DTU Library

ReCognizing SUspect and PredictiNg ThE SpRead of Contagion Based on Mobile Phone LoCation DaTa (COUNTERACT): A system of identifying COVID-19 infectious and hazardous sites, detecting disease outbreaks based on the internet of things, edge computing, and artificial intelligence

Ghayvat, Hemant; Awais, Muhammad; Gope, Prosanta; Pandya, Sharnil; Majumdar, Shubhankar

ReCognizing SUspect and PredictiNg ThE SpRead of Contagion Based on Mobile Phone LoCation DaTa (COUNTERACT): A system of identifying COVID-19 infectious and hazardous sites, detecting disease outbreaks based on the internet of things, edge computing, and artificial intelligence

Hemant Ghayvat [a],[*], Muhammad Awais [b], Prosanta Gope [c], Sharnil Pandya [d], Shubhankar Majumdar [e]

[a] *Innovation Department, Technology University of Denmark, Denmark*
[b] *Center for Intelligent Medical Electronics, Department of Electronic Engineering, School of Information Science and Technology, Fudan University, Shanghai, 200433, China*
[c] *University of Sheffield, United Kingdom*
[d] *Symbiosis Centre for Applied Artificial Intelligence and Symbiosis Institute of Technology, Symbiosis International (Deemed) University, Pune, Maharashtra, India*
[e] *Department of Electronics and Communication Engineering, National Institute of Technology Meghalaya, Shillong, 793003, India*

A B S T R A C T

Human movement is a significant factor in extensive spatial-transmission models of contagious viruses. The proposed COUNTERACT system recognizes infectious sites by retrieving location data from a mobile phone device linked with a particular infected subject. The proposed approach is computing an incubation phase for the subject's infection, backpropagation through the subjects' location data to investigate a location where the subject has been during the incubation period. Classifying to each such site as a contagious site, informing exposed suspects who have been to the contagious location, and seeking near real-time or real-time feedback from suspects to affirm, discard, or improve the recognition of the infectious site. This technique is based on the contraption to gather confirmed infected subject and possibly carrier suspect area location, correlating location for the incubation days. Security and privacy are a specific thing in the present research, and the system is used only through authentication and authorization. The proposed approach is for healthcare officials primarily. It is different from other existing systems where all the subjects have to install the application. The cell phone associated with the global positioning system (GPS) location data is collected from the COVID-19 subjects.

1. Introduction

Influenza, more commonly known as "flu', spreads around the world in seasonal epidemics, resulting in about three to five million yearly cases of severe illness and about 250,000–500,000 annual deaths, rising to millions in some pandemic years. For contagious entities to live and replication the succession of contagion in a new host, the virus should leave (not all virus cells) the current confirmed subject and root contagion somewhere else. Diffusion of contagions can proceed via numerous possible ways (Ahmed et al., 2020; Magklaras & Bojorquez, 2020; Urbaczewski & Lee, 2020). Contagious viruses may be transferred either by direct or indirect connection. Direct connection happens when someone is exposed to an infectious source, such as a handshake, kiss,

physical relation, breathing of transmittable viruses particles released by sneezing or coughing. Indirect connection happens when the virus is competent to resist and survive in the oppressive atmosphere outside the host body area for an extended duration and keep on infectious once definite occasion rises. Nonliving substances that are habitually infected comprise door handles, table chair, keyboard mouse, etc. Consumption of infected food or water item is also an indirect way of contagion. In 1918 Spanish flu, then the 2003 SARS pandemic, and now COVID-19, these viruses are airborne and can, therefore, promptly transmittable infect large groups of people (Ahmed et al., 2020).

Thus there is a need for a system that can detect, identify, and track outbreaks in real-time or near real-time to help contain the spread of disease. In some research, a method is used to collect user locations via a

user's mobile device and correlated with sites in which a disease was identified. The system first receives a report of a user infected with the confirmed disease, after which the user's close contacts during the incubation period of the disease are alerted. However, the method suffers from several drawbacks. First, a user is only alerted to potential exposure after the fact that once the user had already been likely exposed to the disease. Second, the system relies on receiving confirmation of illness. At that point, the close contacts during the incubation period are backtracked and alerted for possible exposure. Third, the close contacts need to be in close physical proximity to be alerted to potential exposure. Even then, the system does not determine the likelihood of the infection being transmitted. Fourth, the system does not consider the user's personal information, such as age, which can affect the incubation period. Fifth, the method is only useful for tracking diseases that spread from person to person. Diseases that a person does not transmit (e.g., food poisoning) are not detected even though other users may catch it at the source. Sixth, the method does not detect unknown diseases. Finally, the system does not track or alert users to other location-based hazards (Cencetti et al., 2020; Maccari & Cagno, 2020).

Current solutions are not proficient in producing personalized forecasts for confirmed subjects and suspects. Such as, previously designed research approaches fail to forecast if a distinctive suspect in close contact with the confirmed subject will develop an infection but only offer forecasts on the expose of an infection through a broad area. Operations of existing models do not protect users' privacy by using subject interfaces with multimedia data, browsing, text, and calls (Ahmed et al., 2020; Cencetti et al., 2020; Maccari & Cagno, 2020). For example, these kinds of applications utilize AI (artificial intelligence) practices and data collection procedures that authorize a computer to produce a personalized data stream, which breaks the privacy boundary. Such as, applications gather browsing outcome information created on interpretations in search logs that designate website focuses comprised in the search. In recent times, in countries such as India, a mobile application called "Aarogya Setu" was launched to help the citizens identify the COVID19 infected individuals via Bluetooth. The "Aarogya Setu" mobile application is connected with the repository of COVID19 patients governed by government officials (Gupta, Bedi, Goyal, Wadhera, & Verma, 2020). However, the Aarogya Setu mobile app does not provide real-time tracking of the COVID19 suspected individuals based on GPS Spatio-temporal data. Klar R. and Lanzerath D. discussed the challenges and vulnerabilities of COVID19 tracking mobile applications (Klar & Lanzerath, 2020). However, they did not discuss anything related to real-time tracking of COVID19 suspects using GPS Spatio-temporal data. Martin and his fellow researchers' discussed the security aspects of various smart mobile devices (Martin, Karopoulos, Hernández-Ramos, Kambourakis, & Fovino, 2020). However, they did not discuss anything related to tracing and tracing COVID19 suspected individuals using GPS location data.

Hanson and his team discussed and analyzed various mobile health applications and their assistance for COVID-19 management. However, the conducted research work did not provide any details about the real-time tracking of COVID19 suspects via a mobile application (John Leon Singh, Couch, & Yap, 2020). Fahey R. and Hino A. discussed various aspects of COVID19 privacy, social limits based on public health responses. However, the presented research work did not provide any details related to GPS based tracking privacy issues and real-time detection of COVID19 suspects using location data (Fahey & Hino, 2020). Mallik and his team discussed a GPS tracking application to track the ambulances carrying COVID19 health patients. The presented research work discussed a handy application for pandemic situations. However, they did not facilitate real-time tracking of COVID19 suspects using their location data (Mallik, Sing, & Bandyopadhyay, 2020). Franch-Pardo and his fellow research team discussed critical details related to GPS Spatio-temporal data, its usage, and applications. However, they did not propose any system which can facilitate real-time tracking of COID19 individuals based on their GPS location data (van

Franch-Pardo, Napoletano, Rosete-Verges, & Billa, 2020). Skoll and his fellow researchers discuss the ideas of COVID-19 testing and surveillance, digital contact tracing, and mass-testing of COVID19 individuals in the USA. However, they did not propose a real-time system to detect COVID19 suspects using GPS Spatio-temporal data (Skoll, Miller, & Saxon, 2020). Menni and his team discussed the real-time tracking of COVID-19 symptoms reporting and prediction related concepts. However, they did not discuss anything about GPS based tracking of COVID19 individuals using their location data (Menni, Valdes, & Freidin, 2020).

The majority of GPS network's mobility data represents (Bibri, 2018; Jamal & Habib, 2020; Kontokosta & Hong, 2021; Lu, An, Hsu, & Zhu, 2019; Silva et al., 2021) a similar pattern, such as the smart societies. Smart societies are categorized as mutually connected and mutually dependent through various network arrangements. The interconnecting physical layer of these networks extends orders of magnitude greater than the overlaying network's growth rate. This generates the bulk of data with surprising rates compared to the past. This produces "Mobility IoT," which is closely associated with the Big data setting(typically storage, data pre-processing, and inference) over massive GPS datasets that impose stringent computational resource demands. The above Bigdata issues will require a radical and modern approach to address the emerging problems and keeping up with the expected flood of received Bigdata.

