



Pandemic tele-smart: a contactless tele-health system for efficient monitoring of remotely located COVID-19 quarantine wards in India using near-field communication and natural language processing system

Vishal Balasubramanian¹ · Sapthagirivasan Vivekanandhan^{2,3} · Venkatesh Mahadevan⁴

Received: 28 October 2020 / Accepted: 7 October 2021 / Published online: 27 October 2021
© International Federation for Medical and Biological Engineering 2021

Abstract

Efficient remote monitoring of the patient infected with coronavirus without spread to healthcare workers is the need of the hour. An effectual and faster communication system must be established wherein the healthcare workers at the remote quarantine ward can communicate with healthcare professionals present in specialty hospitals. Incidentally, there is a need to establish a contactless smart cloud-based connection between a specialty hospital and quarantine wards during pandemic situation. This paper proposes an initial contactless web-based tele-health clinical decision support system that integrates near-field communication (NFC) tags and a smart cloud-based structuring tool that enables the quick diagnosis of patients with COVID-19 symptoms and monitors the remotely located quarantine wards during the recent pandemic. The proposed framework consists of three-stages: (i) contactless health parameter extraction from the patient using an NFC tag; (ii) converting medical report into digital text using optical character recognition algorithm and extracting values of relevant medical-parameters using natural language processing; and (iii) smart visualization of key medical parameters. The accuracy of the proposed system from NFC reader until analysis using a novel structuring algorithm deployed in the cloud is more than 94%. Several capabilities of the proposed web-based system were compared with similar systems and tested in an authentic mock clinical setup, and the physicians found that the system is reliable and user friendly.

Keywords Telemedicine · Mobile health · Remote consultation · Clinical decision support systems · Electronic medical records

1 Introduction

Coronavirus disease 2019 (COVID-19) is a communicable disease caused by severe acute respiratory syndrome coronavirus 2. As on 15th of June 2021, the disease has infected

over 175.84 million people, resulting in a death of 3.80 million people globally [1]. Identified first in December 2019 in Wuhan, this disease mainly spread fomites and by small droplets produced when those infected coughs, sneezes, or talks [2]. The coronavirus travels to the deeper regions of the lungs and causes inflammation, leading to difficulty in breathing, dyspnea, and possibly pneumonia. The World Health Organization declared that the 2019–2020 coronavirus outbreak as a public health emergency of international concern [3] on 30 January 2020 and a pandemic on 11 March 2020 [4]. There have been various disease control protocols adopted by several governments across the world to stop the spread of the disease. But the issue with most of them is that they are prone to breaches since tracing of suspected patients and monitoring of infected patients is challenging.

Since the disease spreads rapidly, breaking the chain of transmission is of prime importance. Few countries have

✉ Sapthagirivasan Vivekanandhan
sapthagiri.ece@gmail.com

¹ Department of Electronics & Communication Engineering, Rajalakshmi Engineering College, Chennai 602105, India

² Department of Biomedical Engineering, Rajalakshmi Engineering College, Chennai 602105, India

³ Medical Devices and Healthcare Technologies Department, Engineering R&D Division, IT Service Company, Bengaluru 560066, India

⁴ The Queen Anne School of Management, Sydney, NSW 2000, Australia

adopted the strategy of “T3” that is “Test”, “Trace”, and “Treat”. In this method, people with symptoms and those suspected of exposure to the virus are tested, and if found positive, their activities are backtracked to trace the individuals with whom they have come in contact. It was initially used for surveillance of diseases like “malaria” and has been extrapolated by various countries for COVID-19 [5]. This enables efficient testing rather than testing people randomly across a large country. Infected patients are then continuously monitored in quarantine wards until they test negative again. This method requires a system to visualize the progress made by the infected patients and the people who have met them. Answering to the technology needed for tracing of individuals, the Government of India came up with a mobile application named “Aarogya Setu” that leverages Wi-Fi, GPS, and Bluetooth to track the movement and interaction of a possible infected person with other people [6].

In this regard, several governments across the globe implemented a complete lockdown in their countries, thereby banning free movement of people in public places except for essential purposes. In the process of lockdown during pandemic, three levels of disease transmission control protocols are followed generally [7].

- The first level ($L1$) of which is self-quarantine or self-isolation at home, where the people with symptoms like that of COVID-19 isolate themselves from the public on moral grounds.
- The second level ($L2$) is quarantine in special wards such as a mobile quarantine ward, which is located outside the hospital, where the movement of people who were exposed to the disease is strictly restricted. They are periodically visited by front-line healthcare professionals such as physicians and nursing staffs to monitor the health condition of the person under quarantine.
- The third level ($L3$) is isolation, where the COVID-19-infected patient is completely isolated and offered treatment in quarantine wards (in hospital) with or without life assistive device’s support.

Mostly, in $L2$ and $L3$, the primary person of contact with the patients is the healthcare workers (front-line healthcare professionals such as physicians and nursing staffs and other supporting workers such as sanitation workers) who are easily exposed to the disease [8]. The spread of the disease can occur between an infected person and a healthcare worker, even if the infected takes all the essential precautions [9]. If a front-line healthcare professional gets sick with an infectious disease, the likely spread might occur to other healthcare workers or vulnerable patients within the healthcare facility [10]. This can include patients with a deteriorated immune system and may be at-risk for grave complications. Hence,

the healthcare workers often resort to self-isolation and social distancing from their family members and the public, in general, leading to depression over a period [11]. Cases of death of front-line healthcare professionals due to exposure to the disease have also been reported in several countries [12] especially in the USA, over 510,242 healthcare workers have been infected by COVID-19 during this pandemic and that 1653 have died [13]. These numbers keep rising even though governments across the world have provided physicians and other front-line professionals with personal protective equipment and standard operating procedures for patient diagnosis, treatment, and management. These concerns create panic within the healthcare community during this pandemic, which raises demand for safe and reliable system while testing or treating the COVID-19-infected patients.

On the other hand, in order to isolate the infected people from the public, countries like India have come up with mobile quarantine wards, wherein, railway coaches are converted into isolation units along with ventilators and necessary support systems [14]. The healthcare workers visit these mobile quarantine wards periodically to monitor the condition of the infected patients, to get the second opinion in the critical situations, and to update the senior physicians and officials present in the main specialty hospitals. There are several challenges in the practice of mobile quarantine wards faced internally. The major challenge that acts as a hindrance to this concept is the difficulty in proper maintenance and monitoring of the patient’s health records during the quarantine period [15]. Although, the mobile quarantine wards might be helpful to isolate the people from the rest of the public, it still does not eliminate the contact between the patient and the front-line healthcare professionals. Similarly, the data obtained from patients are usually comprehensive and narrative. This makes it difficult to interpret and visualize the various parameters of a patient for comparison and faster treatment.

To address these issues, we propose a novel framework that integrates near-field communication (NFC) chip and a smart cloud-based analytical tool to achieve a contactless or minimum-contact patient monitoring and clinical decision support system. The NFC chip will enable the medical officer to pull out the medical history of the corresponding patient from the cloud database easily. There are several reasons and advantages in usage of NFC chips, which can be a viable alternative for presently used wristbands.

- Firstly, every patient is given a separate NFC chip that acts as a linking medium to the patient’s database which minimizes the interaction of front-line healthcare professional with the patient [16–18].
- Secondly, since every patient has a unique NFC chip, the chance for misdiagnosis is reduced. Throughout the history

of healthcare, the failure to properly identify the patients has led to severe medication errors, transfusion errors, testing errors, and wrong person procedures. Cases of patient misidentification and misdiagnosis have been reported in countries such as the UK and the USA [19, 20].

- Finally, the usage of NFC chips solves the issue of protecting the physician from unwanted exposure to the infected patient.

The above reasons make NFC tags a viable solution in the previously discussed “Test” and “Treat” methods in “T3”. NFC tags can be extensively used for people in *L2* and *L3*.

The cloud-based smart structuring tool will assist the healthcare professionals to effortlessly organize the data and integrate it with the existing patient database. This tool leverages “Keyword Extraction” techniques and natural language processing (NLP) for structuring of narrative clinical texts in a structured format. It also has an inbuilt optical character recognition (OCR) application that aids in the conversion of text within the images (medical reports) into digital format [21]. Traditional medical records are available in the form of hard copies, thus making the data in them difficult to be integrated with the database. Thus, the proposed system which integrates NFC, NLP, and OCR technologies can assist the physicians in an improved manner: to monitor the progress and at the same time by maintaining minimum interaction with the people quarantined in *L2* and the treatment progress of the patients who are in *L3* and prescribe a treatment plan appropriately.

2 Methodology

The proposed system, of monitoring isolated and quarantined patients in *L2* and *L3*, respectively, consists of the stepwise processing of the input data which can either be structured or unstructured. These steps can lead to better visualization of data and easy monitoring of the progress of the patient under quarantine. The proposed system mainly focuses on the testing and treating stages of the “T3” strategy. Therefore, people who have tested positive must be monitored continuously for the development of symptoms. On the other hand, the infected patients undergoing treatment must be monitored for signs of recovery. The clinical workflow and detailed process flow of the proposed pandemic tele-health system are depicted in Figs. 1 and 2, respectively.

2.1 Architectural flow

2.1.1 Contactless connection of patient record using the NFC chip

The NFC chip or NFC tag used in this system will act as connecting tool to the patient’s cloud database. Every

patient in the mobile quarantine ward will be given an NFC chip which will link to that corresponding patient’s database. Contrary to the current scenario, where the patients are given with wristbands for identification, the front-line healthcare professional need not come in contact with the infected patient since the NFC chip can be placed near the bedside of the patient. The proposed passively powered NFC tag (NTAG, NXP, Dutch) works seamlessly with our NFC reader ICs, which are used in more than 90% of all NFC-equipped mobile handset models [22]. NTAG21x NFC solutions can store NFC data exchange format (NDEF) formatted data, making them fully compatible with every NFC-enabled phone for contactless smart cards.

The major reason for using an NFC tag and not any other similar technologies such as long-distance RFID chips is that most of the market available smartphones have inbuilt NFC reading capabilities. The inbuilt NFC reader eliminates the need to purchase a RFID reading module, thereby making the proposed model more feasible and economical. Another reason is that, in the case of long-distance RFID tags, the chances of interference from other nearby RFID tags is very likely, which might result in incorrect patient identification.

The front-line healthcare professional can thus connect to the database of the corresponding patient by bringing their NFC-enabled mobile device closer to the NFC tag. Figure 3 depicts the process of uploading medical records to the cloud-based server through an NFC-enabled mobile device. From now on, the physician can add additional records of data to the system which can be in the form of both structured data and unstructured data such as narrative text, images of scanned report, images, and videos of radiological. So, every time the physician connects with the database through the NFC tag, a new record is created in the cloud along with the corresponding date and time.

