risk level for the users in different colour codes. | Web 2.0, mobile, Bluetooth, GPS technologies. For data analysis, the application uses classification, association rule mining, and clustering are used. | Aarogya Setu app is available for both iOS and Android users. It uses Bluetooth and location data of users that has been designed in such a way that it informs the users through notification if they cross paths with a Covid-19 positive person. | It calculates a risk level of infection based on symptoms and produces precautionary messages, depending on proximity distance. The idea is also considered in the proposed application. The information on the symptoms considered in the self-assessment test and the test’s questions has been collected. |
| Menni, Valdes, Freidin, et al. (2020) | A mobile application is developed and evaluated that allows individuals to report Covid-19 symptoms and predict whether they are Covid-19. A total of 2,618,862 participants entered potential Covid-19 symptoms using the mobile application. | Web 2.0 and mobile technologies.              | This study analysed the self-reported symptoms data of the 2,618,862 people statistically and put forwarded a mathematical model to predict Covid-19 disease. The researcher used the loss of taste and smell, excessive fatigue, cough, and loss of appetite symptoms in the model which are the most | This model has been applied to symptom data of smartphone users (805,753) to predict possible infection. It predicted that 140,312 (17.42%) participants were highly likely to have Covid-19. The model is integrated into the inference engine of the proposed system. |

(Continues)
| Reference | Research topic | Technology/ methodology | Product / result | Contributions to our study |
|-----------|----------------|-------------------------|------------------|----------------------------|
| Menni, Valdes, Freydin, et al. (2020) | Another study is conducted by the same researchers (Menni, Valdes, Freydin, et al., 2020). | Web 2.0 and mobile technologies. | Researchers have developed another mathematical model that combines symptoms to predict individuals with Covid-19. | This model has been applied to more than 400,000 people. It predicted that roughly 13% participants were highly likely to have Covid-19. The model is integrated into the inference engine of the proposed system. |
| Michelen et al. (2020) | The researchers assessed the most observed symptoms and its severity in Covid-19. | Symptoms observed in Covid-19 have been assessed in the systematic review studies and the largest cohort studies. | The researchers stated that “anosmia” is a stronger sign of a Covid-19 diagnosis than “fever” data reported by people in the community. He also stated that the strong determinant symptoms at the onset of Covid-19 were “fever”, “cough”, “loss of smell” and “shortness of breath”. Researchers reported other symptoms with a certain specificity, such as “shortness of breath”, “headache”, “anosmia”, “diarrhoea”, “sore throat”, “fatigue”, “rhinorrhea”, “abdominal pain” and “anorexia”. | The symptoms with severity levels are considered in the inference engine's rule base of the proposed system. |
| Christiano (2020) | The researchers compared the coronavirus symptoms in terms of prevalence, incubation time, treatment, and recovery by dividing Covid-19 cases into severity categories such as “mild”, “moderate” and “severe”. | Literature survey is done. | The researchers discussed various significant research studies which reports the vast majority of Covid-19 cases fall into the least severe category; 81% Mild to Moderate, 14% Severe, and 5% Critical. | The coronavirus symptoms according to severity categories such as “mild”, “moderate” and “severe” are considered in the rule set of the inference engine of the proposed system. Also, treatments and recovery details are used as user advices in the proposed system. |
| Yu et al. (2020) | 105 infected children admitted to Wuhan Children's Hospital were considered for this retrospective study. The epidemiological, clinical laboratory, and outcome data were extracted from the medical records of these patients. | A supervised decision-tree classifier is developed to data analysis. They have used standard F1-score [4] to evaluate the performance of the classifier. | The researchers successfully identified mild and severe paediatric patients over the clinical route due to two properties (namely, Direct Bilirubin (DBIL) and alanine transaminase (ALT)). | Since paediatric patients are beyond our scope, we could not use the DBIL and ALT parameters, which were suggested as determinants in Covid-19 in the proposed system. |
| Zens et al. (2020) | The researchers identified possible unreported symptoms as well as the distribution pattern of Covid-19 symptoms. | A tool is developed by DESIGN-IT GmbH as Apple iOS and Google Android mobile applications, in collaboration with the University Medical Center | An app-based daily self-reporting tool is developed (Covid-19 Symptom Tracker), and the data gathered from its users is assessed. The tool helps to identify novel symptoms of | Researchers found the loss of smell and taste as the cardinal symptom; also, diabetes is a very important risk factor for the symptomatic case of Covid-19. Therefore, ||
Consult your doctor to get an over-the-counter pain reliever that’s best for you.

Try to keep your temperature at normal levels such as by applying a cold towel, taking a warm shower, and so forth.

**Government Health Authority Module (GHAM):** The module users are able to list all hospitals by cities in the country. The users are able to view the number and list of Covid-19 patients of a hospital after choosing the hospital in the list. The users are able to display latest statistical data of current number of Covid-19 patients seen in the country, in any selected country, in the world. The users can also view statistical data on total infected patients, recovered patients, patients who have died, patients currently infected (test positive) and currently potential patients (based on unverified online test estimation results on the entire created examination cards by community) on the map. If a SU allows other users to access her/his patient examination cards and other data on the system, other users can then reach it. No data inferencing has been applied to this user account.

**Admin:** Admin is responsible for the authorization of users, management of the system users' screens, and other configuration settings. No data inferencing has been applied to this user account.

Since page limitations are required in publications and the HeM and SuM modules are the main focus of the proposed system, only these modules are discussed in the rest of the article.

### Table 2 (Continued)

| Reference          | Research topic                                                                 | Technology/methodology                                                                 | Product/result                                                                 | Contributions to our study                                                                 |
|--------------------|--------------------------------------------------------------------------------|----------------------------------------------------------------------------------------|--------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------|
| Freiburg and Kliniken Ostallgaeu-Kaufbeuren, Fuessen Hospital. | Covid-19 and to estimate the predictive value of certain symptoms. | similar data considered in this tool (e.g. age, gender, postal code, past medical history, daily symptoms, SARS-CoV-2 test results) were also taken into account in the proposed system. |
| Sudre et al. (2020) | As a single symptom in Covid-19 cannot predict the severity of the disease or the need for special medical support, the researchers asked if documenting symptom time series over the first few days informs outcome. | Unsupervised time series clustering over symptom presentation was performed on data collected from a training dataset of completed cases enlisted early from the Covid Symptom Study Smartphone application, yielding six distinct symptom presentations. | The longitudinal clustering of symptoms can predict the need for respiratory support in severe Covid-19 in advance. | The 14 reported symptoms with severity levels are used in the inference engine's rule base of the proposed system. |

- SG_07: Consult your doctor to get an over-the-counter pain reliever that’s best for you.
- SG_08: Try to keep your temperature at normal levels such as by applying a cold towel, taking a warm shower, and so forth.

**Government Health Authority Module (GHAM):** The module users are able to list all hospitals by cities in the country. The users are able to view the number and list of Covid-19 patients of a hospital after choosing the hospital in the list. The users are able to display latest statistical data of current number of Covid-19 patients seen in the country, in any selected country, in the world. The users can also view statistical data on total infected patients, recovered patients, patients who have died, patients currently infected (test positive) and currently potential patients (based on unverified online test estimation results on the entire created examination cards by community) on the map. If a SU allows other users to access her/his patient examination cards and other data on the system, other users can then reach it. No data inferencing has been applied to this user account.

**Admin:** Admin is responsible for the authorization of users, management of the system users’ screens, and other configuration settings. No data inferencing has been applied to this user account.

Since page limitations are required in publications and the HeM and SuM modules are the main focus of the proposed system, only these modules are discussed in the rest of the article.

