CrossMark

An intelligent system for predicting and preventing MERS-CoV infection outbreak

Rajinder Sandhu[1] · Sandeep K. Sood[1] · Gurpreet Kaur[1]

Abstract MERS-CoV is an airborne disease which spreads easily and has high death rate. To predict and prevent MERS-CoV, real-time analysis of user's health data and his/her geographic location are fundamental. Development of healthcare systems using cloud computing is emerging as an effective solution having benefits of better quality of service, reduced cost, scalability, and flexibility. In this paper, an effective cloud computing system is proposed which predicts MERS-CoV-infected patients using Bayesian belief network and provides geographic-based risk assessment to control its outbreak. The proposed system is tested on synthetic data generated for 0.2 million users. System provided high accuracy for classification and appropriate geographic-based risk assessment. The key point of this paper is the use of geographic positioning system to represent each MERS-CoV users on Google maps so that possibly infected users can be quarantined as early as possible. It will help uninfected citizens to avoid regional exposure and the government agencies to manage the problem more effectively.

Keywords MERS-CoV · Cloud computing · Big data · Bayesian belief network · Geographic positioning system (GPS) · Information granulation

✉ Rajinder Sandhu
  rajsandhu1989@gmail.com

  Sandeep K. Sood
  san1198@gmail.com

  Gurpreet Kaur
  preetgurgk@gmail.com

1 Computer Science and Engineering Department, Guru Nanak Dev University, Regional Campus, Gurdaspur, Punjab, India

🕿 Springer

1 Introduction

Healthcare is a very substantial issue to be managed effectively for any nation's development. One of the most vital challenges for any government healthcare agency is to protect its citizens from disease outbreak which can spread from one citizen to another through air or other common mediums [1]. Viruses affect humans in large numbers as these are microorganisms that can propagate very easily. Corona virus is one of the common viruses which can cause multiple diseases in humans as well as animals. In 2015, its outbreak in South Korea has caused more than ten deaths, and thousands of citizens got infected [2]. MERS-CoV infected person is having primary symptoms such as respiratory diseases with fever, cough, shortness of breath (SoB), and diarrhea. Some of the secondary symptoms of MERS-CoV infection are acute renal failure, multiorgan failure, acute respiratory distress syndrome (ARDS), and consumptive coagulopathy [3]. The MERS-CoV source is not yet confirmed; however, many studies suggested that this is an animal virus. Reports have confirmed the transmission of the virus from human to human, but proper intermediate hosts for its transmission are still unknown which further increases its risk [4]. MERS-CoV is a novel corona virus that was first identified in the residents of Middle East countries in late 2012 with a death rate of more than fifty percent in infected citizens. As the studies suggested, MERS-CoV spreads using common mode of contacts from one individual to another [3]. MERS-CoV is declared as a public health threat; therefore, the government's healthcare agencies are committed to provide affordable medical facilities and suggestions to the infected citizens. One of the most important steps to cure MERS-CoV is to stop its infection rate and transmission process by early identification of infected patients as well as providing precautionary measures to uninfected citizens. Identified infected patients should be quarantined as early as possible so that they do not contribute further in the infection spreading. The healthcare sector and the government agencies are adopting cloud computing with data mining approaches to control different infectious diseases [5]. Cloud computing can offer benefits such as improvement in quality of medical services, increased risk responsiveness, collaborating mobile apps, large storage space, and reduced capital investment.

Cloud computing is an emerging technology which provides information technology (IT) resources over the Internet. It is different from traditional IT hosting services because it delivers IT resources on demand without the need of any capital investment with pay-as-use pricing model [6]. With the increase in the use of digital patient's records in hospitals, large amount of data related to patients, doctors, medication, and the staff are generated. To store, analyze, and effectively represent the outcomes of data mining on this rapidly generated data with minimal time, hospitals need bigger IT infrastructure. Hence, IT expenditure in hospitals is continuously increasing day by day for installation and maintenance of hospital management systems. As anticipated, in the near future, the traditional IT infrastructure setup will not be sufficient to provide necessary quality of service (QoS) parameters. Growth of healthcare systems using cloud computing is emerging as an effective solution with the benefits of more dependable Qos, reduced costs and flexibility [7]. Infectious diseases like MERS and H1N1 happen rarely, so IT infrastructure used to prevent outbreaks of these infectious diseases should be adaptive and scalable in nature. Cloud computing is more suitable