The proposed COUNTERACT framework offers a methodology to address the big data research challenges and provide advanced analytics holistically. The COUNTERACT uses the HYPERBOLIC (Thai et al., 2016) framework for big data analytics. This concept offers a generic computational substrate for data pre-processing, representation, dimensionality reduction, data correlation-clustering, inference, analytics, visualization search-navigation, and decision making in near real-time. The (Bello-Orgaz, Jung, & Camacho, 2016) Big data analytics concept constructs basic pre-processing statistical learning operations and introduces analytics and interpretation for the optimum inference to recognize the infection. The received GPS data may be in raw or networked form. The networked data is correlated data(nodes), and the combinations of complex networks show their correlation paths. To resolve the above issues, we proposed COUNTERACT.

2. Proposed work

The technique and device defined is a way to deliver a switch on epidemics of numerous infections, contagions, or well-being intimidations of multiple types. The method obtains location data from subject and subject systems. Subject systems may apply a GPS data or a cellular network to find location data. The system then develops a correlation server to measure the vicinity of numerous cellphones, subject systems, or supplementary devices, and any other activity of attention. To investigate, in a module, the system analyzes location information linked with local maps and actions. For a location-enabled consumer, the system defines a list of the adjacent encounters with other location-enabled consumers (suspected infections). This location data is reserved for a certain period, and information is clipped for an extended time. In one of the system modules, the period is proportional to the incubation spells of several infections, bacteria, etc. The system collects location information that a specific or 'local origin of contagion' has been recognized either from the user, from a healthcare service professional, or a government body. The COUNTERACT method backpedals all connections for the confirmed subject until the last possible suspect.

The COUNTERACT signals everyone who has been exposed to the contagious site or infected person. Expose is described in a different way for numerous pandemics. For somewhat that comprises airborne microorganisms, in the same four-wheeler automobile, shared workplace, passenger plan, ship, cineplex, bars-eatery, public gathering events where people spaced densely, etc. For microorganisms that spread via physical contacts, such as office meetings, they are used as the signals for

notifying suspects in the housed party. The proposed system discovers potential contagion causes by backpedaling and observing through the branches for each suspect in the close connection cosmos of the confirmed subject. The suspect and confirmed subject mobile location data and backpedaling data are provided to the suitable authorized organization. Individual mobile subscriber controls the present data and backtracking data. Therefore, the COUNTERACT system notifies each consumer about being exposed to contagion, and the consumer regulates whether tracing data would be studied. The suspect's location data is delivered to specialists after applying the filter, censoring the consumer' multimedia, calling, browsing, and text-related data to guard the consumer's privacy. This location information recognizes regions that may be contagious, trails of contagion, separation, etc. The information is useful to acknowledge, oversee, and possibly hold any infection point source. The system offers a map of the infection exposure for each confirmed subjects and suspects separately. The map may define the chronological development of the infection and may support in forecasting its spread. COUNTERACT developed an approach to the real-time recognition, managing, and control of epidemics. By tracing chronological location data, it can offer historic and recently received data. Up until today, not considerable may be prepared. Nevertheless, with pervasive cellular mobile phones and telecommunications networks, the ability to continuously alert consumers' locations have been generated. COUNTERACT system is inevitable and necessary in the present scenarios of virus outbreaks, i.e., COVID-19.

2.1. Methodology

The COUNTERACT system includes a data mining module, database server module (temporary), correlation module, user thresholds (based on proximity and time spent) and alert rules module, geographic tracing module, location reception sub-module, contagion site tracing module, recognition module, mobility behavior model and mapping module, this better presented by Fig. 1. Fig. 1 represents the complete architecture and design of COUNTERACT. Once the user requests the system with the mobile number of the subject for the infection tracking, the system sends the request to the GPS activities' teleservice provider. Teleservice provider gets access information of GPS data from Lawful Interception Box (LIB) and Lawful Interception Controller (LIC), and the data is sent through the internet link to Database Server Module. Database server module data is delivered to different modules and sub-module for analytical processing. GPS data is a big data set. It gets user mobility-related pieces of information for many events and hundreds of hours, so the Data Pruning Module prunes it. Data Pruning Module removes the data, which is not useful either for future use or too distorted. Another analytics module is Data Mining Module, which has two sub-modules: Mobility Behaviour Module and Correlation Server Module. Mobility Behavior Module checks the subject's mobility behavior for the required days and is responsible for forecasting and generating the missing mobility data.

In contrast, the Correlation Server Module identifies the correlations between various mobility events. User Threshold and Alert Rules Module, which has the sub-module Authorities/Users, perform the spatio-temporal analysis to recognize the probability of possible contagion based on proximity and time spent between subjects. The above modules are based on the HP Edge Line Computing Device (Anonymous, 2021).

The mentioned sections are responsible only for the domestics spread recognition and tracing, but it is not complete when the virus becomes pandemic; for this purpose, COUNTERACT introduces Virtual Private Server. Virtual Private Server Based on Edge Computing gets GPS data
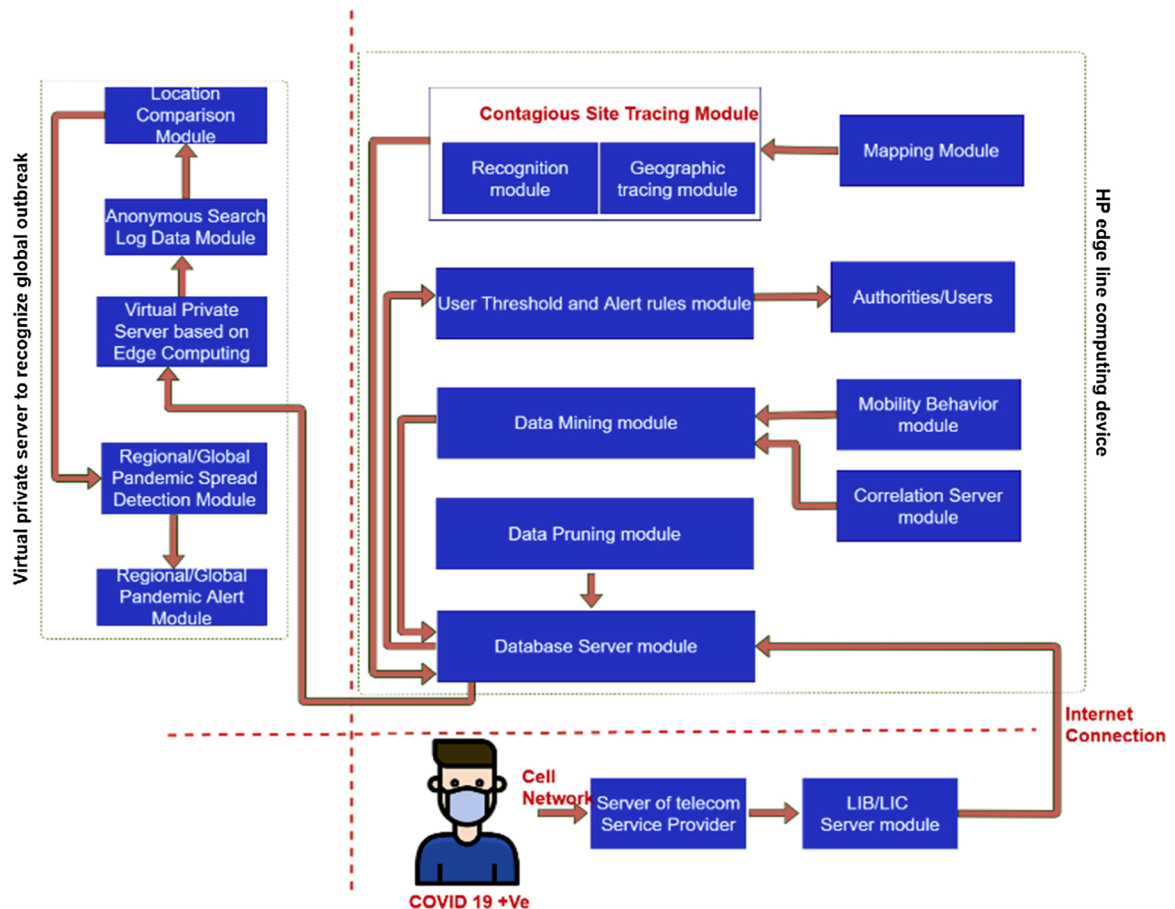

Fig. 1. COUNTERACT system block diagram.

success of various infected confirmed subjects and their respective contagious sites. The privacy of infected subjects is maintained through the Anonymous Search Log Data Module. This GPS data from the Log Search Module is sent to Location Comparision Module to recognize a similar town/area or beyond that. The Location Comparison Module report is delivered to the Regional/Global Pandemic Spread Detection Module to verify the type of spread (Global or Regional). Once the Regional/Global Pandemic Spread Detection Module decides, the Regional/Global Pandemic Alert Module issues alert.