### 4 PERSONAL DATA GATHERED FROM SMARTPHONE USERS ACCORDING TO TRNC LAW

As European Union (EU) stated that data and digital technologies have an important role to play in the struggle against Covid-19 pandemic. According to Guidance pronounced that mobile applications can support public health authorities in monitoring, containing and lifting containment measures of the COVID-19 (Guidance on Apps supporting the fight against COVID 19 pandemic in relation to data protection with the Number 2020/C 124 I/01, Official Journal of the European Union, 2020; European Commission, C2296 Final, 2020). In the application prepared here, volunteerism is essential, using the application is not mandatory for individuals. In fact, volunteering is emphasized in particular all directives which are related with mobile applications, privacy and personal data (2020/C 124 I/0; Official Journal of the European Union, 2020; Directive 2002/58/EC of the European Parliament and of the Council, 2002, Directive on privacy and electronic communications, 2002). The meaning of volunteering is “apps downloaded, installed and used on a voluntary basis by individuals.” The guidance which promulgated by EU addresses only voluntary apps supporting the fight against COVID 19 pandemic (2020/C 124 I/0 Guidance, Guidance on Apps supporting the fight against COVID 19 pandemic in relation to data protection). The applications must furnish with “accurate information to individuals about pandemic, questionnaires for self-assessment and for guidance to individuals (symptom checker functionality), alert persons who have been in proximity for a certain duration to an infected person, in order to provide information such as whether to self-quarantine and where to get tested (contact tracing and warning functionality) and communication forum between patients and doctors in situation of self-isolation or where further diagnosis and treatment advice is provided (increased use of telemedicine) according to the EU Directive” (2020/C 124 I/0 Guidance, Guidance on Apps...
supporting the fight against COVID 19 pandemic in relation to data protection; Directive 2002/58/EC of the European Parliament and of the Council, 2002).

Since the initial real-life usage studies of the system are planned to be carried out in the TRNC, all the data gathered from the participants and suggestions generated by the iE are organized according to the numbered with 89/2007 the Law of Personal Data Protection in TRNC (89/2007 Law of Personal Data Protection, TRNC, 2007). The law is promulgated in 2007 and it is broadly similar with the Directive 95/46/EC of the European Parliament which is repealed with General Data Protection Regulation (GDPR). GDPR is the main regulation of the European countries for personal data protection (GDPR, Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016). Since the law used in the proposed system this was prepared within the framework of the repealed Directive 95/46/EC of the European Parliament (Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995, 1995, on the protection of individuals with regard to the processing of personal data and on the free movement of such data, Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995).
TABLE 3 Some major functional requirements of the HeM, SuM, and GHAM are given below

| FUNCTIONS | HeM | SuM | GHAM |
|-----------|-----|-----|------|
| Sign In/Log In. | ✓ | ✓ | ✓ |
| Displays confirmation and information screens and obtains consent from patient users. | ✓ | | |
| View the list of all patients of a doctor. | ✓ | | |
| View a particular patient profile. | ✓ | ✓ | ✓ |
| Open a new “chat bot” screen to initiate chat a doctor and fill a blank “patient examination card”. | ✓ | | |
| Displays a list of all previously filled patient examination cards, with probability result value of having Covid-19 and suggested supportive treatment steps for the patient. | ✓ | ✓ | ✓ |
| Shares a filled patient examination card to a HE or GHAM users. | ✓ | | |
| Displays all news and performs a search on the news. | ✓ | ✓ | ✓ |
| Displays all the hospitals in the country. | ✓ | | |
| Displays recent number of and list of Covid-19 cases in a particular hospital. | ✓ | | |
| Displays the recent statistical data on the number of Covid-19 cases in the country. | ✓ | ✓ | ✓ |
| Displays the recent statistical data on the number of Covid-19 cases in a selected country. | ✓ | ✓ | ✓ |
| Displays the recent statistical data on the number of Covid-19 cases in all countries. | ✓ | ✓ | ✓ |
| Display statistical data on a map of the total infected patients in the country. | ✓ | | |
| Display statistical data on a map of the recovered patients in the country. | ✓ | | |
| Display statistical data on a map of the death due to Covid-19 in the country. | ✓ | | |
| Display statistical data on a map of the currently infected patients in the country. | ✓ | | |
| Displays estimated data belong to potential patients on a map (the data based on unverified online test prediction results that comes from all patient examination cards created by users). | ✓ | | |

*If the patient gives consent to other type users to access his/her information in the system.

In this study, all data gathered from system users are examined in two different dimensions: temporally and theoretically. All these data and classifications are presented in a table in Appendix A.

According to the current law, the theoretical-based data are divided into two categories. These categories are: (a) personal data and (b) sensitive or special data. The law must have the following principles: “(i) prescribe in detail the processing of specific health data and clearly specify the purposes for the processing; (ii) spell out clearly who is the controller, that is the entity processing the data, and who, beside the controller, can have access to such data; (iii) exclude the possibility to process such data for different purposes than those listed in the legislation and (iv) provide for specific safeguards” (2020/C 124 I/0 Guidance, Guidance on Apps supporting the fight against COVID 19 pandemic in relation to data protection, 2020).

In addition, the time-based data gathered from system users are separated as: (a) the data gathered at first use and (b) the data gathered in continuous use. At first use, the mobile application requires: (i) user demographic details, (ii) health-related measurements (e.g., fever, height, weight, BMI), (iii) the presence of pre-existing medical conditions, (iv) date of recent travel details, if exists, (v) date for self-quarantine or health care taken, if exists, and (vi) date of COVID-19 tests with results, if exists. In continuous use, the users report daily information on the observed symptoms and other medical data (e.g., breath sound data, images of the ear surface/throat/tongue, and video records of user’s general face views) through the patient examination card on the chat bot screen. After inferencing, the users are able to reach the prediction value on the probability of having Covid-19 and some precautionary and personalized supportive action suggestions. Individuals with asymptomatic are also encouraged to use the application. After generating the prediction value and the suggestions by leM, the authentication service of the system shows the results only to the users who are assigned consent by a patient user.

The conditions of processing personal (Art. 5 and 6) or sensitive data (Art. 5 and 7) are different according to the law. However, in both cases the notification obligation (Art. 13) must be fulfilled before the data is received for the data process by the data collectors. The EU requests precaution to be taken to collect information, especially since sensitive data is collected in applications on this issue. For this reason, the users are presented with an information text on the first use in the system. Apart from this, since the system has medical data, it is necessary to obtain explicit consent for each data type by observing the processing conditions of sensitive data.

First of all, a privacy policy has been established within the framework of data protection principles. The issue of privacy is especially the basis of personal data law. International legal rules have been followed in this regard (EU GDPR, 2018). To protect right to privacy, key rules of domestic law (TRNC Constitution, Law of the Personal Data Protection, Regulations, 2007) must be upheld. As much as possible this application complies with the principles of EU law such GDPR and the e-Privacy Directive. The reason EU law is being observed in the study is because it sets
standards in all over of Europe. In terms of international law, the application contains: the proportionality of the measure in terms of duration and scope, purpose limitation, restricted data retaining, minimization, removal, anonymization of data, and application use being voluntary and centered on persons choosing in.

The power given to the state to trace people should be removed once the crisis is over, and an independent authority should be established to ensure the rules are implemented (and to act if they are not). In addition, what is proposed in the “Common EU Toolbox for Member States” published on 15 April by the eHealth Network (2020), and in the letter from the European Data Protection Board (EDPB) to the Directorate General for Justice and Consumers (EDPB, 2020) should be taken into consideration by Member States.

The application, which we have developed within the framework of national law, has been made in accordance with personal data law, both international law is required. In the application a determining factor for persons to confidence the application is representing that they remain in control of their personal data. In order to achieve this the following have been noted in the design of the application: the installation and using of the application is voluntary, different application functionalities should not be bundled so that the individual can provide his/her consent specifically for each functionality, proximity data information should be stored on the individual's device. If those are to be shared with health authorities, they should be shared only after confirmation that the person concerned is infected with the COVID-19 and on the condition that he/she chooses to do so, health authorities should provide the individuals with all necessary information related to the processing of his or her personal data (in line with Articles 12 and 13 of the GDPR and Article 5 of the privacy directive), the individual should be able to exercise his/her rights in particular, access, rectification; deletion under the GDPR.

In particular, responsibility of information, the rights of persons and the issue of consent are considered separately. Registration is not complete to the program until the informational text is read. In addition, the individuals' transparent, detailed, and clear consent is taken earlier for all varieties of data processing activities.

In the text of consent, it is stated that the control over the data is always on the individuals that they have the right to remove, change or alteration them and the conditions of whether the data is to be shared with those responsible. (See 89/2007 Law Art. 7/2, 3) As seen in informational text the information provided simply reachable and provided in clear language. It is also vital to adopt sufficient security measures and privacy policies ensuring that personal data are not revealed to unauthorized parties (Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995, 1995).