2.1.2 Cloud-based smart structuring tool

The smart cloud-based structuring tool is used to extract text contents present in images of medical reports and structure the general medical data, using NLP to fit them into the database and obtain large volumes of rich medical data.

Conversion of images of medical reports into digital text using OCR

Traditionally, the medical reports (such as laboratory reports, radiological findings, and other reports) are predominantly either printed on papers or stored in the form of digital narrative summary. Often majority of these reports printed on paper needs to be converted as images for storing in digital health records to maintain security, privacy, and confidentiality to avoid the data to be intercepted in its transmission and be immune to hacker attacks. The conversion of the text present inside an image into a text file is important for further structuring and storing in database seamlessly. In

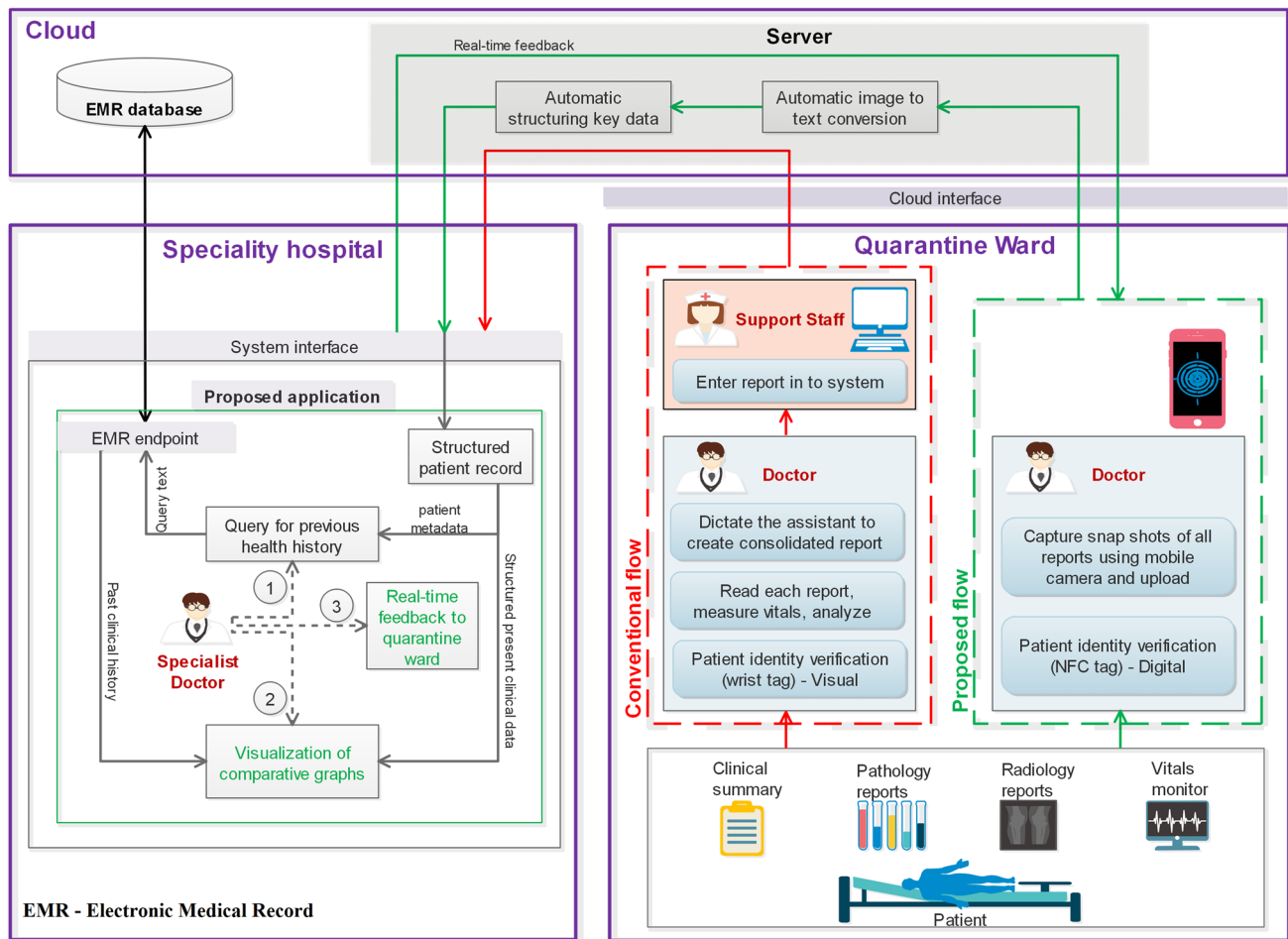


Fig. 1 The clinical workflow of the proposed pandemic tele-health system with comparison of conventional workflow

the proposed system, we leverage Microsoft’s OCR Engine which is available as an application programming interface (API) under its computer vision suite [23].

Novel structuring of narrative text by extracting keywords using NLP The unstructured text obtained from the image through OCR and the narrative text given as input by the front-line healthcare professional is processed to extract the required keywords so that it can be integrated with the database. The entire paragraph is broken down into sentences using the natural language toolkit (NLTK) function. After which, the words or phrases in that sentence are marked, based on the part of speech (POS) using rapid automatic keyword extraction (RAKE) algorithm. This is called as POS tagging. Now, the identifiers for every field value are defined, and the tagged words are now structured and sorted into their respective fields. A total of 57 medical parameters that are most noted with patients suffering from COVID-19 are included in this system for automatic detection and extraction.

There are three steps involved in the process of the proposed novel keyword extraction algorithm. It is a combination of POS tagging through RAKE algorithm and extraction of medical parameter values using keyword identification and locating through bag of words (BoW). Also, the reason for selection of this algorithm is explained in the *Comparison of Keyword Extraction Algorithm* subsection of the “Discussion” section. The first step is the selection of clinically relevant data through scissoring of paragraphs into lines. This ensures that the memory occupancy is low, and the speed of execution is high. In the second step, the use of a rule-based POS tagger (Brill tagger) is applied on clinical data captured and uploaded. In addition, the medical parameters are located using BoW, and their values are searched in the neighboring text elements. The list of medical parameters along with the keywords is given in the Supplemental Table S1. In the third step, the extracted keywords need to be placed in the respective fields. We acknowledge that the identifiers are applied extensively to place the word in the appropriate field by using all the possible combinations of

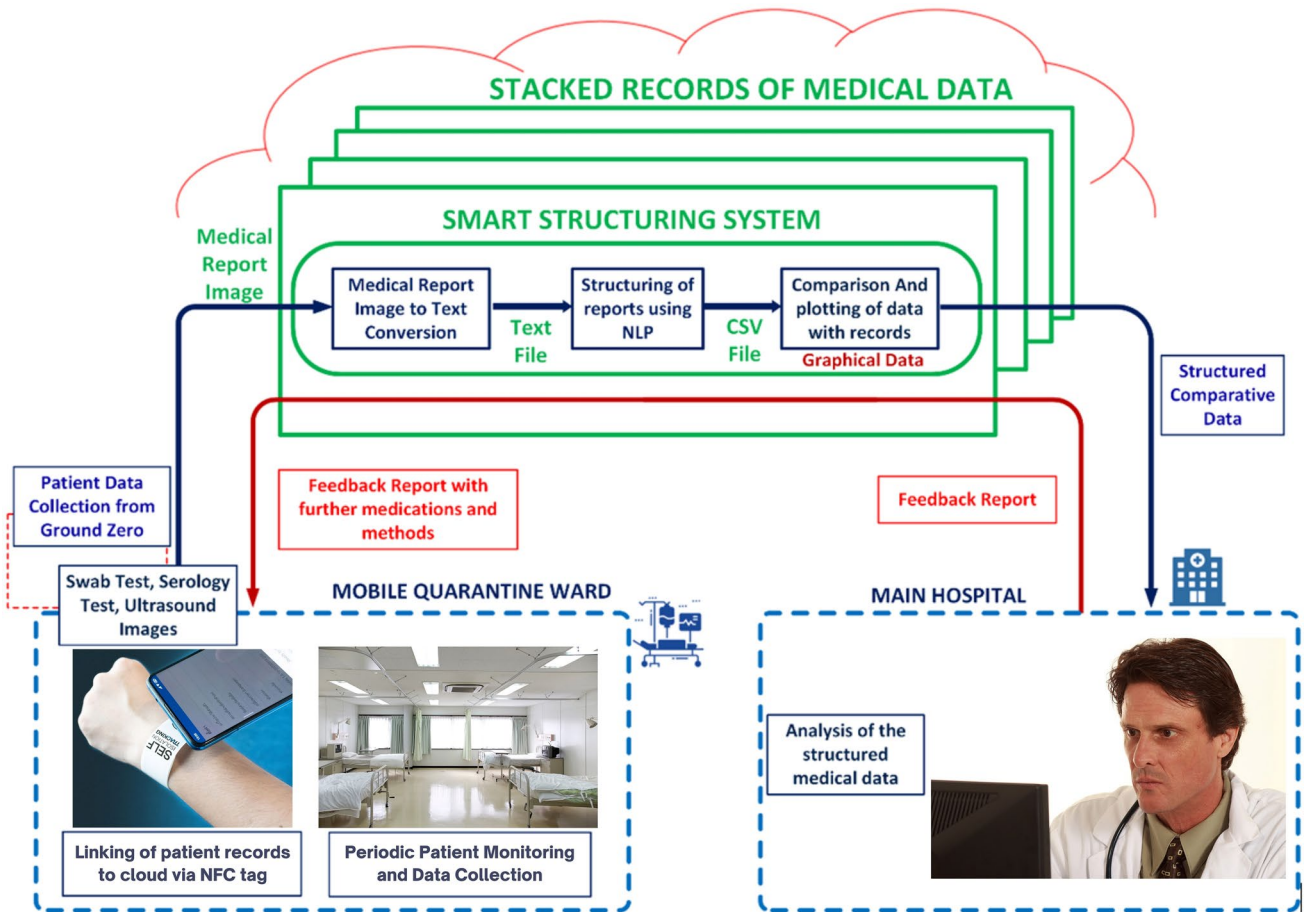
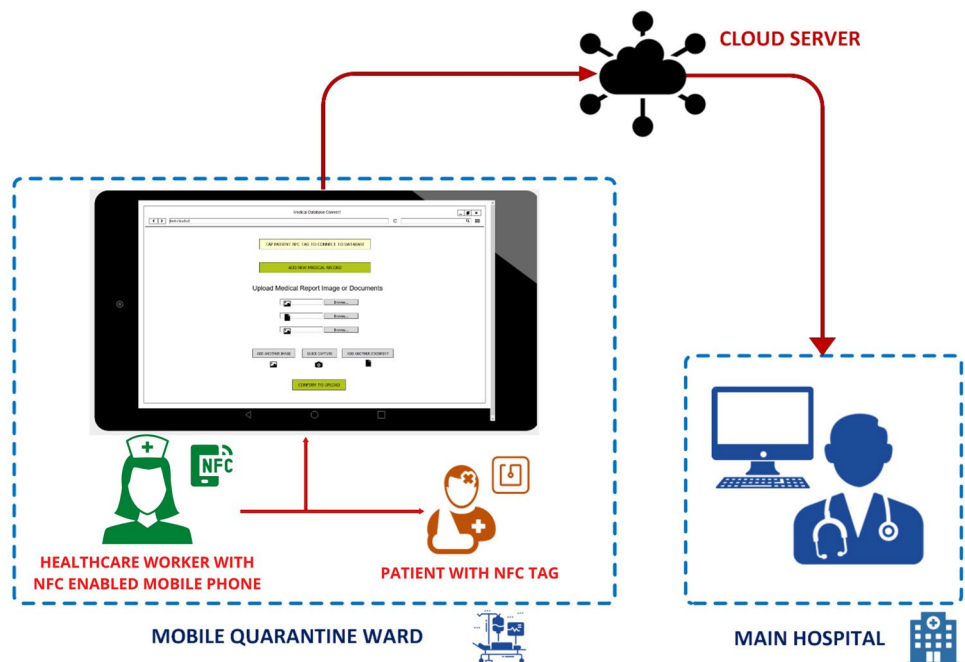