User location data is accessed through user cell phones and GPS. Cell phone location information is collected from a mobile phone base station, and this location data is stored in the database. The correlation module augments location recognition data (bus, cineplex.etc) by adding tree connection to the other users' who have concurred location. Correlation of location is the function of time and distance, location overlapping, and cross path for adequately enough time with proximity. Such as, waiting at the bus stop may not activate a proximity correlation, but being on the bus to a commute for 15 min in the enclosed area may.

The contagion tracing module comprises deep learning models trained-tested to obtain location data related to the tree of registered users and produce a prediction of suspects whose possibility was exposed. The geographic tracking module defines coverage intensities for the suspected users who have not been recognized as a possibility to get contagious. The geographic tracking module can use the location data linked with both confirmed (infected) subjects and suspected subjects to precisely define how frequently each subject has been exposed to the infected subject. The geographic tracking module collects suggestions of possibly diseased subjects from the recognition module. It connects the infected subjects' user location data with that of the suspected subjects to define coverage intensities for the suspected subjects. Suppose location data for subjects specify that suspected subjects were recorded in identical geographic regions as confirmed subjects inside a similar time stamp. In that case, the geographic tracking module can raise a coverage (congestion) level linked with the suspected subject. The geographic tracking module can decide a contagion level for each suspected subjects based on risk elements comprising the total number of encounters each subject had with confirmed infected subjects, and the period that each subject was exposed to a particular subject, the passed time between coverages to confirmed subjects, or both the conditions. The risk level can be exponentially varied for each encounter a suspected subject has with a confirmed subject conferring to each encounter's contagion risk features. The risk level of encounters can be characterized as a scale value from zero to a hundred "contagion risk grade, " with zero signifying no coverages to the confirmed subject. A meeting with a longer duration and proximity will have a high increase in grades than a short time. The geographic tracking module splits a geographic area into a range of cell sites. The geographic module defines coverage intensities for each subject separately by recognizing that if the suspected subject' cell-site data coincide with confirmed subject data location within the same range of time and proximity. The contagion tracing module produces singularized subject forecasts for suspects that the geographic tracing module classifies as probably to develop the infection. The contagion tracing module can generate an infection warning report (a text message or phone call) to the suspected individual to notify the suspect about his recent mobility caused by his exposure to contagion, and they will have symptoms soon. The mapping module offers location data plotting of the contagion coverage according to the suspected subject location data.

Subject spatial data is observed over timestamp values from subject mobile phones. COUNTERACT system relies entirely on mobile phone location as well as GPS data, so in that case, the system designed Mobility Behavioral Modelling. The behavioral mobility modeling receives historical location data to apply a deep learning model to it. This model generates the location data of a particular duration with acceptable accuracy. The correlation module produces this behavior-based mobility data projection. Once the mobility pattern is predicted, the correlation module sends the data to the database module. Database module associated with each mobile subject stores the location of the subject. This user database contains the usual mobility behavior and anomaly (unique) mobility pattern measured by mobility modeling. Such as, the subject may have different behavior on every second Friday as compared to other weekdays.

The contagion tracing module could also be extended for the preventive alert notifications to the individuals. It can alert a healthy subject to prevent exposure by the confirmed subject. It identifies the measures that a healthy subject can take to escape an encounter within a confirmed subject's period and proximity. For example, suppose the contagion tracing module recognizes the healthy subject's mobility location data and behavioral mobility pattern. In that case, they will have a cross-path or encounter with the confirmed subject, the contagion tracing module sends an alert notification. This alert notification contains a description of the risk and informing the healthy subject to change the route. Data mining module interfaces with user risk assessment and warnings to produce notices through the alert module. According to the context, the Alert module sends notifications to labeled receivers (suspected subject) once the defined threshold or vigilant settings are encountered. The threshold or vigilant settings comprise past data, present data, in-depth learning analysis, and behavioral mobility patterns.

The sub-module of the geographic tracing module is the location reception sub-module, which is liable for the reception of the subject's location from the mobile phone or GPS. The location reception sub-module is also responsible for translating location data into the preferred format, such as location data from the cell tower, and GPS has a different display. The correlation module uses the location reception sub-module data to map the subject's locations to actual locations (local train, cineplex, etc.).

Fig. 2 represents a layered design of the proposed COUNTERACT system. The GPS based tracking module of the COUNTERACT system starts functioning by acquiring GPS-spatiotemporal data of COVID19 patients from a mobile handset. The COUNTERACT system performs the in-depth analysis, tracking, and tracing of the acquired GPS data of COVID19 individuals. Furthermore, a notification module generates real-time notifications for emergencies and notifies the health-experts and house members. The services module provides location tracking, tracing, and tagging services to the COUNTERACT system. The user management and control module handle the access rights, authentication, and authorization mechanisms of users and various report generations, such as the total number of suspected COVID19 individuals, etc. Hadoop technology is used to query the COUNTERACT back-end based on time-stamp and location details of COVID19 suspects. The collected data will be stored on a Big data technology-based COUNTERACT cloud computing platform (Mongo DB back-end technology). The data logger module is responsible for keeping time-stamp based GPS data of various COVID19 individuals. The application layer is responsible for integration, aggregation, and interfacing of the acquired GPS Spatio-temporal data with the COUNTERACT system mobile application. It also provides real-time location information of COVID19 suspects via a COUNTERACT mobile application.

3. System modeling and performance evaluation

In the undertaken study, we analyzed mobility data network data such as latitude and longitude of COVID-19 persons, their visited locations, the distance between two mobility data devices (within 1.25-meter distance) for 15 days. The researchers have also analyzed mobility patterns through simulation modeling to detect the possibility of COVID-19 spread:
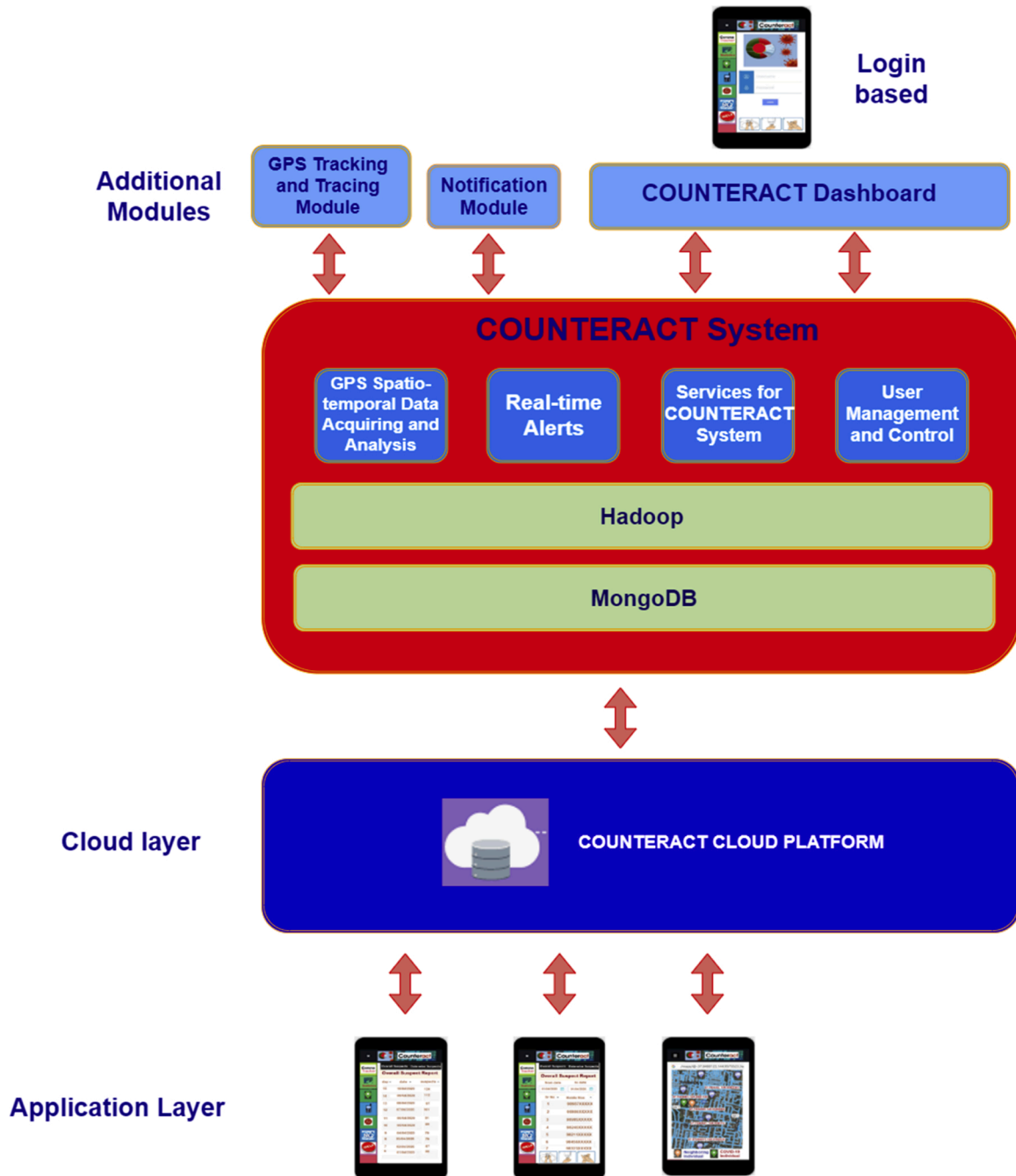
Fig. 2. A layered representation of a COUNTERACT system.