As required by the law, an informative message is displayed for all data collected from users that it will be shared with medical professionals and the government officials when if it is necessary. The gathered data will not only be retrieved, but also processed and some newly inferred data will be generated. Data owners have the right to request their data to be permanently deleted, corrected and updated. If the predicted value for the possibility of having Covid-19 is greater than 85%, it is stated that this value will be shared with GHAM in the information and permission texts shown at the first use of the system. Since the presence of people with high Covid-19 risk in the public sphere may pose a risk to the public interest. TRNC 89/2007 law of the personal data protection (Art. 8) and GDPR (Art. 9 / g) state that personal data and sensitive data can be processed when there is public interest. All requirements are detailed in the information and consent forms. During the processing of patient data gathered by the HEs, legal obligations will be taken into consideration. According to the Law, “The Council of Ministers upon the recommendation of the Council in accordance with a decision taken unanimously may take decisions including regulations on the processing of sensitive data in matters of public interest” (Art. 7/4). Information and consent texts are given as Supplementary material.

## 5 | MOBILE APPLICATION OF THE PROPOSED SYSTEM

Before discussing the technical contributions, the user interactions and data gathered from SUs using the mobile application of the system is discussed through a scenario in this section. The scenario presents the generation of a predicted score of the probability of having Covid-19 for a potential case. In addition, it also presents a set of precautionary and personalized supportive action suggestions deduced by LeM according to the estimation result of the examination card of the user. If the user takes these suggestions into consideration and applies them in a short time, it may provide positive effects for that user in the further days. The ontology, LeM and the semantic rules are detailed in the next sections.

**Scenario:** A patient user, "Ahmet Deniz" (id: 456789, age:45), clicks the “Sign In” button (depicted in Figure 2(a)) to register to the system and generates a new account for himself. Moreover, the user can see a list of the doctors in the hospitals closest to the area they reside in next. The user is also able to choose a doctor to consult the doctor at this point.

As seen in Figure 2(b), the patient first selects “Prof. Dr. Alex Jones” in the doctor's list to get an online consultation and fill out a new examination card and starts a chat. If doctor is suitable for the request, they can response the consultation using the chat bot screen (as shown in Figure 2(c)). Alternatively, if the chat cannot be initiated for any reason (e.g., if it is not the appropriate time, the doctor may not be available), the user can fill in the blank horizontal patient examination card that appears at the top of this chat screen for the doctor to review it later. The card is shown in a dashed red line in Figure 2(c). Various medical data icons are seen on this card such as: observed symptoms icon, fever icon, ear region surface icon, throat/tonsil icon, general face view icon, respiratory icon, and so on. By selecting these icons, various medical data from the user are collected through the user's phone camera and microphone. The first icon asks for the medical symptoms that user has recently observed.
in himself. The user clicks this icon and comes up with a list of symptoms. Each symptom indicates three severity options such as mild, moderate, severe. As shown in Figure 2(d), the user selects the symptoms that he/she observed instantly according to his/her severity. In our example, the user enters the symptoms as “Loss of Smell and Taste,” “Persistent Cough,” “Fatigue,” “Abdominal Pain,” “Inappetence,” and “Diarrhoea” and loads the information to the card (as shown in Figure 3(a)). In addition, entire information entered by the user on the chat bot screen is appeared for the doctor’s review later. In the further steps, the system guides the user to measure the fever, to take images of the throat/tonsil/ear, to take video of the general view of the user’s face, and to record the respiratory sound, and so on.

FIGURE 2 The patient “Ahmet Deniz” (id: 456789) opened a “New Examination Card” that assists in gathering data. (a) The user clicks on “Sign In” to register to create a new account. (b) A list of the closest hospitals’ doctors is seen to select a doctor and to ask a consultation, by the user. (c) After choosing a doctor, an empty “chat board” screen with a horizontal “new examination card” appears. (d) Recent symptoms observed are asked and selected by the user.

FIGURE 3 Loading the symptoms selected, the data collection of recent fever and images of ear of the user. (a) The symptoms are selected and loaded by the user to the chat screen. (b) User enters his fever value as 37.8°C. (c) The fever value of the user loaded to the chat screen. (d) Ear region surface pictures for left and right ears are asked.
As seen in Figure 3, the data gathered about symptoms and all other data to be asked in the following stages are sent to the system database via Wi-Fi. The data are collected as follows:

1. **Measuring the fever**: In this step, the user can use any verified thermometer device for home use and enter the fever medical data manually to the system. In our example, the user selects his/her fever value as 37.8°C from the dropdown fever menu as shown in Figure 3(b) and loads this information to the card (depicted in Figure 3(c)).

2. **Image of ear surface**: The surface images of right and left ears are asked and taken through the camera of smartphone by the user (depicted in Figure 3(d)). The flash of the phone is used. Here, taking the surface images of right and left ears on a daily basis will help to define or measure the presence, amount, viscosity, colour, and similar features of any ear infection. It was thought that it will be useful for HEs to see ear surface images in order to prevent ear infection from being confused with Covid-19 disease. In our example, the user uploads his left/right ear surface images which are taken through the camera of smartphone by the user from the menu as shown in Figure 3(d) and loads the images to the examination card (Figure 4(a)).

3. **Image of the throat/tonsil**: The system asks to the user to upload his throat/tonsil images as seen in Figure 4(b). The user takes an image of the throat/tonsil and uploads it to the card (depicted in Figure 4(c)). The flash of the phone is used. It is observed that high-quality images (≥5 MP) can be obtained.

4. **Taking video of the general view of the user’s face**: After this step, system asks taking and uploading video of the face view of the user via mobile camera in a well-lit home environment (depicted in Figure 4(d)). Thus, the doctor can examine the image of the user’s face for signs of faintness, cyanosis, reddening, red eye inflammation, and so on. Especially, it allows to recognize red eye and a blue appearance around the face and lips, which are two most important Covid-19 symptoms. In our example, the user uploads a general view video of his face which is taken through the camera of smartphone by the user from the menu as shown in Figure 4(d) and loads the face video to the card through the chat bot screen (depicted in Figure 5(a)). As a note, in the process of collecting any image and video data from users, household members can assist the users during the processes through the smartphone camera.

5. **Lung respiratory sound record**: Breathing voice data (is needed to understand the capacity and current conditions of the lungs) is recorded by the user himself via the microphone of the smartphone using a sound recording application. In our example, the user collects breath sound records which are taken through the smartphone’s microphone in a quiet area by the user from the menu as shown in Figure 5(b) and loads it to the card. The chat bot screen runs the online Covid-19 test automatically, then presents the unverified online test prediction result and some precautionary and supportive action suggestions, for the user through IeM (seen in Figure 5(c)). Finally, the system provides a list of all past patient examination cards as shown in Figure 5(d).

Up to this point, it has been explained what types of data need to be collected and how a user needs to fill out a new patient examination card to run an online Covid-19 test. The next sections discuss the symptom-oriented Covid-19 disease ontology, how the system generates precautionary and personalized supportive action suggestions to guide a potential Covid-19 patient at home.

**Figure 4**  Loading the medical data of ear/throat/tonsil images and asking the face view video of the user. (a) Ear region surface photos are loaded to the chat bot screen by the user. (b) Throat picture is asked. (c) Throat picture is taken and loaded to the chat bot screen by the user. (d) The video of the user’s face view is asked.
6 | SYMPTOM-ORIENTED COVID-19 DISEASE ONTOLOGY FOR SELF-PRE-DIAGNOSIS—ONTCOV19

Protégé editor (Knublauch et al., 2004; Noy et al., 2000) is used to develop the Symptom-Oriented Covid-19 Disease Ontology (OntCoV19), which is an open source software for building ontology. OntCoV19 involves numerous concepts, properties and individuals based on the profile data of users, daily medical conceptual data (e.g., symptoms), supportive and preventive treatment suggestions, and the key metrics and factors effected on statistical prediction of Covid-19, and so on.