Fig. 2 The process flow of the proposed cloud-based pandemic tele-health system in which the patients are to be monitored at a mobile and a remote quarantine ward

Fig. 3 The process of uploading of medical records to the database by the healthcare worker at the mobile quarantine ward by reading the patient’s NFC tag



words, which we call keywords. For example, to place the value 120 mmHg in the field of systolic blood pressure, we rather instruct the program to use {Systolic-Blood-Pressure, SBP, BP, etc.} as the identifiers. Similarly, to find the age of the patient, we tend to use various identifiers such as {'yo', 'years', 'year', 'old', 'aged', 'age', 'yr', 'yrs'} and {'male', 'man', 'boy', 'woman', 'female', 'girl', 'transgender', 'unclassified', 'third-gender'} for gender detection. By doing so, we tend to improve the performance of the system by increasing the efficiency of tagging. The flowchart of the NLP model is shown in Fig. 4. However, the keywords can be created in accordance with the current COVID-19 pandemic situation, and healthcare workers are encouraged to use medical parameters and serological test findings commonly found in COVID-19.

Now to locate the field value, we set a search threshold value. This means that the program will search for the value of the identifier within a range of $\pm n$ words. For example, if $n = 2$, to find the value of systolic blood pressure, the program will search for a number within the range of 2 words on either side of the identifier. The search operation stops when a special character such as a conjunction is encountered. This can be better understood as shown in Fig. 5.

The value after being placed in the respective fields is also being placed in the database. Defining the locating

parameters of every keyword to be extracted is of vital importance to all concerned. Keeping this in mind, an exhaustive list of identifiers to locate the keyword in deployed, physicians are encouraged to follow the structure of report as depicted in Fig. 6. This is the dataset that has been followed throughout this work and was prepared after consultation with several physicians.

It is important to note that the elements such as date of birth, sex, vital signs, and medications have a limited number of options that are easy to select from a list or enter quickly in an agreed-upon format. But in reality, none of the unstructured data is easily corralled into a format that can be fed into algorithms and fed back to clinicians in an intuitive manner, yet providers desperately need better ways to understand and leverage these critical sources of information to care for their patients.

2.1.3 Integration

The last part of the proposed step before the deployment of the system would be the integration of the OCR and the NLP into the cloud and a creation of a web application for uploading the records on the mobile quarantine ward side and visualization of the data on the hospital side. As the

Fig. 4 Steps involved in “keywords” extraction from the medical report, which was converted into text files from image format using OCR

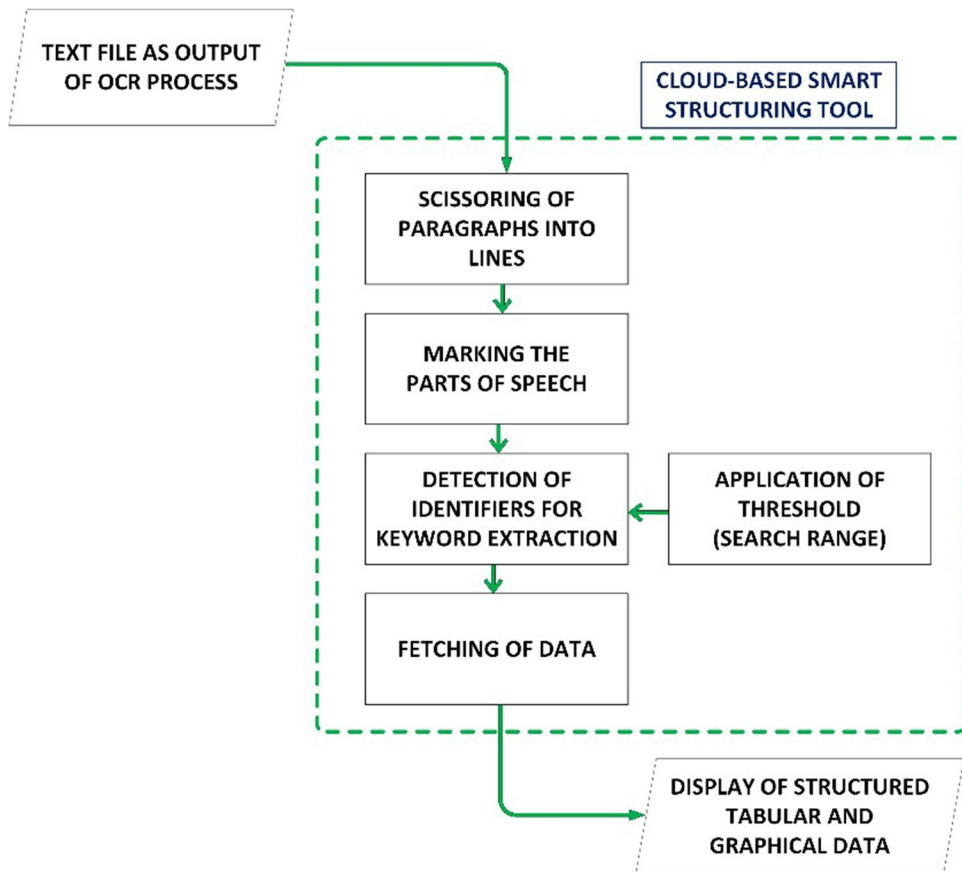


Fig. 5 Extraction of “parameter values” for each medical data through keyword detection and neighbor search algorithm

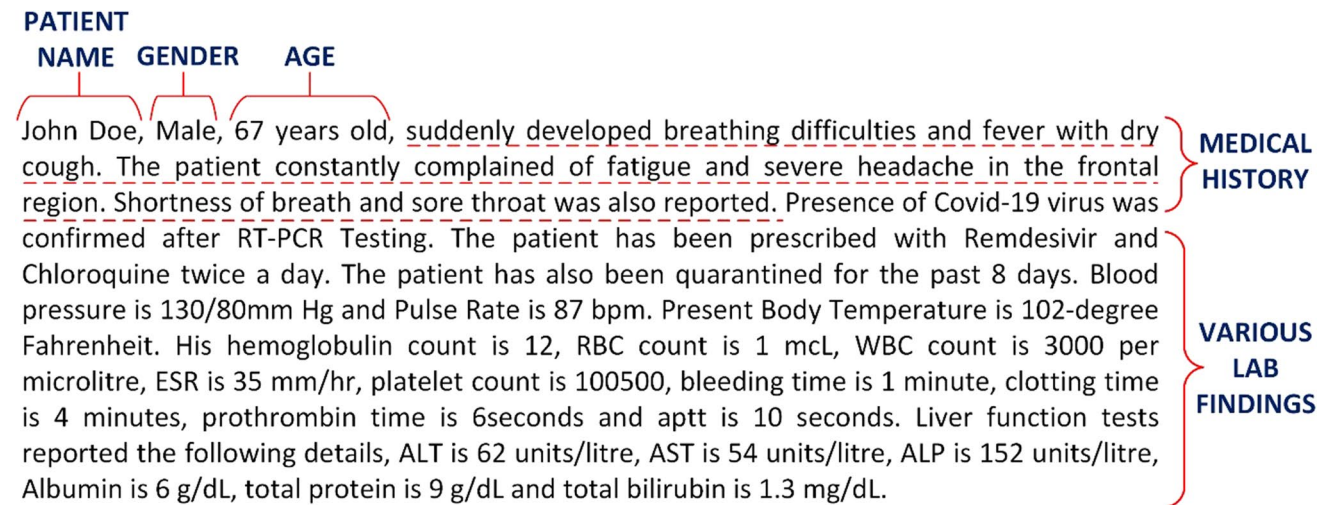
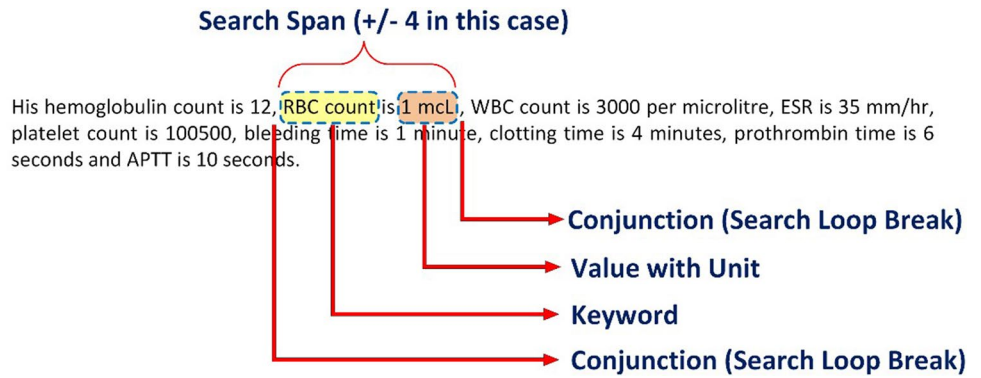


Fig. 6 Identification of various keywords from narrative text of the medical reports uploaded using POS tagging algorithm and represent as tabular data for ease of reading and adding them to the EMR database

end users are physicians, the user interface (UI) must be as user friendly as possible. The graphical UI of the proposed system is depicted in Figs. 7 and 8.

On the quarantine ward end, the application denotes the date and time at which the previous record was updated, so that the front-line healthcare professional will come to know if it would be the appropriate time to upload the next record and that they can maintain a frequency at which they update records. Since the content they upload can be in the form of images or a saved document, they can effortlessly upload them with a click of the button. If the content to be uploaded is from the mobile device, the user can add more upload field by clicking on the “add an image/document from device” button. The “capture image from device camera” button enables them to add images directly from their mobile camera. It is interesting to note that this feature will come in handy to upload images of the readings from the support system, in the case of patients who are in L3 of isolation, or to upload pictures of the patient conditions.