Table 1 Incentive for moving healthcare information to cloud computing

Parameters	Description
Collaboration	Through cloud technologies, the information can be shared quickly with authorized physicians, healthcare, and the governmental agencies situated at various geographic regions in real time
Rapidity	Cloud computing provides on demand, fast, and convenient access to stored data, whereas up gradation of computational infrastructure can be done with minimal service interruption
Mobility	Using mobile and cloud computing together, healthcare services can be achieved anywhere and anytime
Reduced costs	The client can pay only for the actual resource utilization of shared infrastructure without investing in new hardware and its maintenance
Large storage space	Cloud computing provides virtually infinite storage capacity. It supports long-term storage of electronic medical records (EMRs) and electronic health records (EHRs)

for such a varying demand system because it provides virtually unlimited resources on demand with a pay-as-use pricing model [8]. Infection outbreak preventing systems should be highly responsive and provide the desired service with high accuracy. Requirement of cloud service provider to provide agreed QoS facilities high system agility and generates appropriate compensation if requirements are not met [9]. The characteristics defined above act as a motivation to use cloud computing in healthcare infectious disease management system. Some of the significant health service benefits of shifting to cloud computing are listed in Table 1.

To contain the outbreak in its initial stages, ideally the government should test each user for the MERS-CoV infection which is not possible in the normal situations. Hence, the primary aims of the proposed system are

- To provide an initial diagnosis to all users depending on their respective symptoms.
- To distinguish population at risk using geographic positioning system (GPS).
- To secure the personal information of MERS-CoV-infected citizens to prevent mass panic paranoid.

To achieve these objectives, a mobile cloud system is proposed which stores personal as well as MERS-CoV's symptom-related information over the cloud. Initially stored information is processed using information granulation so that private data of users can be truncated and a unique case number is provided to each registered user. Using information granulation, the proposed system can achieve confidentiality of personal information related to patients. For all the future communications between users, physicians, and the government agencies, the unique case number provided by information granulation is used. Bayesian belief network (BBN) is used for initial diagnosis of users which will classify users as possibly infected or uninfected. All possibly infected users will be tracked by GPS from their respective mobile phones. Regions on the map are classified into different categories depending on the level of infection activity in that region. Google map assisted risk assessment, and re-routing is provided to citizens using this activity map. Using mobile application designed in

the proposed system, most of the users can check their MERS-CoV-infection status by submitting symptoms for the sake of curiosity or peer effect. It will assist the government to scan population of infected users very efficiently and efficaciously. All communications and information are stored over cloud storage which can be easily shared and analyzed using multiple data mining tools.

The rest of the paper is organized as follows. Section 2 provides work related to MERS-CoV infection and usage of cloud computing in biomedical applications. Section 3 proposes the system which is used to access and prevent MERS-CoV infections. Before concluding paper in Sect. 5, experimental results are presented, and performance analysis is conducted in Sect. 4.

2 Related work

Early studies conducted for MERS-CoV infection are for its source identification, transmission medium identification, and hazard appraisal. Integration of cloud computing for predicting and preventing MERS-CoV is still in its early days. Hence, related section is divided into two subsections which are MERS-CoV and biomedical informatics in the cloud. The former section explains some of the published works in MERS-CoV virus infection and vaccination, whereas the latter section describes the use of cloud computing in many biomedical applications by various researchers.

2.1 MERS-CoV

In 2013, Gautret [10] studied the effect of travel history on the outbreak of MERS-CoV infections in the world. There can be human-to-human transmission of MERS-CoV infection, so the risk of infection increases to Hajj pilgrims visiting the Middle East countries. According to him, the first case of MERS-CoV was identified from South Arabia. In 2013, Hon [11] discussed that pilgrims from various countries, who visited South Arabia to perform Hajj were identified as infected with respiratory viruses. Infected pilgrims increased the risk of epidemic not only in their home countries upon their return but also to other countries. To control the epidemic, Hajj pilgrims should be advised proper protective measures, and precautionary actions should be taken when they arrive back home. He suggested that it is necessary to know about travel pattern and health symptoms of Hajj pilgrims to control the epidemic. In 2013, Lu et al. [4] described the challenges in identifying the source, designing MERS vaccines, and developing anti-MERS therapy for MERS-CoV infection. However, they had not provided any proposal for identifying the source and to control the spread of MERS-CoV. In 2013, Omrani et al. [12] carried out research on identifying the transmission mediums used by MERS virus for its propagation such as body contact, air, or water. They found that the intermediate host of transmission of MERS-CoV infection can not be recognized or be symptomatic in most of the cases. To manage this significant problem, they suggested early recognition of MERS-CoV-infected patients and implementation of proper infection control measures. In 2014, Corman et al. [13] presented the effective use of RT-PCR kit for real-time detection of MERS-CoV infections in many patients. In 2014, Al-Tawfiq et al. [14] discussed sources of severe acute