The ontology starts with “Thing” class that is divided into various sub-concepts such as “Case,” “Cov_2_Test_Result,” “Gender,” “Medical_Condition,” “Patient,” “Personal_Protective_Equipment,” “Sense,” “Suggestion,” “Symptom,” “Symptom_Severity,” “Treatment,” and so on. The class hierarchy is shown in Figure 6 in the leftmost rectangle. In addition, among the concepts, many Object Type Properties (OTP) are generated via owl:topObjectProperty such as “has_Symptom,” “hasGender,” “has_Symptom_Severity,” “hasSuggestion,” and so on. These OTPs are shown in Appendix B and also seen on the second rectangle area of Figure 6. The properties are created in order to establish interclass relations and to link OWL individuals.

Moreover, based on the concepts, many Data Type Properties (DTP) are generated via owl:topDataProperty such as “has Age,” “has Fever,” “has Weight,” “has Height,” “has BMI,” “linked Suggestion,” “has Ocular Reaction Severity,” “is Health Personnel,” “Traveling in the Past 14 Days,” “require Hospital Admission,” “is Asymptomatic,” “has Radiological Evidence,” “can NOT Stay In Community,” “number Of Symptoms Answered,” “contact With Someone,” and so on. These DTPs are listed in Appendix B and also seen on the third rectangle area of Figure 6. The values of DTPs ending in the terms “Coefficient_m1” and “Coefficient_m2” are calculated using the two equations described in the next section.

Another semantically structured relation belong to the classes is “OWL individual” declarations. The “Suggestions” concept on the OntCoV19 indicates various suggestions which are created as OWL individuals such as; SG_01, SG_02, SG_03, SG_04, SG_05, and so on. Suggestions deduced by the system are linked to different patient case categories. The “Case” concept on the OntCoV19 indicates various case categories (eight in total) which are created as OWL individuals such as; CASE_01, CASE_02, CASE_03, CASE_04, and so on. The suggestions deduced by the system are sent directly to the HEs and the SUs.

In total, OntCoV19 involves 13 OWL classes, 50 OTPs, 42 Sub-OTPs, 80 DTPs, 55 Sub-DTPs, 117 OWL individuals, 68 SWRL rules. DTP and OTP associations on OntCoV19 are summarized in Appendix B. In the next section, technical details about the semantic rules and inferencing mechanism running on those rules to generate a prediction result on the probability of having Covid-19 and personalized supportive action suggestions for a particular SU are presented as follows.
Semantic Web Rule Language (SWRL) (Horrocks et al., 2004) is used to write semantic rules and offers stronger deductive reasoning capabilities. SWRL combines OWL DL and OWL Lite sublanguages with sublanguages of the Rule Markup Language (RML) (Jacob, 2003). The rules are used to check specified conditions and if they hold some actions are taken. SQWRL (Semantic Query-Enhanced Web Rule Language) (O’Connor and Das, 2009) is an expressive query language, which is based on SWRL and used to query OWL ontologies. Main difference between SWRL and SQWRL is that the SWRL is an OWL rule language and SQWRL is an OWL query language. However, both of them have an antecedent part, which is known as the body, and a consequent part, which is known as the head. SWRL semantics is used for the left-hand side (inside the body) of SQWRL queries but running an SQWRL query does not alter the ontology in any way.

Protégé editor, SWRL, and SQWRL are used to define and verify the semantic rules, which provide user the combination of the problem definition facts and inference of the knowledge base. We have also used the ROWL plugin (Sarker et al., 2018) to develop our SWRL rules and OWL axioms which is implemented on top of Protege’s SWRL Tab plugin. It provides a user interface for entering semantic rules as input. A developer can enter a rule in ROWL tab using the standard SWRL syntax. When the developer enters a rule through this ROWL plugin and transforms it into OWL axioms, then the ROWL either increments the generated OWL axioms or SWRL rules on the ontology.

As it is mentioned before, the system gathers several types of data from the patient users, such as: (1) SUs demographic details (e.g., name and surname, age, located city, etc.), (2) health-related measurements (e.g., fever, height, weight, etc.), (3) the presence of pre-existing medical conditions or any hospitalization, (4) travel details in the past 14 days, (5) symptom-based medical data (e.g., recent symptoms observed, breath sound data, images of the ear surface/throat/tonsil, etc.), (6) date for self-quarantine or health care taken, if exists, and (7) date of latest COVID-19 tests with results, if exists.

However, IeM uses some of these data to run its own SWRL rules (i.e. demographic details, health-related measurements, and recent symptoms observed) to deduce a result value of the probability of having Covid-19 and the personalized supportive action suggestions for a particular SU.

Totally, 68 different SWRL rules are created on OntCoV19 (depicted in Appendix C). The 68 SWRL rules are based on three different themes: (1) Rules on mathematical expressions, (2) Rules on calculation of estimating the probability of having Covid-19, and (3) Rules on precautionary and personalized supportive action suggestions. A portion of entire SWRL rules on Protégé are displayed in Figure 7.
by IeM to deduce the suitable case id and its linked suggestions through the SWRL rules. Later, the system presents the proper suggestions deduced to the SU and the HE as precautionary and treatment supportive activities.

The 68 rules are divided into three different categories in the study and are discussed below.

1) Rules on Mathematical Expressions: With the development of SWRL mathematical expressions resident library, it is possible to implement complex mathematical expressions using SWRL rules as seen follows. An instance rule showing how the SWRL’s built-in function can be used for calculating the BMI is discussed below.

**Rule of BMI:** Patient(?p) ^ has_Height(?p,? h) ^ has_Weight(?p,? w) ^ swrlm:eval(?BMI, "10000*w/(pow(h,2))",? w,? h) -> has_BMI(?p,? BMI).

The "?p" variable indicates to a SU instance with patient account, while the "?h", "?w", and "?BMI" variables indicate to the height, weight, and the BMI values of the patient, respectively. The value of "?BMI" variable is the result of the swrlm:eval function. The "^" symbol is used as logical "AND" operator. The formula of BMI = kg/m² is used to implement BMI, which takes the "?h" and "?w" variables as input. The "?BMI" value obtained after the rule is fired is assigned to patient "?p" through the has_BMI (?p,? BMI) property. For the case of “Ahmet Deniz” (id: 456789, age: 45) discussed above, the data assigned for the patient are shown at the bottom right of Figure 8 before inferencing.

After inferencing, the “has BMI” is calculated as 28.08 as inferred result which seen on the bottom right of Figure 9. In other words, “?p” has “has_BMI” property association that is associated to the result of that rule’s execution.

2) Rules on Calculation of Estimating the Probability of Having Covid-19: As mentioned earlier, in the study (Menni, Valdes, Freidin, et al., 2020), the researchers collected potential Covid-19 symptoms from a total of 2,618,862 participants via a mobile app. The application is called as symptom tracker is a free smartphone application that was launched in the United Kingdom on 24 March 2020, and in the United States on 29 March 2020. The researchers associated only loss of smell and taste, cough, fatigue, and skipped meals with a positive test result in the US cohort. They performed stepwise logistic regression in the UK cohort, by arbitrarily dividing it into training and test sets (with the ratio- 80:20) to classify independent symptoms most strongly correlated with Covid-19, adjusting for age, sex and BMI. The researchers ultimately suggested that the combinations of symptoms such as loss of smell and taste, fatigue, persistent cough, and loss of appetite are important to achieve better results in this symptom-oriented prediction model.

The model is given below (Menni, Valdes, Freidin, et al., 2020):

\[
\text{Prediction Model 1 } \text{in (Menni et al., 2020a)} = -1.32 - (0.01 \times \text{age}) + (0.44 \times \text{sex}) + (1.75 \times \text{loss of smell & taste}) + (0.31 \times \text{severe persistent cough}) + (0.49 \times \text{severe fatigue}) + (0.39 \times \text{skipped meals})
\]

where, if an individual report that he has observed all these symptoms in himself, all symptoms are taken as 1, if not 0. In the gender parameter, “1” represents male participants and “0” represents women. The result value is then converted to the predicted probability through the formula \(\exp(x)/(1 + \exp(x))\), then the model is assigning estimated Covid-19 cases for probabilities >0.5 and controls for probabilities <0.5 (Menni, Valdes, Freidin, et al., 2020).