It is also important to verify if the record is being added to that corresponding patient’s database. This is done by capturing and uploading the facial image of the patient every time a report is taken. The facial picture of the patient that was captured on the day of admission of the patient to the ward is used for comparison with the facial pictures that are captured in the subsequent days of the visit by the healthcare professional. This is done using Microsoft Azure’s Facial Analysis Cognitive Services API [24]. A similarity score is generated between the two images, thereby adding an additional layer of precaution so that there is no wrong patient identification.

At the hospital end, the stack of data is processed, and the previously mentioned OCR and NLP algorithms are applied to get structured data that can be visualized as shown in Fig. 8. A hospital with multiple wards, even in multiple locations, can monitor the progress of the patients in a visually appealing and in a faster manner since data is structured and graphical.

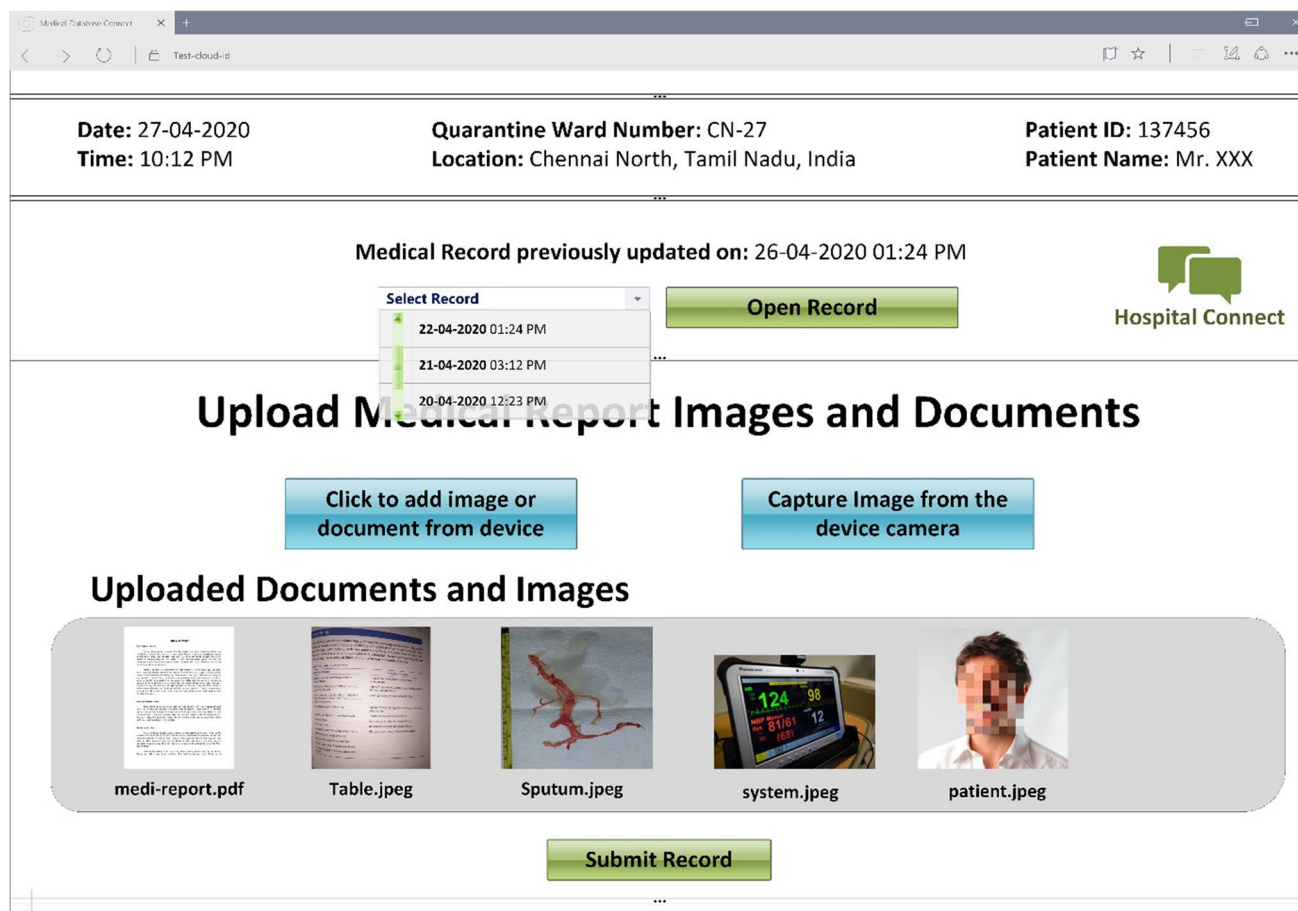


Fig. 7 User interface design of the proposed application on the quarantine ward end to capture and upload clinical parameters and send to a specialty hospital

They can select the patient to be evaluated and can pull out patient's entire medical records from the cloud-based database which can be clearly seen and understood as shown in Fig. 9. In order to monitor the variation of any medical parameter such as readings from serological tests and blood pressure, the physician at the specialty hospital can make use of the smart decision support function and effortlessly get a comparative graph with basic statistical parameters, such as mean, median, high, low, and standard deviation, that shows the variation of the selected parameter with time. This can be an indication of how the medication has been working.

The smart decision support option present within the proposed web-based application enables easy exchange of instructions between the specialty hospital and the mobile quarantine ward. The front-line healthcare worker present in the ward can thus seamlessly communicate with the professional in the hospital who can instruct them with suitable methods and medications. For example, instructions such as a drug or dosage change can be recommended through real-time feedback feature present in the application.

2.2 Implementation and validation approach

2.2.1 Data description

The dataset used in the proposed study consists of 1200 medical case records those were synthetically created by our empanelled clinicians. These medical records were printed in papers and scanned with 300 dots per inch resolution and stored as .bmp images in the 8-bit gray scale format. The dataset contains the aggregated sum of 29,994 medical parameters with an average of 25 parameters per record. The detailed list of parameters considered for the proposed work is depicted in Supplemental Table S1.

2.2.2 Validation

Algorithm validation The next important and logical step would be to evaluate the performance and reliability of the developed system before actual deployment. Validation of the system is done in two stages, once after the medical report image is converted into text and once after the

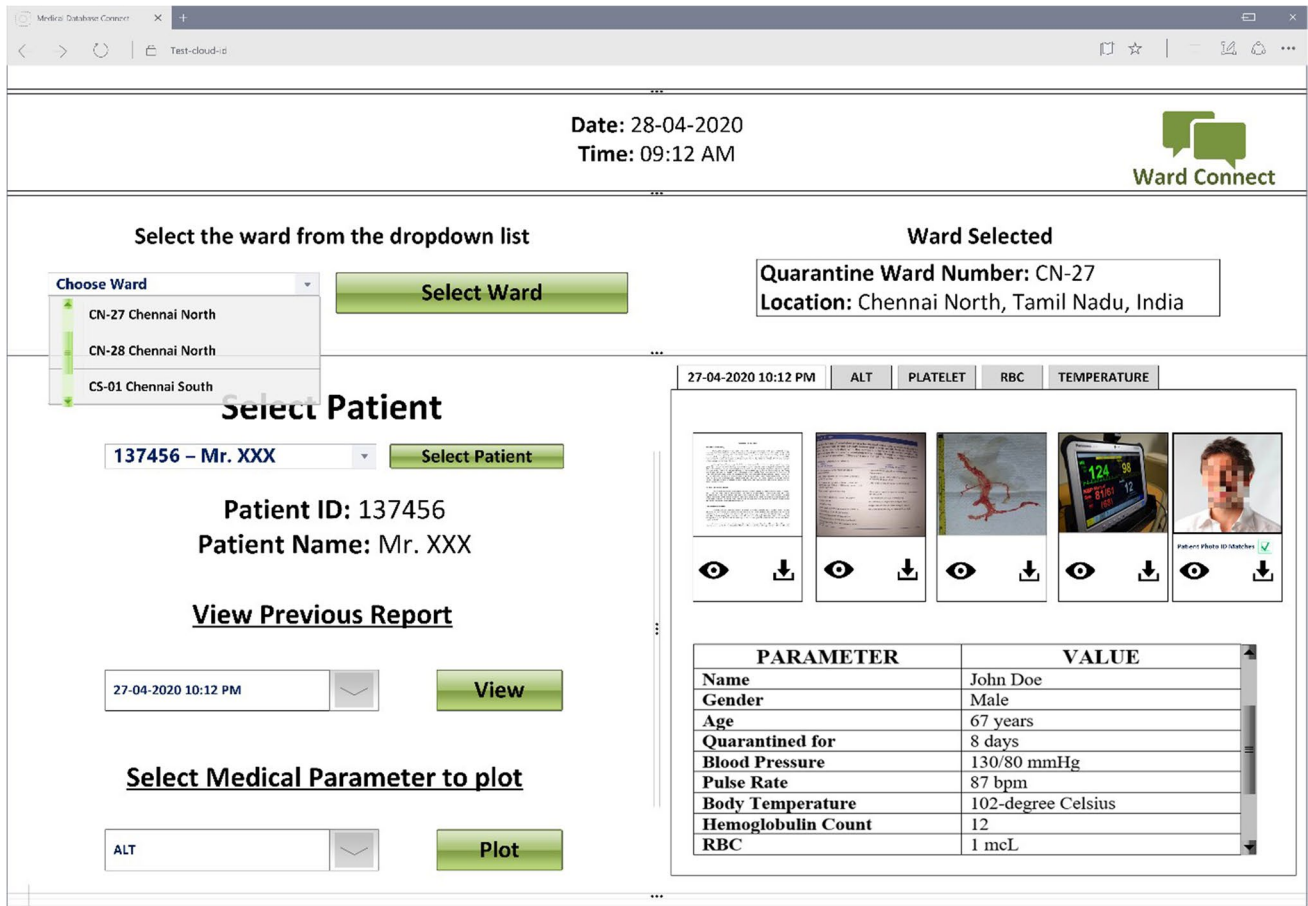


Fig. 8 User interface design of the proposed application on the specialty hospital end where medical data are structured and displayed in tabular form, whereas the images can be either viewed or downloaded

narrative text is structured. This is done to detect the errors at each step in the cloud-based smart structuring tool and to optimize the same. This is done using the formula as given in Eq. 1 and Eq. 2:

$$\eta_{OCR} = \frac{Data_{correct}}{Data_{total}} \tag{1}$$

$$\eta_{system} = \frac{Data_{keywords_correct}}{Data_{total}} \tag{2}$$

where η_{OCR} and η_{system} are the accuracy of the OCR engine and the developed system, respectively, $Data_{total}$ is the total number of input data records, $Data_{correct}$ is the number of input records where the data has been recognized correctly by the OCR engine, and $Data_{keywords_correct}$ is the number of input records whose value for given keywords have been extracted accurately.