3.1. COVID-19 pandemic direct infection suspicion model

A mobility matrix $Mob^{mobility}$ is constructed using the global mobility network data, with elements $M_{p,q}^{mobility}$. Where, $M_{p,q}^{mobility}$ represents the daily mobility movements of the COVID-19 subject and the number of individuals who came closer to the COVID-19 subject and have crossed the set threshold boundary of 1.25 m. It also represents COVID-19 subject's relocation movements at a particular location p register at a time t, day d, and last day d-1 to mobility data relocation movements for the incubation period (for COVID19, we considered d = 15 days) (Gupta et al., 2020). The calculation of the direct infection suspicion $DS_{(t,d)}$ can be represented by Eq. (1),

$$DS_{(t,d)} = \frac{\left[\sum_{d=1}^{15} \left(C19_p^{mobility} \times \sum_{n=1}^{N} NS_{n,p}^{mobility} \right) \right] \times A_{area}}{D_{C19_p, NS_{n,p}}} \quad (1)$$

$C19_{p,q}^{mobility}$ = COVID19 subject Mobility at location p;

NS_n = The list of neighboring healthy individuals who have traveled near a COVID-19 subject for 10 min at location p;

n = 1, 2, 3…. N;

d = day, 1,2, 3…15;

t = time when COVID-19 subject and neighboring healthy individual come within the vicinity of the set threshold limit of 1.25 m.;

n = total number of individuals who came in contact with COVID-19 subject;

A_{area} = average mobility of the COVID-19 subject and the total

number of neighboring healthy individuals based on the radius $r_{p,q}$ at location p;

$D_{C19_p, NS_{n,p}}$ = distance between COVID-19 subject and the neighboring healthy individuals.

It is a fact that the listed parameter values may vary with time, location, and geographical regions. The proposed model provides a probabilistic analysis of the direct suspected COVID 19 cases based on the accessible mobility network data, which is essential information for tracking the suspicious cases in a particular region at a specific time t and day d. However, to enhance the accessible COVID-19 Pandemic Direct Infection Suspicion Model and recognize the possibility of the number of individuals who have traveled near the directly suspected COVID-19 neighboring individual for 10 min or more. We have modeled an Indirect Infection Suspicion Model to identify the indirectly suspected COVID-19 individuals.

3.2. COVID-19 pandemic indirect infection suspicion model

To model indirect COVID19 epidemic dynamics, a scholastic COVID19 pandemic indirect infection suspicion model is used for finding indirectly suspected COVID-19 subjects for the incubation period. Thus, an indirect mobility matrix $iM^{mobility}$ is constructed using the global mobility network data, with elements $M_{p,q}^{mobility}$. Where, $M_{p,q}^{mobility}$ represents the daily mobility movements of healthy individuals who have traveled near COVID-19 suspected in a particular area for more than 10 min (Gupta et al., 2020). It also shows their relocation movements between the area p and q register at a time t, day d, and last day d-1 to mobility data relocation movements for 15 days. The calculation of the direct infection suspicion $IDS_{(t,d)}$ can be represented by Eq. (2),

$$IDS_{(t,d)} = \frac{\left[\sum_{d=1}^{15}\left(iC19_p^{imobility} \times \sum_{n=1}^{N} iNS_{n,p}^{imobility}\right)\right] \times A_{area}}{D_{iC19_p, iNS_{n,p}}} \quad (2)$$

$iC19_{p,q}^{mobility}$ = Suspected COVID19 Subject Mobility at location p,

iNS_n = The list of neighboring individuals who have traveled near a suspected COVID-19 victim for 10 min at location p,

$D_{iC19_p, iNS_{n,p}}$ = distance between suspected COVID-19 subject and the neighboring healthy individuals.

3.3. Mobility behavioral modeling

As soon as sufficient historical movement records are accessible for a subject, movement patterns can be precisely forecasted by the Support Vector Machine (SVM) Classifier. However, subject movements are unsystematic in behavior conditional to the situation; we can still recognize the mobility behavior (Carvalho, Barbastefano, Pastore, & Lippi, 2020). First, the algorithm finds the pattern in the historical records for the classifier to forecast the pattern. The patterns with the matching behavior are defined under the same class annotation. Ç1, Ç2 Ç3… Çn give the class annotation labels for n number of the movement patter$_n$. $\bar{T}1, \bar{T}2, \bar{T}3$ represent the timestamps, $\bar{T}_k$ and where k is the number of timestamps. The locations are shown by $ĭ_1, ĭ_2, ĭ_3…ĭ_f$, and f is the number of sites. Let us assume the pattern for class

$$Ç_1 = [\bar{T}_1 ĭ_1, \bar{T}_2 ĭ_2, \bar{T}_3 ĭ_3, \bar{T}_4 ĭ_4, \bar{T}_5 ĭ_4] \quad (3)$$

$$Ç_2 = [\bar{T}_1 ĭ_1, \bar{T}_2 ĭ_2, \bar{T}_3 ĭ_3, \bar{T}_4 ĭ_4, \bar{T}_5 ĭ_4] \quad (4)$$

$$Ç_3 = [\bar{T}_1 ĭ_6, \bar{T}_2 ĭ_2, \bar{T}_3 ĭ_5, \bar{T}_4 ĭ_3, \bar{T}_5 ĭ_2] \quad (5)$$

$$Ç_4 = [\bar{T}_1 ĭ_3, \bar{T}_2 ĭ_3, \bar{T}_3 ĭ_3, \bar{T}_4 ĭ_4, \bar{T}_5 ĭ_4] \quad (6)$$

These patterns and classes vary from subject to subject. In the primary phase of the research, forecasting algorithms mine location records every 10 min and mark the patterns hourly. Such as, a subject under test

is at location $ĭ_a$, at time $\bar{T}_1$. So at location $ĭ_b$ at time $\bar{T}_2 = \bar{T}_1 + 10$ min, followed by $\bar{T}_3 = \bar{T}_3 + 10$ min, $\bar{T}_4 = \bar{T}_3 + 10$ min, $\bar{T}_5 = \bar{T}_4 + 10$ min for the locations $ĭ_c$ $ĭ_c$ and $ĭ_d$ respectively. The spatiotemporal series represents the pattern, which we denoted as the class annotation. With the increase in the time period, the classified patterns increase, resulting in a sufficient training testing dataset. This large multiclass dataset is solved by SVM. Let us assume that there is a problem with forecasting class annotation for the mobility dataset Dz. The spatiotemporal pattern for the dataset is given as $\bar{T}1 = ĭ_1, \bar{T}_2 = ĭ_2, \bar{T}_3 = ĭ_3, \bar{T}_4 = ĭ_4$, and $\bar{T}_5 = ?$; now the class label for the above pattern needs to be forecasted by $\sum_{r=1}^{n} Ç_r^{r=1}$. The projected mobility model measures the posterior probability for all the defined classes. Out of all the classes, the class with the greatest posterior probability gets the responsibility to recognize the pattern followed by location prediction. Along with the above method, we also consider using the present GPS coordinates and applying them for the movement prediction. The mobility model uses the distance covered, movement direction, and movement speed to forecast the location in this additive approach. The mobility model acquires the GPS coordinates from the subject's mobile phone handset every 15 s over 5 min. Forecast of upcoming location will be prepared for $\Delta ţ = 5$ min, $\Delta ţ = 10$ min, $\Delta ţ = 15$ min and $\Delta ţ = 20$ min.