The same researchers proposed another linear model based on another combination of loss of smell and taste, fever, fatigue, persistent cough, diarrhea, abdominal pain and loss of appetite resulted (Menni, Valdes, Freydin, et al., 2020). The researchers applied logistic regression...
adjusted for age, gender, and BMI to identify other symptoms (as well as anosmia) that may be associated with Covid-19 disease. For this reason, the researchers questioned 10 more types of symptoms (fever, persistent cough, fatigue, shortness of breath, diarrhoea, delirium, skipped meals, abdominal pain, chest pain, and hoarse voice), and then associated with testing positive for Covid-19 in the UK cohort.

Prediction Model 2 in (Menni et al., 2020b) = \(-2.30 + (0.01 \times \text{age}) - (0.24 \times \text{sex}) + (1.6 \times \text{loss of smell & taste}) + (0.76 \times \text{fever}) + (0.33 \times \text{persistent cough}) + (0.25 \times \text{fatigue}) + (0.31 \times \text{diarrhea}) + (0.46 \times \text{skipped meals}) - (0.48 \times \text{abdominal pain})\)
where, values of the parameters were considered by the same researchers as shown above. In that study, the researchers found that the prediction model had a sensitivity of 0.54[0.44; 0.63] and a specificity of 0.86[0.80; 0.90], and a ROC-AUC 0.77[0.72; 0.82]. Cross-validation ROC-AUC was 0.75[0.72; 0.77] (Menni, Valdes, Freydin, et al., 2020).

In both models, the strongest predictor was defined as the loss of smell and taste by the researchers (Menni, Valdes, Freydin, et al., 2020; Menni, Valdes, Freydin, et al., 2020). Therefore, certain SWRL rules to predict the probability of having Covid-19 score based on the Prediction Model 1 and Prediction Model 2 are developed. The rules are constituted and used to calculate the values of relevant Prediction Model 1 and Prediction Model 2 metrics from the input key metrics (e.g. age, sex, the availability of loss of smell and taste, the availability of persistent cough, the availability of fatigue, the availability of inappetence, etc.). While 8 rules are created to obtain Model 1, 10 rules are created to obtain Model 2.

In the rules, the “?p” variable indicates to a patient instance, while the “?gc”, “?lst”, “?pcc”, “?fwc”, and “?smc” variables indicate the values of the gender coefficient for a male/female, the loss of smell taste coefficient, the persistent cough coefficient, the fatigue weakness coefficient, and the skipped meals coefficient, respectively. After running first seven rules by IeM, the Rule #7 gives the result values of first six rules. The Rule #8 gives the result probability of having Covid-19 of a particular patient case. After the score is generated, some other SWRL rules of IeM assign this patient case to one of the following types of cases:

- the estimated “Higher Risk Covid-19 Case” result for probabilities greater than $p > 0.85$,
- the estimated “Medium-To-High Risk Covid-19 Case” result for probabilities between $0.5 > p > 0.85$, and
- the estimated “Control-Needed Case” result for probabilities less than $p > 0.5$.

Table 4 describes the rules created and operating principles developed for Model 1.

In next section, the SWRL rules and the preventive suggestions associated for a patient case are explained in detail. The rules and working principle developed for Prediction Model 1 are also applied similarly in Prediction Model 2.

(3) Rules on precautionary and personalized supportive action suggestions: The SWRL rules have been created to deduce the precautionary and personalized supportive suggestions for all SUs with patient account, who may or may not receive a doctor’s consultation through the system.

| TABLE 4 | The SWRL rules created to calculate the probability of having Covid-19 according to the Model 1 |
|-----------------|--------------------------------------------------|
| **Rules created for the Model 1** | **Explanation** |
| Patient(?p), has_Gender(?p, Male) -> has_Gender_Coefficient(?p, 1.0f) | If gender of the patient class instance (?p) is “Male”, then use “1” for “sex” variable to retrieve the prediction result for Model 1, shown in the Equation (1). |
| Patient(?p), has_Gender(?p, Female) -> has_Gender_Coefficient(?p, 0.0f) | If gender of the patient class instance (?p) is “Female”, then use “0”. |
| Patient(?p) ^ has_Symptom(?p, Loss_of_Smell_and_Taste) -> has_Loss_Of_Smell_Taste_Coefficient_m1(?p, 1.75) | If the patient suffers from the loss of smell and taste symptom, then use “1.75” as the coefficient for that symptom shown in the Model 1 equation. |
| Patient(?p) ^ has_Symptom(?p, Persistent_Cough) ^ has_Persistent_Cough_Severity(?p, Severe) -> has_Persistent_Cough_Coefficient_m1(?p, 0.31) | If the patient suffers from the “Severe” level of persistent cough symptom, then use “0.31” as the coefficient for that symptom shown in the Model 1 equation. |
| Patient(?p) ^ has_Symptom(?p, Fatigue_Weakness) ^ has_Fatigue_Weakness_Severity(?p, Severe) -> has_Fatigue_Weakness_Coefficient_m1(?p, 0.49) | If the patient suffers from the “Severe” level of fatigue symptom, then use “0.49” as the coefficient for that symptom shown in the Model 1 equation. |
| Patient(?p) ^ has_Symptom(?p, Skipped_Meals) -> has_Skipped_Meals_Coefficient_m1(?p, 0.39) | If the patient suffers from the skipped meals symptom, then use “0.39” as the coefficient for that symptom shown in the Model 1 equation. |
| Patient(?p) ^ has_Age(?p, ?a) ^ has_Gender_Coefficient(?p, ?gc) ^ has_Loss_Of_Smell_Taste_Coefficient_m1(?p, ?lst) ^ has_Persistent_Cough_Coefficient_m1(?p, ?pcc) ^ has_Fatigue_Weakness_Coefficient_m1(?p, ?fwc) ^ has_Skipped_Meals_Coefficient_m1(?p, ?smc) ^ swrlm:eval(?resM1, “(-1.32)-(0.01*a)+0.44*gc) + lst + pcc + fwc + smc”) | The {swrlm: eval (?resM1, “(-1.32)-(0.01*a)+0.44*gc) + lst + pcc + fwc + smc”)} statement provides to calculate the result value (?resM1) of the equation for Model 1. The result value will be input for the Rule 8 shown in next. |
| Patient(?p) ^ has_Prediction_Result_Model_1(?p, ?resM1) ^ swrlm:eval(?probOfM1, “(e^ (?resM1))/(1+(e^ (?resM1)))” , ?resM1) | The result value (?resM1) of Rule 7 is then converted to the predicted probability result (?probOFM1) through the formula $e^{(1-\frac{\text{result}}{1+\text{result}})}$. The value is between 0-100%. The resulting probability value is computed in [swrlm: eval (?probOFM1, “(e^ (?resM1))/(1+(e^ (?resM1)))” , ?resM1)], where ?probOFM1 is the result probability, e is exponential, ?resM1 is the output of Rule 7. |
As it is mentioned before, the research study (Christiano, 2020) investigated the patient cases with Covid-19 according to the severity categories such as mild, moderate, and severe. The study discussed that what symptoms make a case severe. The rules are categorized according to five types of case categories (Asymptomatic Case, Mild Covid-19 Case, Moderate Covid-19 Case, Severe Covid-19 Case, and Healed Covid-19 Case). IeM processes user-entered symptoms through SWRL rules and produces a set of precautionary and personalized supportive action suggestions for each case category mentioned above.

The relations between the observed symptoms in Covid-19 infection and the precautionary and personalized supportive action suggestions created in the system are presented in Table 5. In addition, corresponding recommendations and relevant SWRL rules for each case category are presented in Appendix C. For instance, if a user enters his/her fever info between 38.3 and 38.9, and entered at least two of these symptoms with severity level (e.g. chills with repeated shaking, deep cough, fatigue, body aches, muscle pain, general feeling of being unwell), then the user is probably considered a moderate case of Covid-19, which is CASE 03. The following rule is to identify a moderate case of patient according to the recent fever and the symptoms observed. The clarification of this rule is as follows: If “p” is an instance of patient class (Patient(?p)) AND if the number of moderate type of symptom of the patient is detected as greater than or equal to “2” AND if the fever value of the patient is detected as greater than or equal to “38.3” AND if the fever value of the patient is detected as less than or equal to “38.9” THEN the patient “p” is detected as a moderate case and associated with the “Moderate_Case_Of_Covid19_Patient (?p, true)” property.