The variance of the system is also calculated to measure the reliability, when the system is subjected to repeated

measurements. The variance of the system is calculated using Eq. 3:

$$Variance(\sigma^2) = \frac{\sum_{n=1}^3 (m_n - \bar{m})^2}{n - 1} \tag{3}$$

where m_n is the number of keywords extracted correctly in each record in a given attempt and n is the number of attempts. Here, the number of attempts is 3.

The other performance measures such as precision, recall, and F1 score are calculated based on Eq. 4 to Eq. 6.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{4}$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{5}$$

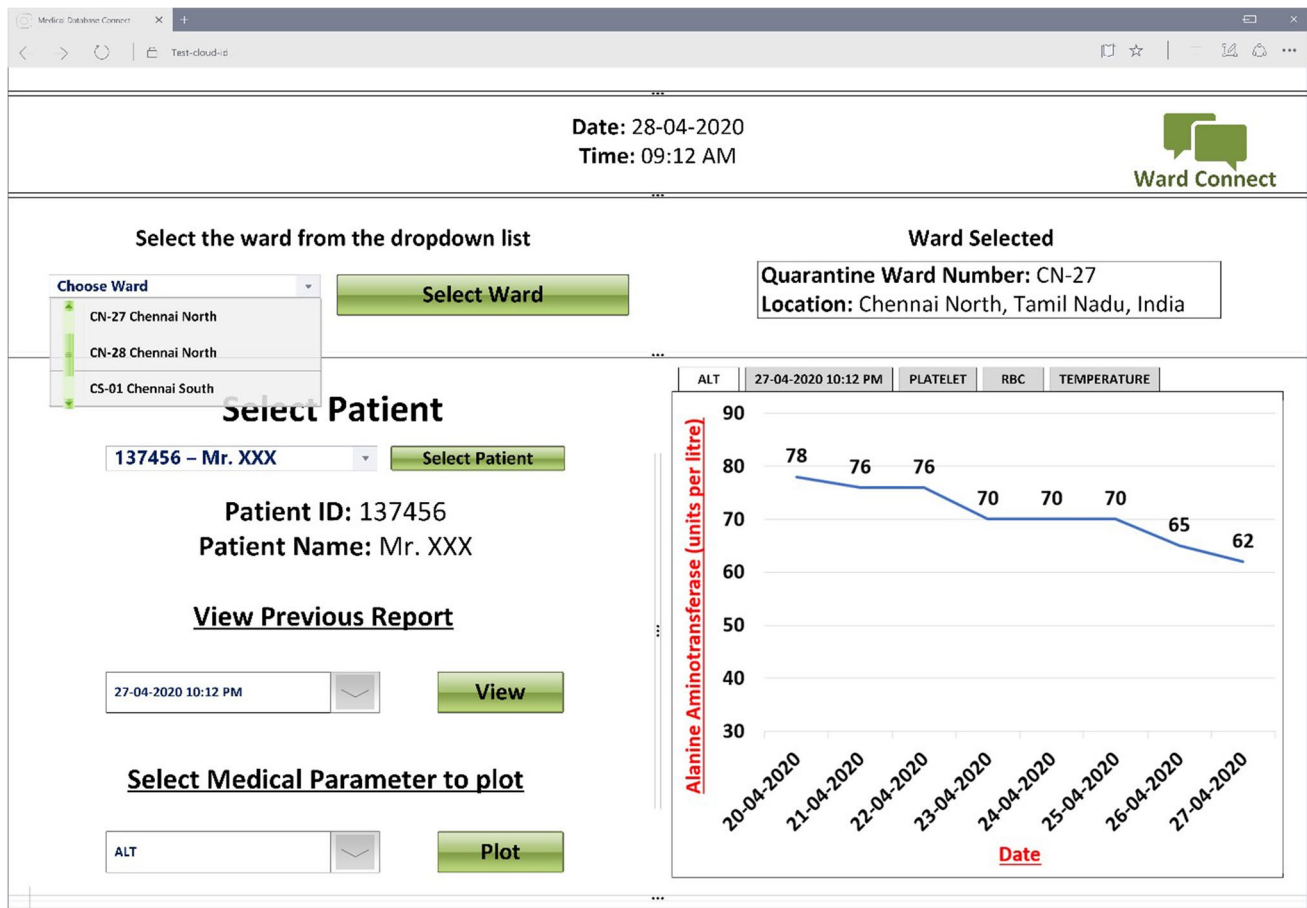


Fig. 9 Smart decision support feature in the application plots key health parameters. The plotting feature enables the doctor to track the variations in a particular medical parameter reading taken at different time intervals

$$F1 \text{ Score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

Clinical validation The developed system is validated clinically by requesting physicians who have experience of working in the COVID-19 hospitals and have treated the infected patients. Four such physicians were identified who then agreed to use this system in an imitation setup with a remote quarantine ward and a hospital-like circumstance. This mock evaluation of the system was based on various performance parameters such as system repeatability, user friendliness, accuracy, latency, usefulness, and adaptability and scalability. The physicians rated the system on a scale of 1 to 5 with 1 being the least and 5 being the highest. The repeatability was measured using Eq. 7 to Eq. 9:

$$\text{Mean}(\bar{m}) = \frac{\sum_{n=1}^5 m_n}{n} \quad (7)$$

$$\text{Standard Deviation} (\sigma) = \frac{\sum_{n=1}^5 (m_n - \bar{m})^2}{n - 1} \quad (8)$$

$$\text{Standard Deviation of the Mean (SDM)} = \frac{\sigma}{\sqrt{n}} \quad (9)$$

where m_n is the number of keywords extracted correctly in a given record in a given attempt and n is the number of attempts. Here, the number of attempts is 5. Standard deviation of the mean (SDM) is then calculated and is taken as the measure of repeatability of the system. Ideally, SDM should be zero. It means that the results have zero variance and that the system possesses high repeatability.

2.2.3 Implementation technology

The proposed web-based application was developed in Windows 10, 64-bit operating system with Intel Core i5 processor, 2.40 GHz, 8 GB RAM. Web pages were created using HTML and integrated with back-end Python code (Python V3.7) through middleware programming. Firebase database was used for storing and retrieving the data. The application was deployed onto the cloud-based server for testing. Microsoft Azure OCR API was used for conversion of images into text. As a part of the smart cloud-based structuring tool, NLTK (Version 3) along with our proposed structuring algorithm was used for identification and structuring of records. Graphs and tables were created using CSS and JavaScript.

Separate web pages with login access were designed for front-line healthcare workers to scan and upload patient’s records from the quarantine ward end and specialist physician to retrieve and analyze at the specialty hospital ends. The decision support features with plot function which is previously mentioned will enable the specialist physician to study the variation of a parameter with time. This will enable them to know the response of the patient to a particular drug or medication. Any abnormal values in a medical parameter will alert the physician. This is done using the isolation forest algorithm for anomaly detection.

3 Results

A basic layout along with the required placeholders was created with the help of the framing option as depicted in Fig. 10. The functions and positions of every element are clearly specified for illustrative purpose. Relative positioning is used rather than global positioning so that the website window remains constant for any screen size. Scroll bars are included for better readability in the case of huge text report images.

In Fig. 11, we can notice the use of the chat option built into the application. Through this feature, effective and easy communication can be established between the specialty hospital and the quarantine ward. On the hospital end, the chat option has smart reply features integrated with it. All these enable effortless transfer of messages between the two entities.

3.1 Algorithm validation

A set of 1200 medical case history reports were created under the guidance of a physician to test the efficiency of the OCR engine. Three trials were conducted, out of which 1195.3 records were correctly recognized on an average as depicted in Table 1. This brings the efficiency of the OCR engine to 99.6%. To test the efficiency of the NLP algorithm;