The distance and direction of the subject's movement in $\Delta ţ$ is measured by the location coordinates of the two consecutive location points (Latitude_A, Longitude_A) and (Latitude_B, Longitude_B). The direction is recorded through the slope value figured from $(Latitude_B - Latitude_A)/(Longitude_B - Longitude_A)$. Different slope values correspond to other direction of the user from the current location, (Slope = 0, subject's movement either west or east), (Slope>0 as well as Slope< 0, then, subject's movement either northeast or southwest), and $((Longitude_B - Longitude_A) = 0)$, subject's movement either north or south). The approximate speed of the subject is calculated by applying the distance traveled in 5 min. So the distance covered by the subject in $\Delta ţ$ is given as speed * $\Delta ţ$. These measured distances and directions are applied in the prediction of location. The mobility production was for the missing spatiotemporal data. Let S be the set of functioning subjects in the contemporary time period ($\bar{T}$) days, and $\mathbb{R}$ (i) denote the COVID-19 subjects in S(n) containing participant each carrying GPS locator ($\bar{G}$). We briefly describe how we divided the data into two segments, one where the data of all the incubation period was available and the other where the data of at least one incubation day was missing:

The condition I: if S(n) GPS locator ($\bar{G}$) → "OFF": Need to turn "ON" $\bar{G}$ to check and verify mobility prediction model ($\acute{P}$); **Condition II**: if S (n) GPS locator ($\bar{G}$) → "ON": a) For $\bar{T}$ >=full incubation period. (mobility prediction model ($\acute{P}$) to detect $\mathbb{R}$ (n) subject around n). b) For $1 < \bar{T} < 12$, the incubation period ($\acute{P}$ predicts the missing $\bar{T}$ and regenerates the pattern followed by $\mathbb{R}$ (n) subject movement around n).

3.4. System approach based on HP edge line computing, Artificial Intelligence, and Internet of Things

Neural Networks (NNs) is a capable methodology for separating specific information from user mobility location data of IoT gadgets conveyed in tricky conditions. In light of its multilayer configuration, NNs is likewise proper for the edge computing condition. This way, in this article, we initially bring NNs for IoTs into the edge line computing. As existing nodes have restricted computational capacity, we likewise plan an offloading methodology to enhance IoT NNs applications with edge processing. In the performance assessment, we test how to implement various NNs events in an edge processing condition with our technique. The assessment results show that our approach outflanks other technological advancements on NNs for IoT. As edge computing unload processing events from the central cloud to the edge close IoT nodes, the transmitted information is tremendously decreased by the pre-processing systems. The edge computation can implement well

when the transitional data dimensions are lesser than the input data dimensions. Each layer of NNs can immediately downsize the transitional data size until enough highlights are found. So, the NNs fit for edge computing since it is conceivable to upload portions of NN layers in the edge line and move the diminished transitional data to the virtual private cloud server.

Another benefit of NNs in edge processing is the privacy aspects in transitional data delivery. Transitional data produced in customary large information frameworks, for example, MapReduce or Spark, contains client privacy (Thai et al., 2016) as the pre-processing stays as information semantics. The transitional data in NNS, for the most part, have diverse semantics contrasted with the source information. For instance, it is challenging to comprehend the primary information with the highlights mined by a NNs filter in the intermediate NNs layer. Subsequently, in this article, we NNs for IoT into the edge line computing to improve learning execution to decrease traffic. We figure a versatile model that is perfect with various NNs. In this way, due to the diverse transitional data size and pre-processing overhead of various NNs, we express a scheduling issue to expand the quantity of NNs tasks with the limited system data transfer capacity and administration ability of edge nodes. Research additionally attempts to ensure the quality of service (QoS) of every NNs administration for IoT in the scheduling. We plan offline and online planning models to take care of the issue. Research performs simulation with different NNs undertakings and given edge processing settings. The exploratory outcomes show that our answer outflanks other improvement strategies on NNs for IoT (Fernandes & GL, 2017; Li, Ota, & Dong, 2018). The primary commitments of this sub-section are summed up as follows. We initially bring NNs for IoT into edge line computing. As far as we could know, this is an inventive work concentrating on NNs for IoT with edge line computation. We figure a flexible model for different NNs for IoT in edge computing. Likewise, this study structures a useful online model to advance the QoS and bandwidth of the edge line computing approach applied.

The COUNTERACT system uses the HP edge line computing device and its features of advanced edge computing. In the HP edge line computing setup, the privacy-preserving edge line servers are denoted by set S, whereas s_e represents an edge server. The HP virtual private server is used as a cloud server, which is denoted by C_s. Sc, B represents the service capacity, bandwidth, and threshold parameter (threshold value prevents or minimizes the network congestion caused by the traffic between the s_e and Cs), and ɥ, respectively. B* ɥ represents sp maximum accessible bandwidth for the communication and traffic between the HP edge line server s_e and Cs. The all deep learning tasks are denoted by the set ┼, whereas the τ_j represents a deep learning task of set ┼ with the bandwidth Bj. τ_j contains ŭj number of layers. Rpj represents the average ratio of the transitional data magnitude produced by the Pth layer [P ϵ [1, ŭj]] with respect to the total input data size. Đij shows the applied input dataset size per unit of τ_j for an assigned edge line server s_e, and the transferring latency is represented by

$$Ł = Đ_{ij} * \frac{Rpj}{Bj} \tag{7}$$

The threshold value of transferring latency is set to Łj to maintain a reasonable quality of service. ₵Pj represents the computational overhead after the Pth layer, and the complete task overhead is $₵_{Pj} * Đ_{ij}$. The projected HP edge line computing architecture commits to allot optimum allowable tasks [$\tau_1, \tau_2, \tau_3 ... \tau_j$] through neural network layers integrated with IoT. This offers a significant reduction in latency. The projected system offers real-time operation and addresses the scheduling issue through the dynamic switching between offline/online models. During scheduling, the offline model investigates the P_{th} layer's parameter Pj^p to support the optimum value of $R_{pj} *₵_{Pj}$ for the edge server sj^p. Now the tasks are arranged in ascending sequence with respect to the input dataset size. The model starts with the scheduling of the smallest input dataset size in the edge line server. Moreover, this model monitors the edge line server resources for the effective and

successful execution of task τ_j. In case the edge line server does not have acceptable bandwidth as well as QoS, the model varies the value of P until it finds the best P for the task execution in s_e. Even after checking all P values scheduling model fails to find a suitable P-value in edge line server s_e; the schedule model postpones task τ_j. The offline model's complexity is given $(|┼| * |S|^2 * P)$. On the other hand, an online scheduling model is used to evaluate the condition for τ_j task implementation. As the scheduling model receives limited features related to the current task, the model relies on past performance features. Indexes β_1 and β_2 represent the maximum and minimum necessary bandwidth of a given task, correspondingly. Therefore, for task τ_j, the system first computes the Pj^p and sj^p. Followed by measuring a value

$$F(Oꝑ_{ij}) \longleftarrow (\beta_2 \cdot ꜧ/ \beta_1) * (\beta_1 * ꜧ) \tag{8}$$

$Oꝑ_{ij}$ is the reserve service capacity of the edge line server, and ꜧ is the constant.

In case

$$(B_{ij}^p - Đ_{ij}^p * R_{pj}^p / Ł)) * (Oꝑ_{ij} - Đ_{ij}^p \cdot {}_{ij}^p) \leq F(Oꝑ_{ij}) \tag{9}$$

and other edge line servers have adequate resources, the scheduling model implements task τ_j, and the estimated relationship of the online scheduling model is ($\frac{1}{(\ln(\frac{\beta_1}{\beta_2})+1)*}$). Fig. 3 represents various GIU based screenshots and report screens of the proposed COUNTERACT system.