Rule for detecting moderate case of Covid19 patients: Patient(?p) ^ (has_Moderate_Covid19_Symptom > = 2)(?p) ^ has_Fever(?p,?f) ^ swrlb:greaterThanEqual(?f, 38.3) ^ swrlb:lessThanEqual(?age,38.9) —> Moderate_Case_Of_Covid19_Patient (?p,true).

Another rule is a link to the previous rule, it assigns patients who meet all the above-mentioned conditions into the CASE 03 category. The clarification of this rule is as follows: if “p” is an instance of patient class (Patient(?p)) AND if the patient is detected as “true” with moderate case of Covid-19 patient THEN the case of the patient “p” is detected CASE_03.

Rule assigns the patients to CASE_03 category: Patient(?p) ^ Moderate_Case_Of_Covid19_Patient (?p,true) —> HAS_CASE (?p, CASE_03). Entire SWRL rules created on the OntCoV19 to produce precautionary and personalized supportive action suggestions according to case categories are given in Table 5.

8 | EXPERIMENTAL STUDIES

8.1 | Dataset used

The researchers (Xu et al., 2020; Xu et al., 2020) collected an individual-based epidemiological data set from national, provincial, municipal health reports, and online reports to support the analysis and tracking of the COVID-19 epidemic in Hubei, China. All data gathered are: (1) key dates (e.g. the date of onset of disease, date of hospital admission, date of confirmation of infection, and dates of travel history), (2) demographic information of the cases, (3) geographic information of the cases, (4) symptoms observed, and (5) additional information about the cases. A copy of the researchers’ dataset is able to be downloaded from figshare, which includes a fixed version of the data set at the time of submission, ranging from 1st December 2019 to 5th January 2020. The data set involves many attributes such as ID, Age, Sex, City, Province, country, latitude, longitude, wuhan(0)_not_wuhan(1), geo_resolution, date_onset_symptoms, date_admission_hospital, date_confirmation, symptoms, lives_in_Wuhan, travel_history_dates, travel_history_location, reported_market_exposure, additional_information, chronic_disease_binary, chronic_disease, source, date_death_or_discharge, location, and so on.

In our experimental studies, we were able to use only a few of these features, which were found suitable to run our system rules, are given below:

- ID - Unique identifier for reported case.
- Age - Age of the case reported in years. When it is not reported, N/A is used.
- Sex - Sex of the case. When it is not reported, N/A is used.
- Date_Onset_Symptoms - Date when the reported case was recorded to have become symptomatic.
- Date_Admission_Hospital - Date when the reported case was recorded to have been hospitalized.
- Date_Confirmation - Date when the reported case was confirmed as having COVID-19 using rt-PCR.
- Symptoms - List of symptoms recorded in the description of the case.

The researchers separated all patient data into two different datasets; the cases reported from Hubei, China, and the cases reported from outside of Hubei, China. The researchers presented two cleaned datasets for other researchers which are named as “clean-hubei.csv” and “clean-outside-hubei.csv”. There were 10,135 case records in the “clean-hubei.csv” dataset. Of these, only 27 cases contained appropriate symptom data (e.g. those with at least two symptoms, age, and important dates entered were selected) to verify our model and rule based IeM of the system.
**Table 5** The patient case categories and 22 different precautionary and personalized supportive action suggestions deduced by IeM

| Patient case categories | Precautionary and personalized supportive action suggestions |
|-------------------------|------------------------------------------------------------|
| **CASE 01**<br>Asymptomatic: No symptoms. | **SG_01**: Continue online tests.<br>**SG_02**: You need to wash hands with soap and water often and for at least 20 seconds.<br>**SG_03**: You need to practice good respiratory hygiene and avoid touching the eyes, nose or mouth with unwashed hands. |
| **CASE 02**<br>Symptoms in Mild Cases of Covid-19: Low-grade fever, dry cough, nasal congestion, runny nose, fatigue, sore throat, headache, new loss of taste and smell. | **SG_01**: Continue online tests.<br>**SG_04**: You need rest.<br>**SG_05**: Stay home and self-isolate even with minor symptoms such as cough, headache, mild fever, until you recover.<br>**SG_06**: You need to drink plenty of fluids.<br>**SG_07**: Consult your doctor to get an over-the-counter pain reliever that's best for you.<br>**SG_08**: Try to keep your temperature at normal levels such as by applying a cold towel, taking a warm shower.<br>**SG_09**: Have someone bring you supplies. |
| **CASE 03**<br>Symptoms in Moderate Cases of Covid-19: Fever >38.3 – 38.9, chills with repeated shaking, deep cough, fatigue, body aches, muscle pain, general feeling of being unwell. | **SG_01**: Continue online tests.<br>**SG_04**: You need rest.<br>**SG_06**: You need to drink plenty of fluids<br>**SG_07**: Consult your doctor to get an over-the-counter pain reliever that's best for you.<br>**SG_08**: Try to keep your temperature at normal levels such as by applying a cold towel, taking a warm shower.<br>**SG_09**: Have someone bring you supplies.<br>**SG_10**: Stay home and self-isolate even with mild or moderate symptoms such as cough, headache, mild fever, chills with repeated shaking, deep cough, fatigue, body aches, muscle pain, general feeling of being unwell, until you recover. |
| **CASE 04**<br>Symptoms in Severe Cases of Covid-19: Fever >39, deep cough, fatigue, body aches, breathing difficulties, chest discomfort, confusion / unresponsiveness, bluish face / lips, possible gastrointestinal issues: diarrhoea or nausea. | **SG_11**: You need rest in isolation in the hospital environment.<br>**SG_12**: You need to seek medical attention / hospitalization but call by telephone in advance if possible and follow the directions of your local health authority.<br>**SG_13**: You may need to get serum/fluid through IV.<br>**SG_14**: You may need to get support to breathe, so oxygen support in hospital. |
| **CASE 05**<br>Survive from coronavirus and no symptom | **SG_01**: Continue online tests.<br>**SG_02**: You need to wash hands with soap and water often and for at least 20 seconds.<br>**SG_03**: You need to practice good respiratory hygiene and avoid touching the eyes, nose or mouth with unwashed hands.<br>**SG_06**: You need to drink plenty of fluids.<br>**SG_15**: You need to take precautions to prevent re-infection.<br>**SG_16**: You need to use personal protective equipment.<br>**SG_17**: You need to stay at home and avoid crowded places.<br>**SG_18**: You need to take precautions to prevent re-infection.<br>**SG_19**: You need to use personal protective equipment.<br>**SG_20**: You need to stay at home and avoid crowded places. |
| **CASE 06 (Model 1):** Higher Risk Covid-19 Case<br>Probability of Model 1, if >85%, (Equation (1)) Higher Risk Covid-19 Case | **SG_12**: You need to seek medical attention / hospitalization but call by telephone in advance if possible and follow the directions of your local health authority.<br>**SG_16**: You need to use personal protective equipment.<br>**SG_18**: You are at higher risk of Covid-19 according to your Online Covid-19 test result, please do not be panic!<br>**SG_19**: Please observe your health conditions and symptoms closely.<br>**SG_20**: Consult your doctor for a physical examination as soon as possible.<br>**SG_21**: Have Real-Time PCR (RT-qPCR) test for SARS-CoV-2. |
| **CASE 06 (Model 2):** Higher Risk Covid-19 Case<br>Probability of Model 2, if >85%, (Equation (2)) | Same suggestions in **CASE_06**. |
| **CASE 07 (Model 1):** Medium-To-High Risk Covid-19 Case<br>Probability of Model 1, if >50% and < 85%, (Equation (1)) | **SG_01**: Continue online tests.<br>**SG_02**: You need to wash hands with soap and water often and for at least 20 seconds. |

(Continues)
In addition, there were 8392 case records in the "clean-outside-hubei.csv" dataset. Similarly, uncertain symptoms notes (e.g., respiratory symptoms, flu-like symptoms, lesions on chest radiographs), those with less than 2 symptoms data, or age data with "N/A," and so on, were excluded from the dataset. Only 142 case records were found suitable for our study.