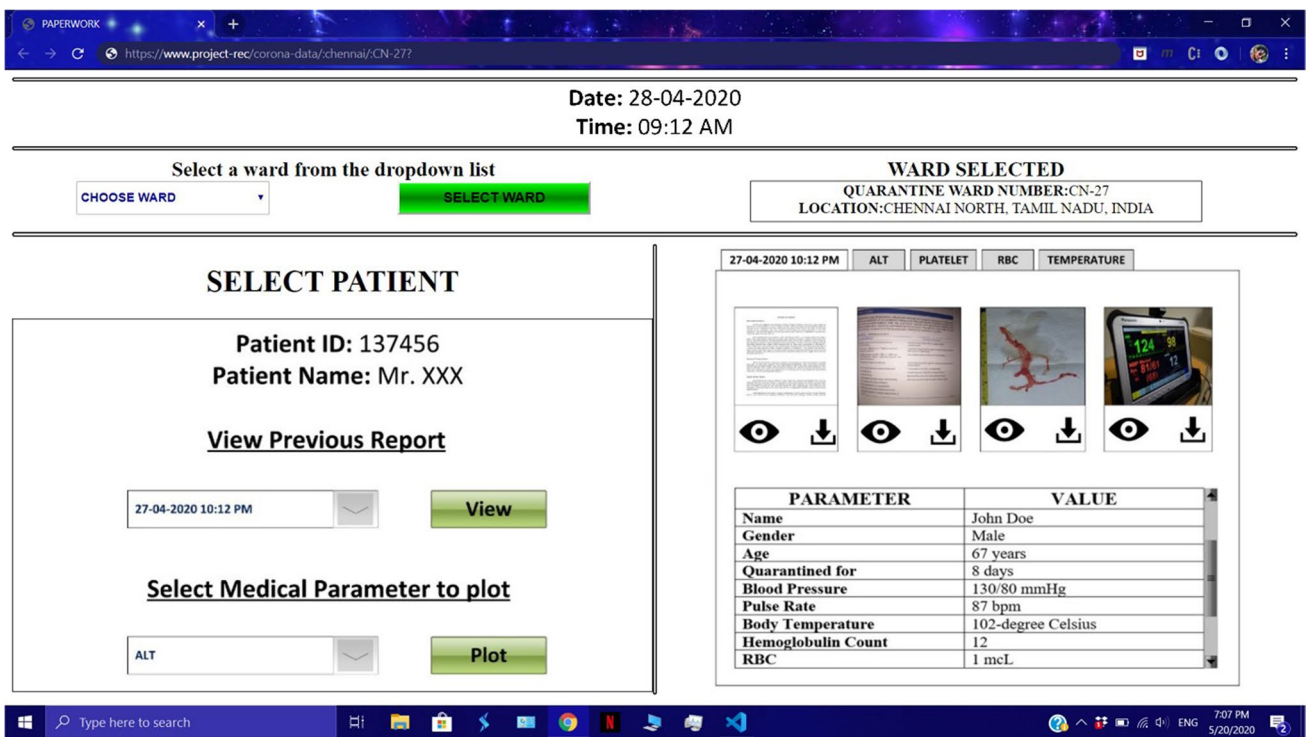
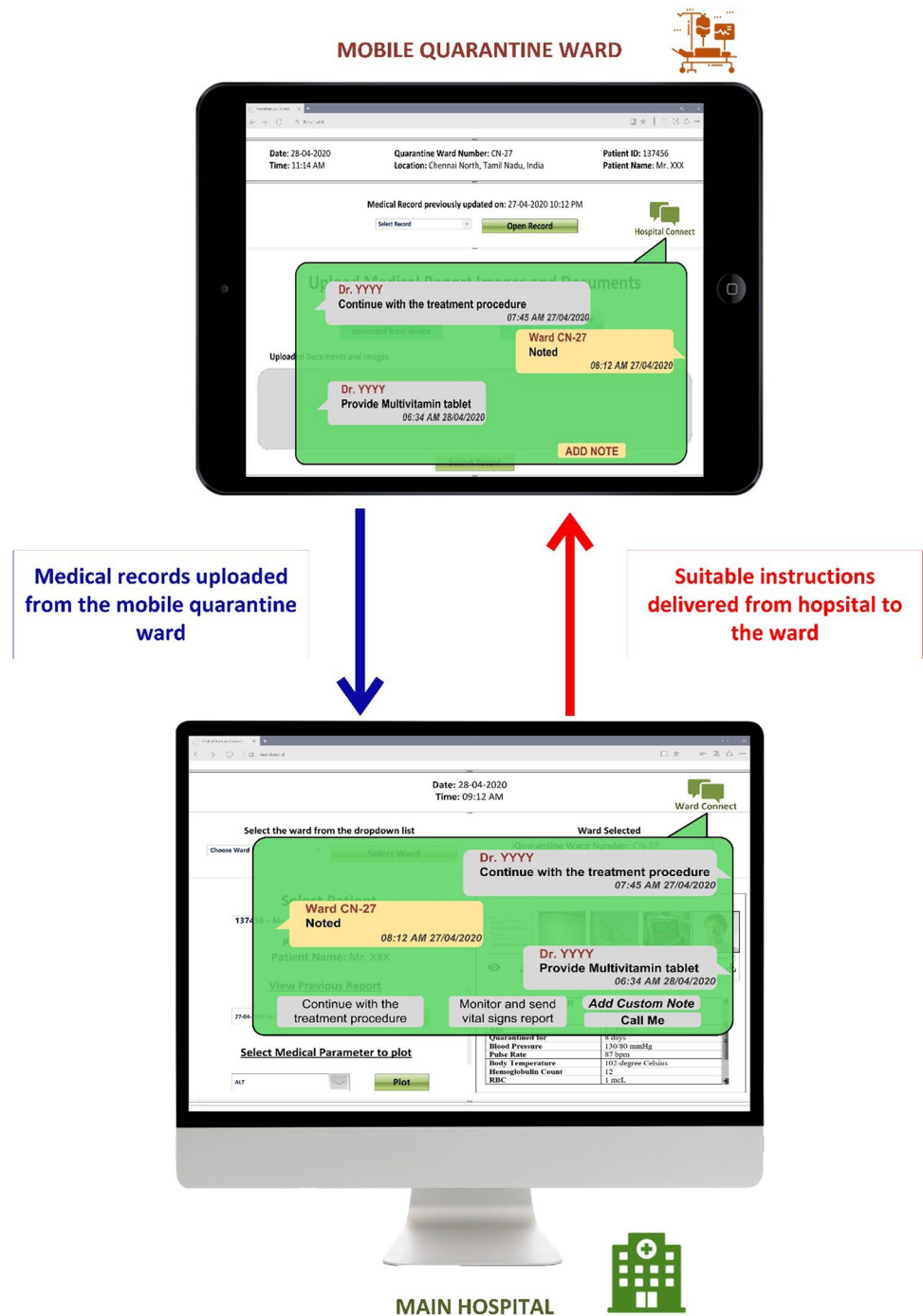


Fig. 10 Visualization of structured data obtained by proposed smart structuring tool on the specialty hospital end

Fig. 11 Messaging applet built within the web application to empower communication between the quarantine ward and the specialty hospital. The applet has an inbuilt smart answer suggestion system enabling faster and easier communication on both the ends



the same 1200 medical case histories were directly fed into the system as text files. The results were then manually compared with the original data. The algorithm was considered to be successful if and only if all the medical data obtained were accurate. Three trials were conducted, and on an average, 1129.3 records were extracted, and keywords were classified correctly. The calculated root means squared coefficient of variation (RMS CV) was 1.3%. It was found that the parameters extracted using the smart structuring algorithm

were highly reproducible and have an efficiency of about 94%. The memory size of the entire developed system was found to be 1.8 megabytes. Hence, this acts as a trade-off between accuracy and size. In addition to that, the number of keywords extracted correctly in each record along with its value was evaluated, and the mean and variance of the system was calculated using previously mentioned Eq. 3. The system was found to have zero variance. Hence, the developed system can be deemed reliable.

Table 1 Efficiency of the proposed smart structuring algorithm: comparison of parameter extraction in multiple attempts

Key parameters/performances		“Ground Truth” parameters	Number of parameters correctly identified by the proposed method		
			Attempt 1	Attempt 2	Attempt 3
Total number of medical case records ($n = 1200$)	The aggregated sum of parameters extracted from 1200 case records	29,994	29,922	29,925	29,923
	Average number of parameters per case record (\pm Std. Dev)	24.995 (± 3.219)	24.935 (± 3.133)	24.938 (± 3.137)	24.936 (± 3.134)
	Minimum and Maximum number of parameters per case record [min–max]	[20–30]			

“Ground Truth” parameters were manually extracted by clinical expert

3.2 Clinical validation

A mock clinical setup was established, and the system was used by four physicians who then filled out a questionnaire regarding the usage of the proposed system. The physicians uploaded various medical records and tested the system by repeating with the same medical record. On a scale of 1 to 5, with 1 being the lowest and 5 being the highest, they gave their reviews. Their views were collated, and a mean score was calculated (shown in Supplemental Table S2). On an average, the physicians rated the user friendliness of the system at 4.8. The accuracy of the algorithm was rated at 4.6. Although there were not any misidentifications of the keyword’s values, in some cases, one or two keywords were not identified. The physicians found the computation speed of the system to be quite fast without any latency. The inbuilt chat option was found to have a latency of 2.7 s. The processing of the uploaded records (character identification and structuring) took about 4.8 s. All the time metrics were measured with the help of an assistant who used a stopwatch to measure the time intervals. In addition to that, all the 4 physicians felt the need for such a system and marked it as a must-have system during pandemic-like situations for patient monitoring. On an average, the physicians rated 4.2 for system adaptability and scalability. Since the developed system addresses only those medical parameters that are relevant to the novel coronavirus disease, the physicians felt the need to scale the system to other domains of medicine and surgery.

The repeatability of the system was calculated for the records that were uploaded to the system by the physicians. A single report was uploaded 5 times to verify if all the attempts produced the same result. It was calculated in terms of SDM as seen in Eq. 9. The SDM produced a value of zero. This means that the system is repeatable and would produce the same result when tried any number of times. Hence, the

developed system was clinically validated by physicians who felt the need to have it in the hospitals treating COVID-19 patients.

4 Discussion

Front-line healthcare professionals and other healthcare workers at the quarantine ward often visit many patients in a short period of time, which may lead to increased levels of human errors. In the current clinical practice, a nursing staff who notes down the medical parameters of the patient as the physician dictates and the values are later fed into a healthcare system/database. By doing so, there is a possibility of human error while noting down the values and while uploading them in the system. The proposed system reduces the proximity of interaction between the healthcare worker and the patients, while increasing the number of patients a single healthcare worker can monitor. Since the developed system increases the speed of communication and in an organized way between these professionals, one front-line worker will be able to monitor several patients. This reduces the number of front-line health providers who will actually work from the mobile quarantine facility while reducing the possibility of error. Perez et al. in their work have developed a similar NFC-enabled system built specifically for patient care in the ICUs in a hospital [25]. It is expected that their system would work well in a smaller set of patients and would be ineffectual in the case of pandemic like situations, since the system requires several front-line professionals (nurse, auxiliary nurse, primary care physician) for data validation. This slows down the speed of monitoring. In the case of our developed solution, since it is coupled to a smart structuring and communication tool, it becomes a reliable application to use during pandemic situations owing to its speed and ease of monitoring.

The NLP along with the novel smart structuring algorithm present as a part of the cloud-based smart structuring tool proved to be useful for extracting much key medical information from the medical texts. In our proposed system, we have taken into consideration about 57 medical parameters that are most noted by patients suffering from COVID-19 including basic information such as name, age, and blood group. Other details that were extracted include diseases, symptoms, and medications. When the physician visits a particular patient the next time, the same set of tests may be done for the patient. These medical parameters are stored every time a test is taken, and a report is drafted. The physician can then know the history of a particular parameter of the patient with just a single click. The physician can understand more about the reaction of the body to the previously prescribed drugs. The graphical representation of key health parameters helps the physician to get deeper insights about the present condition of the patient. Vilic et al. in their work leveraged NLP to create a timeline of vital signs for critically ill patients [26]. Their system works well only in situations where all the data are available readily in the digital format. However, our proposed system takes into pre-existing hard copies of records as well, making it a ready to deploy system in any situations.

The proposed system aims to act an open-ended framework to which several modules can be incorporated in the future. One such module would be the addition of sensors such as vital monitoring sensors which can be added by adjusting and adding the necessary software and hardware architectures. In addition, the proposed system can be extended to several other parameters depending on the need and the field of study and examination. Thus, it makes the proposed system be exhaustive and need-oriented and scalable depending on the application.

4.1 Comparison of keyword extraction algorithms

It is important to compare the various keyword extraction algorithms to justify the selection of the proposed one in this system. For this, parameters such as accuracy, precision, and recall must be studied. Comparing the performance of the proposed novel combination keyword extraction algorithm to other similar approaches is not an easy task, due to the lack of medical patient history and data classification systems that use a general purpose semantic annotator like the one we used. Hence, a comparison with relatively similar systems is made by citing the review works of several researchers. A comparison of graphical methods such as TextRank and statistical methods such as TF*IDF and POS using RAKE algorithm was done by Martin Dostal and Karel Jezek [27]. They have shown that the RAKE algorithm works better than the others when large datasets are involved. Garcia et al. in their work of classifying biomedical literature using bag

of words (BoW) have obtained high accuracy, precision, and recall [28]. They have compared it with other similar systems and have proven better performance. A comparison of text classification systems with each of their pros and cons was done by Aliwy and Ameer [29]. These prove that the proposed novel combinatorial RAKE and BoW algorithm has relatively higher performance, both in terms of speed, accuracy, and simplicity, in comparison with other algorithms and approaches. In the future, an extensive analysis of the performance of different algorithms is to be made for patient clinical data.