4. Performance evaluation, discussion, and observation

For implementing IoT and AI applications, the above mentioned computational resources (MSI GP73 Leopard 17,3 8RF-648NE, Nvidia GeForce GTX 1070, 8 GB, Intel® Core™ i7-8750H processor, 16 GB DDR4 RAM, HP edgeline server EL1000) have been used. The edge computing was introduced because the backtracking and suspect finding based on the user's location data had the challenges of big data handling, high computation, real-time processing, privacy-preserving, and uploading information to the HP virtual private server. The present system applied Keras Framework for Low short Term Memory (LSTM) neural network on Google Co-laboratory Open Source Python Tool. As presented in Fig. 4(a), two LSTM networks were selected, LSTM1 and LSTM2. The blue color plot represented the reduced data size ratio, and red color plots show the computational overhead. The deep learning neural networks are tested up to seven layers, but the results of up to five layers are considered because of the performance consideration. These layers were had different combinations of neurons and activation functions. Fig. 4(a) represented the number of neural networks layers reduced that applied input spatiotemporal data size. In contrast, the computational overhead increased by an increase in the number of layers. The maximum number of backtracked contacts tasks was up to 850. The input data size of each spatiotemporal backtracked contact was 774KB to 1700KB. The applied HP edge line server had excellent bandwidth and latency management. Fig. 4(b) shows the different computational time required while backtracking the contacts with superficial LSTM layers and layer scheduling concept. Layer scheduling over performed as compared to classical layer structure. When the COUNTERACT software was installed on the smartphone (Samsung Galaxy S9), the system tool 57 min. to prepare the infection report, while when the same system installed on the computer (MSI GP73 Leopard 17,3 8RF-648NE, Nvidia GeForce GTX 1070, 8 GB, Intel® Core™ i7-8750H processor, 16 GB DDR4 RAM) the computational time was 13 min., and computer system with HP edge line server EL1000 has reduced the computational time to 4 min. The same experiment was performed on the Smart Phone Samsung Galaxy S9 cell phone. The computational time was 15–18 times higher compared to a dedicated HP edge line server setup. Fig. 4(c) represents the performance of the selected inline algorithm for task scheduling as comparing the other two
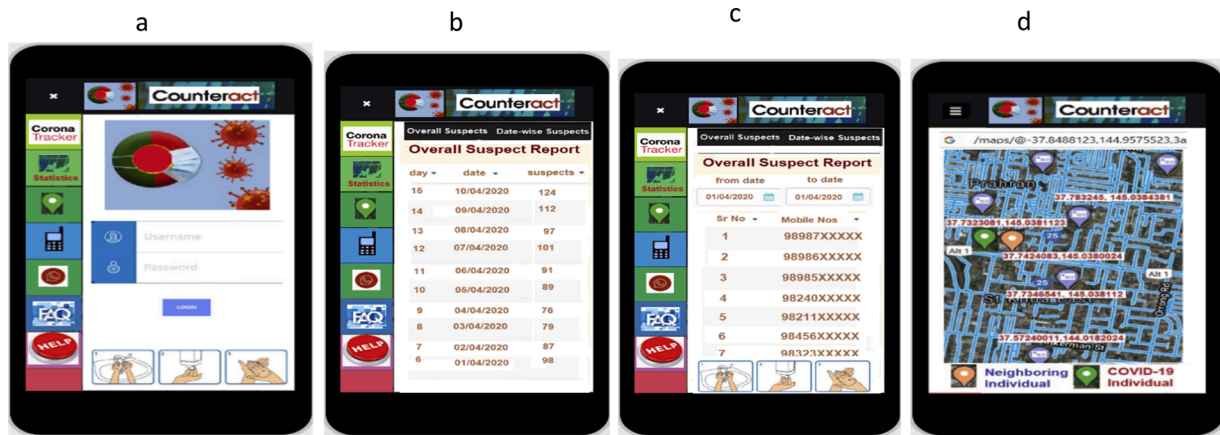
Fig. 3. Screenshots of COUNTERACT graphical user interface.

(a)   (b)   (c)
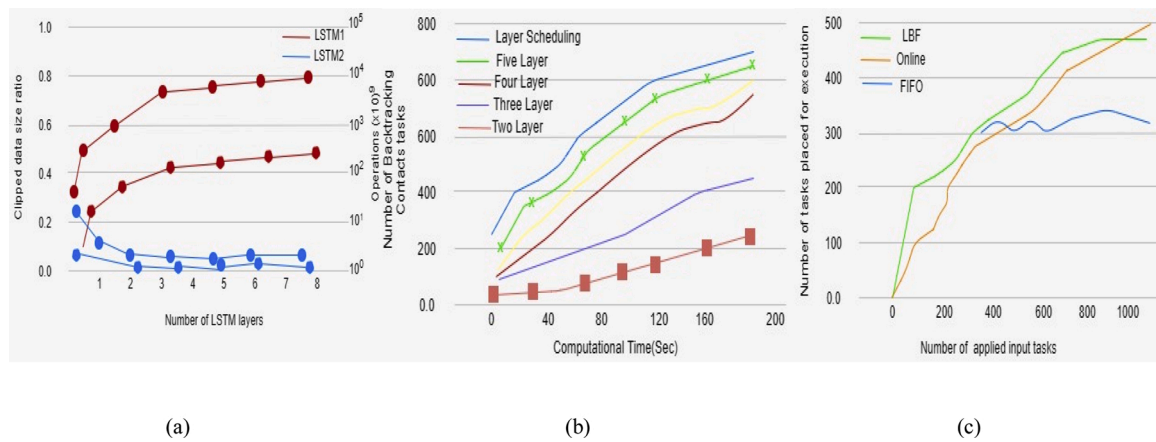
Fig. 4. (a) Reduced data size ratio vs Number of Deep Learning Network Layers, (b) Performance of Layer Scheduling and classical layers for various task load to computational time (c) Online Task Scheduling algorithm evaluated, compared with FIFO, LBF under different applied tasks and deployed tasks.

First in First Out (FIFO) and Low Bandwidth First Deployment (LBF). FIFO model showed the deployment of the task until it ran out of bandwidth and network resources. FIFO recorded plot values from 384 tasks. LBF was on the similar trend as FIFO in terms of the task deployment approach. As LBF recognized the deficiency of bandwidth, it pulled out the task when maximum bandwidth demand arises. The selected online algorithm for the task scheduling performed significantly well when the input task was near 465 (outperformed FIFO) and 987 (Outperformed LBF).

We collected 126 confirmed COVID-19 subjects' spatiotemporal data in the incubation period for backtracking to conduct the recognition of contagion spread and hazard location. Only 23 % of subjects have all 15 days of data available; the rest subjects data were missing for some days. In this case, we had an immediate need to have gold standard data where spatiotemporal data free from noise and bias, so we prepared data for 15 days with our controlled subject. We annotated this control subject as a COVID-19 subject who has a regular office routine. We also annotated the following actors for the controlled subjects spatiotemporal data: H1=COVID-19 subject's (our Gold standard) friend, H2= Shopkeeper where our subject used to get breakfast, H3= Cinema boy, H4, H5= Persons sitting next to COVID-19 subject one on the left and other on the right, H6, H7= Persons sitting next to COVID-19 subject in bus one on the left and other on the right, H8=COVID-19 subject's neighbor on the right side, H9=COVID-19 subject's neighbor on the left side, H10=Food delivery boy, H11=Milk delivery boy, H12-H13=COVID-19 subject's house-made, H14-H15=COVID-19 subject's roommate. Below, Fig. 5(a) shows the mobility and interaction of the controlled subject to the other

subjects that came across in daily living activity on one particular day, via the simulation on MATLAB R2018a.