In summary, to perform our experimental studies, 169 patient data were found suitable in total from the epidemiological dataset (Xu et al., 2020; Xu, Gutierrez, et al., 2020). The initial tests of the system were applied on these 169 patient cases.

8.2 | System verification studies

As it is mentioned before, the system was tested on 169 Covid-19 positive patient data. The online Covid-19 diagnostic test results produced by iE-M were compared with the real PCR test results of the same cases. In addition, a set of suitable suggestions deduced by iE-M for each case were compared with the written suggestions of a collaborated HE. Only 10 out of 169 cases randomly selected as sample cases are presented as a table in Appendix D due to page limitation. In addition, the score values deduced by iE-M, the actual PCR test results, the suggestions deduced by iE-M, and the suggestions produced by our collaborated HE, for each case are also presented.

In evaluation of the proposed system, three type of verification studies are applied, which are: (1) verification of the developed SWRL rules, (2) verification of the deduced suggestions, and (3) verification of deduced Covid-19 risk prediction scores and results.

(1) Verification of Semantic Web Rules of iE-M: As the first verification, manually verified 169 patient cases were sent to the system ontology to verify all created SWRL rules of iE-M. The user demographic data and symptoms data of each case were sent to the ontology for deducing the probability of having Covid-19 score through iE-M. In addition, iE-M deduced some precautionary and supportive action suggestions as an output for each case. The suggestions deduced for each patient case are kept for further evaluations in the next steps.

(2) Manual Verification of the Deduced Suggestions by Our Collaborating HE: As the second validation, a set of precautionary and personalized supportive action suggestions deduced by iE-M for each case were compared with the suggestions generated manually of our collaborated HE.
HE evaluated each of the suggestions deduced by IeM for each case. In addition, the HE determined a set of suggestions for each case using standard diagnostic procedures and professional reasoning after evaluating the observed symptoms and written health conditions of each cases in the data set. Then, the accuracy of the system in generating suggestions for cases was evaluated. The following methodology was used for calculating the system accuracy.

- The suggestions that are proposed by both the system and the HE is defined as the True Positive (TP).
- The suggestions proposed by the system but not proposed by the HE is defined as False Positive (FP).
- The suggestions that are not proposed by the system and are not actually suitable for the patient case are the True Negatives (TN).
- The suggestions that are not proposed for the patient case by the system but actually should have been proposed to the patient case are defined as the False Negative (FN).

Considering these metrics, the accuracy of the system are calculated based on the given formula:

\[
\text{accuracy}_{\text{suggestions}} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} = 1
\]  

Accuracy determines how accurate the system is, and how accurate the suggestions are proposed by the system (Shung, 2018). During the verification phase with our collaborated HE, 5 missing suggestions in our rule set were observed in the first round of tests and the missing suggestions were integrated to the appropriate SWRL rules in the system ontology. Therefore, in case of similar cases, these 5 suggestions will not be skipped anymore as a result of the inferencing. The accuracy for the first-round tests was calculated was equal to 0.93. Results in the second round showed that the system produced accurate and complete suggestions for all cases. In the second round of tests, the accuracy of the system performance increased to 1.00 after adding the 5 skipped suggestions to the system.

As an example, the suggestions of the system and our HE for only 10 patient cases is presented as a table in Appendix D.

**3) Verification of Deduced Test Scores and Results by IeM:** As the third validation, IeM predicted two scores for probability of having Covid-19 risk for each case based on the Prediction Model 1 and Prediction Model 2. These scores deduced by IeM for each case are then used to decide final Covid-19 risk prediction result as positive / negative. According to the study of Menni et al. (Menni, Valdes, Freydin, et al., 2020; Menni, Valdes, Freydin, et al., 2020), the models used are assigning “estimated Covid-19 cases” for probabilities >0.5 and “control Covid-19 cases” for probabilities <0.5 (Menni, Valdes, Freydin, et al., 2020). Therefore, our system also applies the same strategy while producing the final Covid-19 risk prediction result by IeM. Ultimately, the result of each case as positive / negative was compared with the actual PCR positive / negative test results (Figure 10).

The threshold line (red line) is set at 50% on the result scores graph in Figure 10. Those cases above the line are predicted as “estimated Covid-19 cases”, while those below the line are considered “control Covid-19 cases” to be followed. Brown and blue lines show the scores for probability of having Covid-19 risk deduced by IeM according to the Model 1 and Model 2, respectively. Since all 169 cases in our data set are positive Covid-19 patients (via PCR test), the black line shows the score as 100% for each case. In other words, the 169 cases are positive cases confirmed by PCR tests in the clinical setting, as presented in the original data set.

As it can be seen on the graph, most of result scores deduced by IeM have been obtained above 50%. In other words, the system concluded these cases as “estimated Covid-19 cases.” Of 169 positive patient cases, 10 patients were predicted as negative according to the Prediction Model 1, while 24 patients were predicted as negative according to the Prediction Model 2. The system achieves 0.94 accuracy when operating the SWRL rules according to the Prediction Model 1, while it achieves around 0.86 when operated according to the Prediction Model 2.

As an example, the scores deduced by IeM and real PCR test results for only 10 patient cases is presented as a table in Appendix D.

9 | DISCUSSIONS

It is a known fact that the world still does not have sufficient information about Covid-19. With the knowledge obtained from real Covid-19 patients hospitalized, it seems quite challenging to reach the final solution in a short time. First of all, the major problem about the disease is that those with mild symptoms or asymptomatic ones spread the Covid-19 virus and its derivatives to the whole community. The world has been warned about that and we need to take some precautions (e.g. attention to social distance, quarantine rules, attention to hygiene, etc.) to prevent this risk from becoming more common in the community. As a result, significant data and findings obtained from recent research studies will be the main building blocks of Covid-19 struggle systems in the future. Therefore, such the self-pre-diagnosis systems to predict and monitor Covid-19 infection can be helpful to alert individuals and the governmental health authorities, to take precautions and reduce the risk of contamination among the individuals.
A search for “coronavirus” or “Covid-19” on Apple’s App Store reveals few apps definitely centered around the outbreak and almost no spam (Kif Leswing, 2020). iPhone developers mentioned to CNBC that, Apple company has rejected the submission of mobile apps that aren’t from well-known institutions such as governments or hospitals in relation to the Covid-19 pandemic. Thus, due to this policy enforced by Apple, as well as the fact that the necessary approvals from our government’s health authorities have not yet been obtained, the mobile application of the proposed system could not yet be able to release for the usage of smartphone user in TRNC.

As it is mentioned before, in system evaluations, an individual-based epidemiological data set (Xu et al., 2020; Xu, Gutierrez, et al., 2020) was used when predicting infection probability in smartphone users using their demographic and symptoms observed. In order to test the accuracy of IeM, not only positive cases but also asymptomatic individuals with positive PCR result data were included in the data set. The findings obtained and limitations observed from the experimental studies are explained in detail below.

9.1 Limitations

Menni et al. have revealed the most significant symptoms (e.g. “Loss of smell and taste”, “fever”, “abdominal pain”, etc.) of Covid-19 risk by examining millions of patient cases (Menni, Valdes, Freidin, et al., 2020; Menni, Valdes, Freydin, et al., 2020). In addition, the researchers proposed two models based on these symptoms which are used in the proposed system rules. Our experimental studies show that there was no significant difference between the results of IeM and the actual PCR test results in our data set provided that some key symptomatic findings were observed in the cases studied.

As it is mentioned, the system involves a set of patient case categories (CASE 01 - CASE 08). The patient case categories indicate 22 different precautionary and personalized supportive action suggestions which are deduced by IeM for cases. In order to run the appropriate categories, the system must send adequate and appropriate demographic and symptom data to IeM as input. However, the researchers who collected the data set could not enter complete and consistent symptom data for some cases (Xu et al., 2020; Xu, Gutierrez, et al., 2020). Thus, some of our predefined case categories of the system could not be deduced and assigned by IeM for the 169 cases. Consequently, IeM could not generate the suitable precautionary and supportive action suggestions for suitable cases belong to these categories.