4.2 Focusing on “T3” Strategy (“Test”, “Trace”, and “Treat”)

The proposed system will aid in proper implementation of the “T3” strategy, specifically in the treating stage. Tracing of infected individuals can be done using mobile applications such as “Aarogya Setu”, developed by the Government of India. The application creates a user database to form a network of information that can alert citizens and officials of potential victims of the COVID-19 pandemic. It leverages Bluetooth range as a proximity sensor under which a person can likely be infected with coronavirus when come in contact with an infected person. Data is shared between devices when they come in contact to each other, thus enabling the government to be notified for further testing.

Other front-line professionals in quarantine wards can now update the condition of an infected patient while maintaining minimum contact using the NFC tags. The records can be sent to the specialist physicians present in specialty hospitals who can then assess the direction in which the treatment must be taken forward. The records can also be appended to the database so that the progress of a patient can be monitored easily and remotely by the specialist physician. Structuring of medical records is swiftly done by the NLP and novel structuring algorithm. Thus, the proposed system provides a viable solution to the present pandemic situation.

4.3 Use case scenarios

The proposed system can be applied to a wide range of situations. In the case of people in L_2 , where the exposed person is kept in special quarantine wards, the proposed system can be used to monitor the status of the person to check for symptoms of COVID-19. In the current scenario, in states of India such as West Bengal, where monitoring devices such as thermometers are handed over to them for self-monitoring and reporting, the proposed system can be used to visualize and analyze various parameters of the person remotely. Brian and Malcolm in their work reported an NFC integrated system where NFC tags are given to infected people. Their system is particularly useful to monitor if or not the patient

has developed symptoms of COVID-19 [30]. Similarly, the proposed system reduces the need for the healthcare workers to visit the houses of the quarantined people, thereby minimizing contact and exposure.

For the patients who are in $L3$, the proposed system can be used extensively to monitor the progress of the patient and thus the response of the patient to the given medications and diet. It is of utmost importance to adjust the drug dosage and formulation from time to time depending on how the person responds to a given medication. Since the system is scalable, in the future, wearable sensors can be attached to the patients to enable continuous and real-time monitoring of vital signs. Specifically, in the case of patients who are on life support systems, healthcare professionals can be alerted to abnormal variations in patient vital signs and receiving instant updates that enable them to quickly identify and prevent potentially harmful situations. A short review of the various developed remote patient monitoring and clinical decision support systems is summarized in Table 2.

As depicted in Table 2, several systems exist for monitoring patients and their progress. Like the work of Andreea et al. [31], the physicians found the system very useful and easy to use owing to its simple and minimalistic user interface and experience. Unlike the system developed by Lai et al. [32], which focuses only on fatigue and related symptoms, the developed system here focuses on a wide range of symptoms primarily noticed in COVID-19 patients. In addition to the work of Al-Naggar et al. [34], the proposed system has very low latency. Like the work of Novo et al. [36], the system can be scaled to any domain of medicine and surgery, but without any much change to the framework.

Data extraction through NLP is a relatively new technology in the healthcare domain. Anil et al. [37] in their work have leveraged the Kaiser Permanente Southern California (KPSC) clinical information extraction system. This system uses computationally hungry algorithms such as the Word2Vec and similar transformer models. They have demonstrated the efficiency of about 99%, while extracting data from electronic medical records of patients with prostate-related diseases, specifically prostate cancer. Similarly, Thomas et al., [38] in their work with the use of NLP to annotate drug product labeling, have achieved an accuracy ranging from 64 to 77%. Again, the technology used here is memory-wise heavy, leading to lower computational speeds. A detailed comparison of performance metrics of the proposed system against various similar systems is highlighted in Table 3.

In future, the proposed system can be integrated with existing electronic health records (EHRs) or vendor neutral archives (VNA) for seamless data storage and retrieval. In the current scenario, the system stands alone and not tested in the actual clinical environment. This was developed solely to address the difficulties in monitoring the

patients in the COVID-19 pandemic situation. In addition to this, it should also be Health Insurance Portability and Accountability Act (HIPAA) compliant and ensure privacy, security, and portability of health information. In addition, the system is built keeping in mind the regulatory compliance to avoid any potential breach of sanitary rules in critical care units. Usage of Hadoop and Oracle has ensured the same. Also, the system must be scaled to other domains of medicine and surgery. Currently, the proposed smart structuring algorithm addresses and extracts values only for those medical parameters that are related to the novel coronavirus disease. In future, it can be extrapolated according to the need.

5 Conclusion

In the present scenario, there is a need to create a centralized dashboard that can help various health and government officials to visualize the spread of the virus and progress of the patients for a given treatment. By this, outliers can be easily detected, and the medications can be altered according to the patient. It must also be ensured that the healthcare workers maintain minimum contact and exposure to the virus. The proposed web-based system that deploys NFC tags proves to be a viable solution to this problem. In this way, maximum utilization of the workforce can be enabled with minimum exposure to the virus, thus reducing its spread. Ease of visualization of the medical reports is of utmost importance since it enhances the decision-making capabilities of the physician. This is achieved using the smart cloud-based structuring tool that uses OCR and NLP to convert the input into structured digital data. Thus, several capabilities of the proposed web-based system are found to have an efficiency of over 94%. The system was tested in an authentic mock clinical setup and evaluated by physicians who felt it user friendly and reliable. The system showed a low latency of 2.7 s. In addition to that, since it has an open-ended architecture, the system can be scaled depending on the future needs. Also, as a future development of this system, real-time sensors can be deployed to monitor the patients continuously, thus alerting the physician instantly in case of abnormal and abrupt variations of vital signs in the case of people on life-support systems. Since the range of NFC transfers is less than a few centimeters, data breach in air is not possible, thus making the proposed system more secure.

Since the study focuses more on the activities employed by the various governing bodies in India, at present state, the proposed system is more suitable for the current Indian situation, and in the future, the system can be modified to suit the needs of the other countries depending upon their requirements.

Table 2 Review of various pre-existing patient monitoring and clinical decision support systems

Authors	Algorithm/tech used	Use case	Results obtained
Andreea et al. [31]	Arduino Uno and ECG sensor	Pregnant women	Pregnant women monitoring using mobile app
Lai et al. [32]	Patient fatigue level monitoring using self-reporting, interactive app	Juvenile cancer patients	<i>Percentage of usability: 95% of patients, 71% of clinicians</i>
Reiss et al. [33]	Left ventricular assist device, wireless controller, cloud analytics, physician workstation	Patients with heart failure	<i>Percentage of usability: 76.8% of patients</i>
Al-Naggar et al. [34]	Arduino Mega for collection of data from several sensors through Wi-Fi	People living in remote locations	<i>Accuracy: 99.2%; latency: 14 s in Wi-Fi and 68 s for data transfer</i>
Sathya and Kumar [35]	Blowfish algorithm for encrypting medical data and Ciphertext policy attribute-based encryption for authorization	Patients using internet connected biosensors	Fast and secure data transmission
Novo et al. [36]	Cardiovascular parameters tracking system using sensors	Patients in early stages of and those with cardiovascular disorders	Tracking of all the parameters related to cardiovascular disorders made it possible to detect them in advance
Proposed system	(i) An NFC chip to connect with the patient's medical records. (ii) A novel cloud-based structuring algorithm with inbuilt OCR Engine	Monitoring of patients in remote quarantine wards during pandemic situations	User rated metrics (out of 5) Accuracy: 4.6, user friendliness: 4.8, importance of system: 4/4 physicians agreed, system adaptability: 4.2, latency: 2.7 s for message transmission, 4.8 s for processing of records Algorithm metrics Accuracy: 94%, system size: 1.8 megabytes

Table 3 Comparison of evaluation and performance metrics of similar NLP algorithms that use patient-authored textual data

Author	Data source	Data type	Number of documents	Evaluation and performance metrics
Freifeld et al. [39]	Twitter	Tweets with mentions of 23 drugs and 4 vaccines and resemblance to adverse events	4401 tweets	Automated, dictionary-based symptom classification had 72% recall and 86% precision
Gupta et al. [40]	Online community forum for asthma, ENT, adult type II diabetes, acne, and breast cancer	Sentences	680,071 sentences	Extracts symptoms and conditions with an F-measure of 66–76%
Jimeno-Yepes et al. [41]	Twitter	Tweets with 2 out of 3 entity types—diseases, symptoms, or pharmacological substances	1300 tweets	Highest performing model (Micromed + Meta) had precision, recall, and F-measure as 72%, 60%, and 66%, respectively
Karmen et al. [42]	Online public mental health message board	User posts within a 20–200 word interval	1304 posts	Average precision of 84% and an average F-measure of 79%
Liu & Chen [43]	3 online patient forums for diabetes and 1 for heart disease	Patient discussion posts	1,072,474 posts	Average F-measure of 90% for drug entity extraction and average F-measure of 80% for medical event extraction
Nikfarjam et al. [44]	Twitter and a health-related social network	Tweets and user posts for 81 widely used drugs	1784 tweets and 6279 posts	86%, 78%, and 82% for precision, recall, and F-measure, respectively
Eshleman & Singh [45]	Twitter	Tweets with mention of 200 commonly prescribed drugs	157,735 tweets	Precision exceeding 85% and F-measure over 81%
Kilicoglu et al. [46]	PubMed	Abstracts of manuscripts	-	0.55 precision, 0.34 recall, and 0.42 F1 score
Proposed System	Patient medical records	Medical data extracted from scanned medical records	1200 records contain aggregated sum of 29,994 parameters	Precision: 94% Recall: 91% F1 score: 92%

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11517-021-02456-1>.

Acknowledgements We would like to acknowledge Dr. Harish Balasubramanian and his team from Calcutta Medical College and Hospital, for his inputs regarding the standard operating procedures followed by the hospital in treating COVID-19 patients and for his inputs on medications followed.

Declarations

Conflict of interest All authors declare no competing interests.