Fig. 5(b) shows, on the same day, when H1 meet the controlled subject, H1 fulfilled the condition of getting infected; based on the co-ordination, we discovered that H1 spent more than 27 min with a distance of less than 2 m approximately. The performance of the infection prediction was the function of spatiotemporal data completeness in 15 days of incubation, that's why for the 15 days available data, system performance was optimum. The real number of infected people was done by contacting and confirming suspected people who came in contact with the particular COVID-19 subject. In contrast, the prediction solely depended on the spatiotemporal data-based backtracking. When we collected GPS-based spatiotemporal data, the data had high bias as the proximity is approximate. We did not get the exact distance between the infected subject and suspect, so it caused the system to consider the distance threshold 2 m $\pm$ 2 m. This caused the system to produce many false-positive (FP) predictions of suspected infected subjects. As (a) where all days data was available and we did not implement behavioral mobility prediction, we recorded the least number of false-negative (FN) 0.010 as compared to the other three categories [(b), 2–3 days data missing, FN = 0.13], [(c), 4–8 days data missing, FN = 0.34], [9–12 days data missing, FN = 0.39]. Whereas for the same reason as (a) had all the incubation data available, it recorded the highest false positive 0.10 as it considered standard deviation distance 2 m. The proposed system was designed in such a way so that it would not miss any infected subject, i.e., false-negative should be as small as possible. Fig. 6(a) represents the ROC curve of all four confusion matrix, where complete
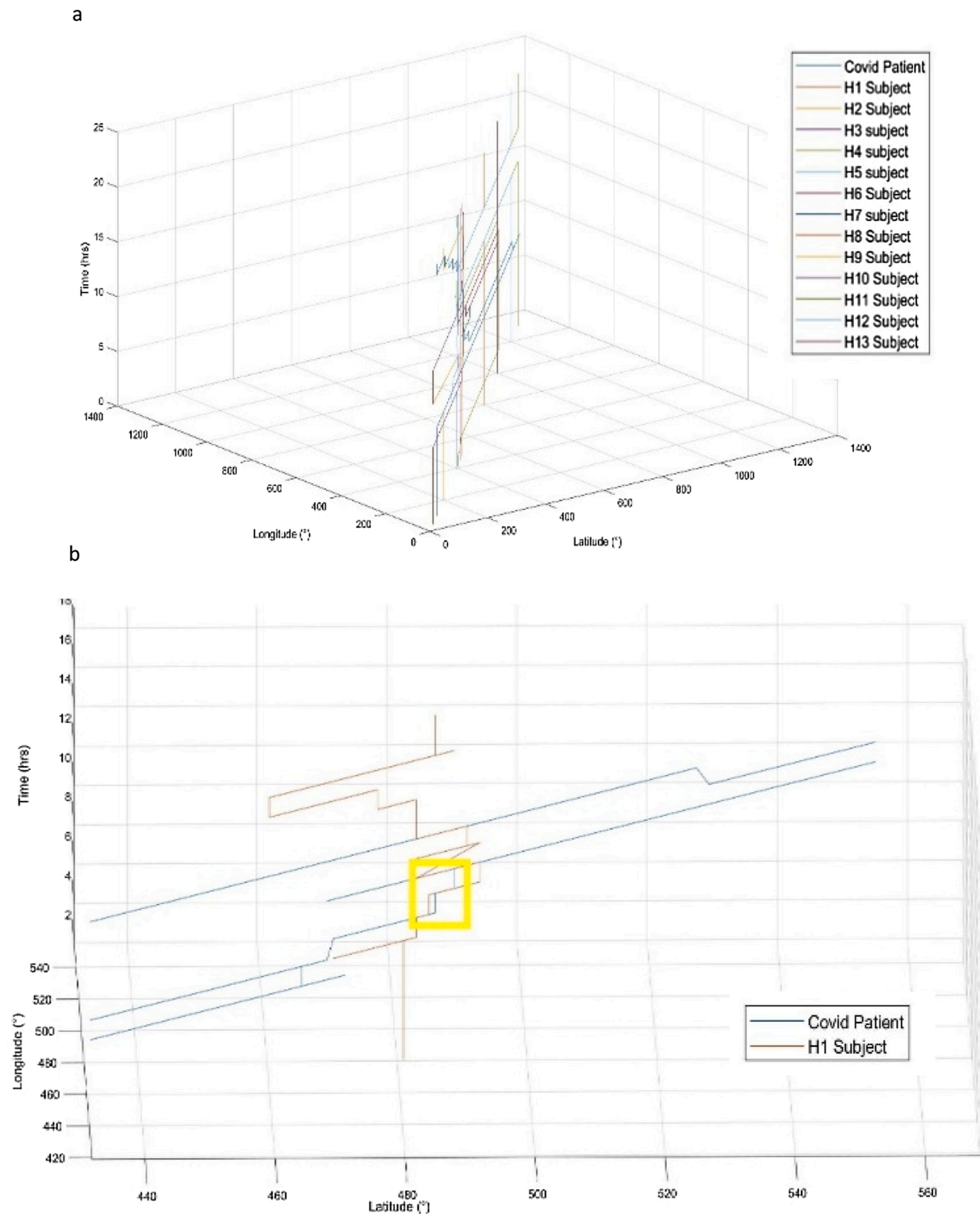
Fig. 5. (a) 3D COVID-19 Subject moments relative to other subjects, (b) 3D moment of COVID subject relative to his friend (H1), and yellow box highlight the time duration they spend together, that caused him (H1) to get effected with COVID-19 as well.

days data secured a promising curve through exceptional accuracy 0.945, sensitivity 0.90, and specificity 0.99; on the other hand, performance went down with the decrease in incubation days data as [2–3 days data missing, accuracy 0.84, sensitivity 0.81 and specificity 0.87], [4–8 days data missing, accuracy 0.67, sensitivity 0.69 and specificity 0.66], [9–12 days data missing, accuracy 0.61, sensitivity 0.62 and specificity 0.61]. As the proximity of suspect subjects with COVID-19 subject goes beyond the threshold

distance, the chance of getting infected reduces with increase in distance and decrease in time spent together. Whereas when the infected subject and suspected subject at proximity less than 2 m, there is a higher probability of infection even when they spend less time together. Fig. 6(b) shows 3D probability distribution with a change in proximity and time spent together. The scale axis defines the score of chances of infection from -10 (which is safest) to +10 (highest probability of getting infected).

a
ROC Curves Comparison Analysis

b
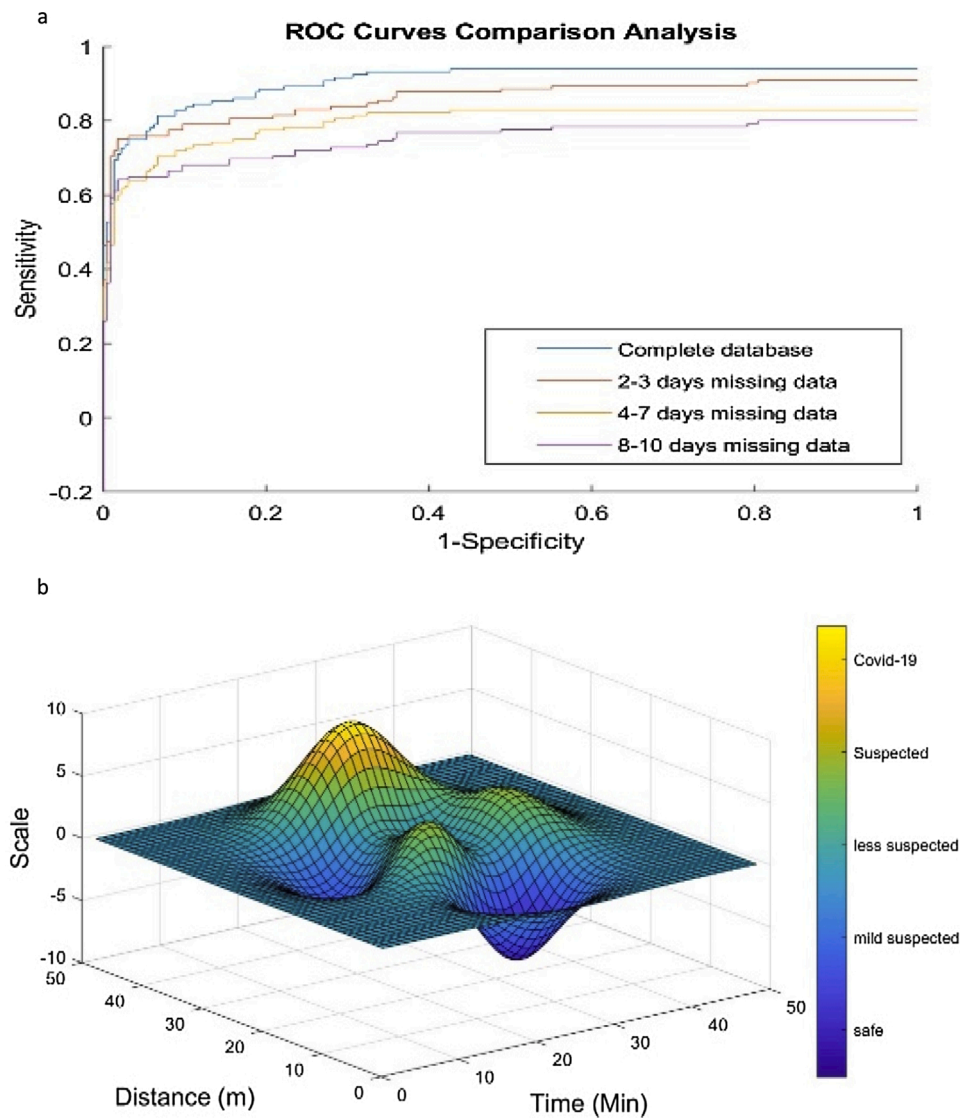
Fig. 6. (a) ROC curve of COUNTERACT system performance on a prediction of spread among healthy subjects, (b) 3D Simulation of the contagion spread among healthy subject from a controlled COVID-19 subject to the threshold value of proximity and time spent.