In addition, the data field of presence of pre-existing Covid-19 infection or hospitalization were missing in their data set. However, some of the predefined case categories of the proposed system also need to use these data to produce proper suggestions for suitable cases who had Covid-19 background in past. Therefore, IeM could not able to assign any case to the CASE 05 in the data set and produce any suggestions for any case according to it.

Moreover, measured fever values of the cases in the original data set were also missing in majority of 169 patient case. Instead of fever values, the presence / absence of patient fever was noted in the field of symptoms. The presence of the term “fever” meant “the participant has a fever” and the absence “the participant has no fever. Due to the CASE 02, CASE 03, and CASE 04 require specific measured fever data, IeM could not found any case from these categories and generate any suggestion for any patient.

To observe that IeM is actively working in these categories as well, after obtaining the necessary approvals, we aim to begin collecting our data set in a real clinical environment, in accordance with our symptom-oriented patient diagnosis form presented in Appendix E. The step of gathering data based on the required data fields (according to our patient form) from real patients in the TRNC and obtaining the necessary legal
permissions to run with these data has been initiated. In addition, it is expected that the necessary legal permissions must be obtained from government authorities in order to spread this mobile application in TRNC and encourage the use of individuals.

In fact, for any data set, if the demographic and symptoms data is complete, IeM already has sufficient semantic rules to determine any suitable case category for any patient and can infer the proper suggestions for that patient.

While all these findings in this study are promising, more randomized and longer-term studies on real-time patient data are needed to evaluate the effectiveness of the proposed system. Therefore, as the next step of our study, we planned it as a case-control study by collecting data both through the mobile application of our system and with full real patient data. Designing such expert systems convenient for the clinical characteristics of each community is of great importance in terms of the accuracy of follow-up data in Covid-19 management.

10 | CONCLUSION

In this study, a knowledge-based self-pre-diagnosis system is designed to predict Covid-19 risk in smartphone users using their demographic and symptoms data. The potential users are smartphone users, healthcare experts and government health authorities. The system does not only share the data gathered from the users with experts, but also analyzes the symptom data as a diagnostic assistant to predict possible Covid-19 risk. In addition, the system deduces and presents a set of proper precautionary and supportive action suggestions to users.

The system has its own an ontology knowledge-base and semantic rules and can be used as a predictive tool of Covid-19 risk in distant areas from clinics or hospitals. The major parts of the system are: (1) Symptom-oriented Covid-19 disease ontology for self-pre-diagnosis—OntCoV19, (2) semantic Web rule knowledge base, and (3) an ontology-based inference engine (IeM). The system ontology (OntCoV19) is formalized in OWL and implemented as a system ontology. In addition, the rule knowledge base is formalized in SWRL and stored in the OntCoV19. The rules examine the data gathered from smartphone users such as user demographics, observed symptoms, certain medical measurements, presence of pre-existing medical conditions or any hospitalization, latest PCR test results, and so forth.

The system was tested on 169 Covid-19 positive patient data. In the experimental studies, three type of verification studies were applied: (1) verification of the developed SWRL rules, (2) verification of the deduced suggestions, and (3) verification of deduced Covid-19 risk prediction scores and results. The test results produced by IeM were compared with the real PCR test results of the same cases. All results are also verified manually by a collaborated HE. In addition, a set of suitable suggestions deduced by IeM for each case were compared with the written suggestions of a collaborated HE. Finally, one detailed case study has been discussed to present how the system gathers medical data from smartphone users via its mobile application.

As a result, the result findings of our experimental studies are promising in demonstrating the applicability, effectiveness and efficiency of the proposed system. Such a system can be extensively used in home by smartphone users. It is hoped that the system will become a model in determining the size of the ever-increasing Covid-19 spiral, in patient follow-up and treatment processes.

In future studies, an image processing mechanism is planned to add as a separate module to automatically detect Covid-19 patients using digital chest x-ray images and then determine the improvement level of the patients over time. In addition, using deep learning approaches, it is planned to examine the respiratory sound data entered daily by users through the mobile application of the system in order to predict Covid-19.

CONFLICT OF INTEREST

The author declares that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

DATA AVAILABILITY STATEMENT

no shared data currently.

ORCID

Duygu Çelik Ertuğrul @ https://orcid.org/0000-0003-1380-705X
Demet Çelik Ulusoy @ https://orcid.org/0000-0002-5542-9467

ENDNOTE

1 https://figshare.com/articles/dataset/covid19_zip/11949279/4

REFERENCES

Alepis, E., & Lambrinidis, C. (2013). M-health: Supporting automated diagnosis and electronic health records. Springerplus, 2, 103.
Barbosa, A., Carvalho, C., Steibel, F., & Costa, J. (2020). Colombia's CoronApp. https://feed.itsrio.org/colombias-coronapp-5802ceb54b7.
Baysari, M. T., & Westbrook, J. I. (2015). Mobile applications for patient-centered care coordination: A review of human factors methods applied to their design, development, and evaluation. Yearbook of Medical Informatics, 1, 47–54.
World Health Organization. (2020c). Report of the WHO-China joint mission on coronavirus disease 2019 (COVID-19). World Health Organization. (2020b). Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). https://www.who.int/docs/default-source/coronaviruse/whochina-joint-mission-on-covid-19-final-report.pdf

World Health Organization. (2020a). Report of the WHO-China joint mission on coronavirus disease 2019 (COVID-19).

Yu, H., Guo, Y., Xiang, Y., Sun, C., & Yang, T. (2020). Data-driven discovery of a clinical route for severity detection of COVID-19 paediatric cases.

Zens, M., Brammertz, A., Herpich, J., Südkamp, N., & Hinterseer, M. (2020). App-based tracking of self-reported COVID-19 symptoms: Analysis of questionnaire data. Journal of Medical Internet Research, 22(9), e21956.

AUTHOR BIOGRAPHIES

Duygu Çelik Ertuçrul, was born in Turkey. She is an associate professor in the computer engineering department at the Eastern Mediterranean University, Famagusta, North Cyprus via Mersin-10, Turkey. Her research topics are related to the Web and Semantics; Composition and Discovery of Semantic Web Services, Semantic Search Agents, Rule-Based Expert Systems, m-Health and Healthcare Knowledgebase Expert Systems. She is one of the organizers of two international workshops and one international symposium: “IEEE International Workshop on ESAS: E-Health Systems and Semantic Web” since 2006. “Security of Information and Networks (SIN) between 2007-2011” and “IEEE COMPSAC Symposium on Web Technologies & Data Analytics (WEDA) in 2016“. She also supervises several research and development projects supported by the university, government, and industrial companies. She has organized a special issue for Expert Systems: The Journal of Knowledge Engineering and served as guest editor. She has published numerous articles/book chapters/book in several international/national journals and conferences on the topic of Web Semantics and Mobile Medical Healthcare Services and Systems. She is also the author of various books or book chapters about ‘Composition or Discovery of Semantic Web Services’ and ‘Semantic Web based e-Health Services’.

Demet Çelik Ulusoy, is an assistant professor of law at the Eastern Mediterranean University, North Cyprus. She was born in Turkey, on April 23, 1978. She graduated from Faculty of Law (LLB) in 2000. She received her LL.M. degree in 2003 in EMU, Faculty of Law in the field of constitutional law with thesis title “Judicial Review in Turkey Republic of Northern Cyprus”. She received her Ph.D. in 2013 from Ankara University, Turkey with the thesis on “Federal Systems and Judicial Review of Federalism Conflicts”. She joined Faculty of Law, EMU, in 2006 as a research assistant where she became Dr. and an assistant professor in 2013. I am an assistant professor in constitutional law field at Eastern Mediterranean University, Faculty of Law. Governmental system is one of the important issues of constitutional law. Besides being one of my special interests, I have been giving contemporary governmental systems course in the Faculty for 3 years. Except for my books and editors published in the field of constitutional law, I have national-international publications and conference papers. I am already publish a book concerning personal data protection and right to protection of personal daria. She also serve as international joint graduate program coordinator at the Faculty of Law with Bologna University.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Çelik Ertuçrul, D., & Çelik Ulusoy, D. (2022). A knowledge-based self-pre-diagnosis system to predict Covid-19 in smartphone users using personal data and observed symptoms. Expert Systems, 39(3), e12716. https://doi.org/10.1111/exsy.12716