References

- World Health Organisation, Coronavirus disease (COVID-19) dashboard. <https://covid19.who.int/> Accessed 15 June 2021
- Centres for Disease Control and Prevention, Coronavirus Disease (COVID-19) (2020) How COVID-19 spreads. <https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/how-COVID-19-spreads.html>. Accessed 15 June 2021
- World Health Organisation, Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). [https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-\(2005\)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-\(2019-ncov\)](https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov)). Accessed 15 June 2021
- World Health Organisation (WHO) Director-General's opening remarks at the media briefing on COVID-19, <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-COVID-19---11-march-2020>. Accessed 15 June 2021
- World Health Organisation, T3: Test. Treat. Track (2012) Scaling up diagnostic testing, treatment, and surveillance for malaria, https://www.who.int/malaria/publications/atoz/test_treat_track_brochure.pdf?ua=1, April 2012 (accessed 15 June 2021)
- Government of India, Ministry of Electronics and Information Technology (2020) Notification on AarogyaSetu data access, and knowledge sharing protocol. https://static.mygov.in/rest/s3fs-public/mygov_159051652451307401.pdf. Accessed 15 June 2021
- Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D (2020) Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* 181:281–292. <https://doi.org/10.1016/j.cell.2020.02.058>

8. Elmashae Y, Yermakov M, Frank E, Benjamin M, Maier A, Newman N, Reponen T, Grinshpun SA (2019) Exposure of home-attending healthcare workers to aerosolized medications (simulation study). *J Aerosol Sci* 133(2019):49–55. <https://doi.org/10.1016/j.jaerosci.2019.04.006>
9. Matuka O, Singh TS, Bryce E, Yassi A, Kgasha O, Zungu M, Kyaw K, Malotle M, Renton K, O'Hara L (2015) Pilot study to detect airborne Mycobacterium tuberculosis exposure in a South African public healthcare facility outpatient clinic. *J Hosp Infect* 89:192–196. <https://doi.org/10.1016/j.jhin.2014.11.013>
10. Alfaraj SH, Al-Tawfiq JA, Altuwaijri TA, Alanazi M, Alzahrani N, Memish ZA (2018) Middle East respiratory syndrome coronavirus transmission among health care workers: implication for infection control. *Am J Infect Control* 46:165–168. <https://doi.org/10.1016/j.ajic.2017.08.010>
11. Levin J (2019) Mental health care for survivors and healthcare workers in the aftermath of an outbreak. *Psychiatry Pandemics* 127–141. https://doi.org/10.1007/978-3-030-15346-5_11
12. John Hopkins University and Medicine (2020) Maps and trends in coronavirus disease 2019. <https://coronavirus.jhu.edu/data>. Accessed 15 June 2021
13. Centers for Disease Control and Prevention (2020) Coronavirus disease (COVID-19), cases in the U.S. <https://covid.cdc.gov/covid-data-tracker/#health-care-personnel>. Accessed 15 June 2021
14. Government of Indian, Ministry of Railways, Railways Board (2020) COVID-19 - Conversion of coaches to serve as quarantine/isolation facilities. http://www.indianrailways.gov.in/railwayboard/uploads/directorate/mec_engg/downloads/Coaching/2020/Conversion_ICF_Coaches_310320.pdf. Accessed 15 June 2021
15. Evans RS (2016) Electronic health records: then, now, and in the future. *Yearb Med Inform Suppl*.1:S48-S61. <https://doi.org/10.15265/IYS-2016-s006>
16. Gune A, Bhat A, Pradeep A (2013) Implementation of near field communication-based healthcare management system. *IEEE Symp Ind Electron Appl Kuching* 195–199. <https://doi.org/10.1109/ISIEA.2013.6738993>
17. Sujadevi VG, Kumar TH, Arunjith AS, Hrudy P, Prabakaran P (2016) Effortless exchange of personal health records using near field communication. *Int Conf Adv Comput Commun Inf Jaipur* 1764–1769. <https://doi.org/10.1109/ICACCI.2016.7732303>
18. Supriya S, Ramgopal S, George SM (2017) Near field communication-based system for health monitoring. 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology Bangalore 653–657. <https://doi.org/10.1109/RTEICT.2017.8256678>
19. National Patient Safety Agency (2005) Wristbands for hospital inpatients improves safety, Safer practice notice 11. <https://www.england.nhs.uk/patient-safety/national-patient-safety-incident-reports/>. Accessed 15 June 2021
20. Mannos D (2003) NCPS patient misidentification study: a summary of root cause analyses. VA NCPS Topics in Patient Safety. Washington, DC, United States Department of Veterans Affairs. https://www.patientsafety.va.gov/docs/TIPS/TIPS_Jul03.pdf. Accessed 15 June 2021
21. Thompson P, McNaught J, Ananiadou S, (2015) Customised OCR correction for historical medical text. (2015) *Digital Heritage*. Granada, Spain 35–42. <https://doi.org/10.1109/DigitalHeritage.2015.7413829>
22. NXP, NTAG 210/NTAG 212, Overview. https://www.nxp.com/products/rfid-nfc/nfc-hf/ntag/ntag-for-tags-labels/ntag210-ntag212-ntag-21x-nfc-forum-type-2-tag-ic-with-multiple-user-memory-sizes:NTAG210_NTAG212. Accessed 15 June 2021
23. Microsoft Azure, Computer Vision API documentation. <https://docs.microsoft.com/en-us/azure/opbuildpdf/cognitive-services/computer-vision/toc.pdf?branch=live>. Accessed 15 June 2021
24. Microsoft Azure, Face detection API. <https://azure.microsoft.com/en-in/services/cognitive-services/face/>. Accessed 15 June 2021
25. Martínez MP, Dafonte C, Gómez A (2018) Traceability in Patient Healthcare through the Integration of RFID Technology in an ICU in a Hospital. *Sensors* 18:1627–1644. <https://doi.org/10.3390/s18051627>
26. Vilic A, Hoppe K, Petersen J, Kjaer T, Sorensen H (2016) Simplifying EHR overview of critically ill patients through vital signs monitoring. *IEEE J Biomed Health Inform* 1:1. <https://doi.org/10.1109/JBHI.2016.2638123>
27. Dostal M, Ježek K (2011) Automatic keyphrase extraction based on NLP and statistical method. *Proceedings of the DATESO 2011: Annual International Workshop on Databases, TExts, Specifications and Objects*, Pisek, Czech Republic
28. García MA, Rodríguez RP, Rifón LE (2015) Biomedical literature classification using encyclopedic knowledge: a Wikipedia-based bag-of-concepts approach. *PeerJ* 3:e1279. <https://doi.org/10.7717/peerj.1279>
29. Aliwy AH, Ameer EA (2017) Comparative study of five text classification algorithms with their improvements. *Int J Appl Eng Res* 12:4309–4319
30. Ondiege B, Clarke M (2017) Investigating user identification in remote patient monitoring devices. *Bioengineering* 4:76–101. <https://doi.org/10.3390/bioengineering4030076>
31. Robu A, Gauca B, Crisan-Vida M, Stoicu-Tivadar L (2016) Integrated system for monitoring and prevention in obstetrics-gynaecology. *Stud Health Technol Inform* 221:8–12. <https://doi.org/10.3233/978-1-61499-633-0-8>
32. Lai JS, Yount S, Beaumont JL, Cella D, Toia J, Goldman S (2015) A patient-centered symptom monitoring and reporting system for children and young adults with cancer (SyMon-SAYS). *Pediatr Blood Cancer* 62:1813–1818. <https://doi.org/10.1002/pbc.25550>
33. Reiss N, Schmidt T, Müller-von FA et al (2017) Telemonitoring and medical care of heart failure patients supported by left ventricular assist devices—the medolution project. *Stud Health Technol Inform* 236:267–274. <https://doi.org/10.3233/978-1-61499-759-7-267>
34. Al-Naggar NQ et al (2019) Design of a remote real-time monitoring system for multiple physiological parameters based on smartphone. *J Healthc Eng* 2019:5674673. <https://doi.org/10.1155/2019/5674673>
35. Sathya D, Kumar PG (2017) Secured remote health monitoring system. *Healthc Technol Lett* 4:228–232. <https://doi.org/10.1049/hlt.2017.0033>
36. Novo J, Hermida A, Ortega M et al (2017) Hydra: a web-based system for cardiovascular analysis, diagnosis, and treatment. *Comput Methods Programs Biomed* 139:61–81. <https://doi.org/10.1016/j.cmpb.2016.10.019>
37. Anil Thomas A, Zheng C, Jung H et al (2014) Extracting data from electronic medical records: validation of a natural language processing program to assess prostate biopsy results. *World J Urol* 32:99–103. <https://doi.org/10.1007/s00345-013-1040-4>
38. Thomas L, Carol P, Oanh D et al (2018) Evaluation of natural language processing (NLP) systems to annotate drug product labelling with MedDRA terminology. *J Biomed Inform* 83:73–86. <https://doi.org/10.1016/j.jbi.2018.05.019>
39. Freifeld CC, Brownstein JS, Menone CM et al (2014) Digital drug safety surveillance: monitoring pharmaceutical products in twitter. *Drug Saf* 37:343–350. <https://doi.org/10.1007/s40264-014-0155-x>
40. Gupta S, MacLean DL, Heer J, Manning CD (2014) Induced lexico-syntactic patterns improve information extraction from online medical forums. *J Am Med Inform Assoc* 21:902–909. <https://doi.org/10.1136/amiajnl-2014-002669>
41. Jimeno-Yepes A, MacKinlay A, Han B, Chen Q (2015) Identifying diseases, drugs, and symptoms in twitter. *Stud Health*

- Technol Inform 216:643–647. <https://doi.org/10.3233/978-1-61499-564-7-643>
42. Karmen C, Hsiung RC, Wetter T (2015) Screening Internet forum participants for depression symptoms by assembling and enhancing multiple NLP methods. *Comput Methods Programs Biomed* 120:27–36. <https://doi.org/10.1016/j.cmpb.2015.03.008>
 43. Liu X, Chen H (2015) A research framework for pharmacovigilance in health social media: Identification and evaluation of patient adverse drug event reports. *J Biomed Inform* 58:268–279. <https://doi.org/10.1016/j.jbi.2015.10.011>
 44. Nikfarjam A, Sarker A, O'Connor K, Ginn R, Gonzalez G (2015) Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. *J Am Med Inform Assoc* 22:671–681. <https://doi.org/10.1093/jamia/ocu041>
 45. Eshleman R, Singh R (2016) Leveraging graph topology and semantic context for pharmacovigilance through twitter-streams. *BMC Bioinformatics* 17:335. <https://doi.org/10.1186/s12859-016-1220-5>
 46. Kilicoglu H, Rosemblat G, Fiszman M, Shin D (2020) Broad-coverage biomedical relation extraction with SemRep. *BMC Bioinformatics* 21:188. <https://doi.org/10.1186/s12859-020-3517-7>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Vishal Balasubramanian is an Electronics Engineer from Anna University, currently pursuing his graduate degree at the University of British Columbia. He is specializing in digital design, FPGA, and signal processing.

Dr. Sapthagirivasan Vivekanandhan is a Medical Imaging consultant, completed doctorate and post-doctoral in radiology and nuclear imaging division, respectively. He specialized in AI, AR-VR, "5G" in medical imaging.

Prof Venkatesh Mahadevan is the Vice Principal and Dean (Academic Affairs) at QASM with over 30 years of multinational higher education industry experience in India, Malaysia, the UK, and New Zealand.