5. Limitation and future work

The incomplete spatiotemporal data in the instance files stops us from correctly evaluating which movement algorithm best records the spread of authentic infection. However, the COUNTERACT system is an integrated contagion spread recognition approach and significant for community well-being resolutions, etc. Spatiotemporal mathematical prototypes of spread execute finest with added spatiotemporal bigger data size. The event statistics are not entirely complete to conduct the disease spread at its best. One of the reasons for this limited data is the assumption on the mobility of missing incubation bay's data. Therefore, a position variation can fail to be seized if the GPS and the mobile phone are off, and the projected location for a specified day might not resemble the location where the individual has been to. The spatial measure balances seizing applicable signal and particulars, processing parameters magnitude, and measuring out the noise. This assumption of mobility prediction increases the bias in spatiotemporal data.

Moreover, the location data's privacy is a critical aspect that needs to address without solving it; people hesitate to use the software tool. The experiments are ongoing to resolve privacy issues through differential privacy (Qiao et al., 2019) for GPS location data. Differential privacy approaches work on the anonymization approach, which assures the

resultant of computing completed on the applied database will not differ considerably despite even if any subject's data are incorporated in the database.

6. Conclusion

COUNTERACT system comprises reception of subject location data; treating the location data to produce location information, the location information consisting of a point of contact between the confirmed COVID-19 subject and the suspected subjects. Disease trajectory indicated the relation of the neighboring individual with disease trajectory. Location information was sorted and updated regularly in the storage. The update included shaving of location information, of negligible risk according to the proximity, and removing time/date periods where the COVID19 subject spent less time with the neighboring subject. The proposed tool generated a list of adjacent subjects based on the mobile number of COVID19 subjects. COUNTERACT is a computer-implemented infection forecast technique performed by a computing system. It is including identifying, based on the subset of the population who are likely carrying an infection. Recognizing, by the computing system, an exposure level of a user to the contagion based on a correlation of a first location data related with the user with a second location

data related with one or more users in the subset of the population who are likely carrying the infection; recognizing, by the computing system and based on the exposure level, whether the user is likely to be or become ill; providing, for display on a user computing device, a notification indicating that the user has been exposed to the infection. It identifies an action for avoiding exposure to the contagion; and providing a notification alerting the user to the action to a computing device associated with the user. Determining the user's exposure level to the contagion determines that the user was present in a geographic region within an exposure window. There are highly populated developing countries such as India, Brazil; they suffered most during the lockdown phase. These developing economies were finding it hard to manage the primary day to day facilities and livelihood. The proposed system can serve in a new normal condition of the virus outbreak, where instead of shutting down the nation, we can practice social distancing and track the spread of the virus. This is the modern and technologically advanced path for the sustainability of society.

Consent to participate

All participants gave consent to participate in the data collection process. Written informed consent was obtained for publication of this research article.

Ethics approval

The study protocol was designed according to the hospital's clinical study regulation and approved by the Internal Ethics Committee.

Funding

Declaration of Competing Interest

The authors declare no conflict of interest.

References

Ahmed, N., Michelin, R. A., Xue, W., Ruj, S., Malaney, R., Kanhere, S. S., et al. (2020). A survey of covid-19 contact tracing apps. *IEEE Access*.

https://www.hpe.com/us/en/servers/edgeline-systems.html (11th Nov 2020).

Bello-Orgaz, G., Jung, J. J., & Camacho, D. (2016). Social big data: Recent achievements and new challenges. *Information Fusion, 28*, 45–59.

Bibri, S. E. (2018). The IoT for smart sustainable cities of the future: An analytical framework for sensor-based big data applications for environmental sustainability. *Sustainable Cities and Society, 38*, 230–253.

Carvalho, D., Barbastefano, R., Pastore, D., & Lippi, M. C. (2020). A novel predictive mathematical model for COVID-19 pandemic with quarantine, contagion dynamics, and environmentally mediated transmissio18n. *medRxiv*.

Cencetti, G., Santin, G., Longa, A., Pigani, E., Barrat, A., Cattuto, C., et al. (2020). Using real-world contact networks to quantify the effectiveness of digital contact tracing and isolation strategies for Covid-19 pandemic. *medRxiv*.

Fahey, R. A., & Hino, A. (2020). COVID-19, digital privacy, and the social limits on data-focused public health responses. *International Journal of Information Management, 55*, Article 102181. https://doi.org/10.1016/j.ijinfomgt.2020.102181

Fernandes, R., & GL, R. D. S. (2017). A new approach to predict user mobility using semantic analysis and machine learning. *Journal of Medical Systems, 41*(12), 188.

Gupta, R., Bedi, M., Goyal, P., Wadhera, S., & Verma, V. (2020). Analysis of COVID-19 tracking tool in India: Case study of aarogya setu mobile application. *Digital Government: Research and Practice, 1*(4), 8. https://doi.org/10.1145/3416088. Article 28 (December 2020).

Jamal, S., & Habib, M. A. (2020). Smartphone and daily travel: How the use of smartphone applications affect travel decisions. *Sustainable Cities and Society, 53*, Article 101939.

John Leon Singh, H., Couch, D., & Yap, K. (2020). Mobile health apps that help with COVID-19 management: Scoping review. *JMIR nursing, 3*(1), Article e20596. https://doi.org/10.2196/20596

Klar, R., & Lanzerath, D. (2020). The ethics of COVID-19 tracking apps – Challenges and voluntariness. *Research Ethics, 16*(3-4), 1–9. https://doi.org/10.1177/1747016120943622

Kontokosta, C. E., & Hong, B. (2021). Bias in smart city governance: How socio-spatial disparities in 311 complaint behavior impact the fairness of data-driven decisions. *Sustainable Cities and Society, 64*, Article 102503.

Li, H., Ota, K., & Dong, M. (2018). Learning IoT in edge: Deep learning for the Internet of Things with edge computing. *IEEE Network, 32*(1), 96–101.

Lu, M., An, K., Hsu, S. C., & Zhu, R. (2019). Considering user behavior in free-floating bike sharing system design: A data-informed spatial agent-based model. *Sustainable Cities and Society, 49*, Article 101567.

Maccari, L., & Cagno, V. (2020). Do we need a Contact Tracing App? *arXiv preprint arXiv: 2005.10187*.

Magklaras, G., & Bojorquez, L. N. L. (2020). A review of information security aspects of the emerging COVID-19 contact tracing mobile phone applications. *arXiv preprint arXiv:2006.00529*.

Mallik, R., Sing, D., & Bandyopadhyay, R. (2020). GPS tracking app for police to track ambulances carrying COVID-19 patients for ensuring safe distancing. *Transactions of the Indian National Academy of Engineering, 5*, 181–185. https://doi.org/10.1007/s41403-020-00116-8

Martin, T., Karopoulos, G., Hernández-Ramos, J. L., Kambourakis, G., & Fovino, I. N. (2020). *Demystifying COVID-19 digital contact tracing: A survey on frameworks and mobile apps, wireless communications and mobile computing*. https://doi.org/10.1155/2020/8851429

Menni, C., Valdes, A. M., Freidin, M. B., et al. (2020). Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nature Medicine, 26*, 1037–1040. https://doi.org/10.1038/s41591-020-0916-2

Qiao, Y., Liu, Z., Lv, H., Li, M., Huang, Z., Li, Z., et al. (2019). An effective data privacy protection algorithm based on differential privacy in edge computing. *IEEE Access, 7*, 136203–136213.

Silva, J. C. S., de Lima Silva, D. F., Neto, A. D. S. D., Ferraz, A., Melo, J. L., Júnior, N. R. F., et al. (2021). A city cluster risk-based approach for Sars-CoV-2 and isolation barriers based on anonymized mobile phone users' location data. *Sustainable Cities and Society, 65*, Article 102574.

Skoll, D., Miller, J. C., & Saxon, L. A. (2020). COVID-19 testing and infection surveillance: Is a combined digital contact-tracing and mass-testing solution feasible in the United States? *Cardiovascular Digital Health Journal, 1*(3), 149–159. https://doi.org/10.1016/j.cvdhj.2020.09.004

Thai, M. T., Wu, W., & Xiong, H. (Eds.). (2016). *Big data in complex and social networks*. CRC Press.

Urbaczewski, A., & Lee, Y. J. (2020). Information Technology and the pandemic: A preliminary multinational analysis of the impact of mobile tracking technology on the COVID-19 contagion control. *European Journal of Information Systems*, 1–10.

van Franch-Pardo, I., Napoletano, B. M., Rosete-Verges, F., & Billa, L. (2020). Spatial analysis and GIS in the study of COVID A review. *The Science of the Total Environment, 739*, 140033. https://doi.org/10.1016/j.scitotenv.2020.14003