respiratory syndrome coronavirus (SARS-COV) and MERS-CoV with their modes of transmission to humans. They suggested precautionary measures to travelers going to and returning from Middle East countries. All of the further mentioned articles related to MERS are from year 2015. Bhadra et al. [15] developed a real-time isothermal assay test for the detection of MERS virus. Kupferschmidt [16] and Soliman et al. [17] discussed the vaccinations and causes of MERS virus. They also provided first-aid and prevention measures for MERS. Yusof et al. [18] studied the number of camels infected with MERS and their locations in Saudi Arabia Emirate and the United Arab Emirates. Arabi et al. [19] demonstrated the effect of severe neurologic effect of MERS in hospitalized patients. Shalhoub [20] studied the effect of MERS when patients suffer from some chronic diseases. They studied the effect of MERS on a HIV-positive patient for 38 days. Similarly, Alghamdi [21] presented a case study of two MERS-infected users admitted in hospital. Banik et al. [22] argued about the knowledge gap of MERS in doctors and citizens. They also provided its infection rate as well as its difference from other similar viruses. Shehri [23] described the actions taken by health agencies in Saudi Arabia after MERS outbreak. He mentioned benefits and limitations of each action taken by government. Youde [24] listed the relationship between global health governance and MERS outbreak in detailed comparison. No further work has so far been presented for the detection and prevention of MERS-CoV to the best of our knowledge.

## 2.2 Biomedical informatics in the cloud

Cloud computing in healthcare area is still in its early days. Some of the notable contributions are discussed here. In 2010, Doukas et al. [25] demonstrated that mobile application and cloud computing together can enable better sharing, storing, updating, and retrieval of electronic healthcare data. A prototype has been implemented in the form of an android mobile application. The Amazon S3 cloud service was used for evaluating the developed system. In 2012, Rao and Kumar [26] developed predictive models for diagnosis of the Dengue fever and increased the probability of correct detection. Methodology used by them imputed missing values in the dataset using a novel imputation strategy and then generated a decision tree. Wrapper-based feature selection algorithm was proposed to extract the most important symptoms of dengue. In 2014, Thilakanathan et al. [27] demonstrated security protocol that allowed secure sharing of data in the cloud within the context of mobile health applications. They attempted to address the problems of achieving efficient user revocation especially while considering large data sizes. First, they defined a secure data-sharing model and protocol. Second, they demonstrated the feasibility of the protocol through their own developed prototype which combines smart phone, Bluetooth, and cloud computing technologies. Secure data-sharing architecture was adopted by adding a security layer that enables efficient and secure data sharing. In 2014, Kaur and Chana [28] proposed cloud intelligent system for diabetic patients. They used principal component analysis (PCA) for identifying minimum correlative variables from collected attribute variables. Naïve bayes and k-nearest neighbor classifiers were used to classify infected and uninfected users. In 2014, Ali et al. [29] studied a wireless lightweight, body-worn

sensor used in cloud health care for real-time and continuous physiological monitoring. They suggested a data authentication scheme to authenticate data generated by sensors, because they argued that sensors operate in lossy environment. In 2014, Sultan [30] explained the potential of cloud computing for improvement and advancement of life science research. However, they only provided cloud computing advantages in healthcare sector; no new work was proposed.

3 Proposed system

Figure 1 shows proposed system for predicting and preventing MERS-CoV epidemics. It comprises five components: data collection, information granulation, BBN-based initial analysis, GPS-based risk assessment, and communication & information sharing. It contains a cloud storage repository fragmented into multiple clouds known as medical record (M.R.) database. The data-collection component is responsible for compiling the raw information from users using the body-worn sensors and manually recorded data using the mobile application. BBN-based initial analysis classifies users into uninfected or possibly infected using symptoms registered by sensors and user. Personal information about users will be truncated, and a unique case number will be provided for further usage of the system. All compiled information and results generated are stored in a cloud storage repository termed as M.R. database. Cloud storage provides easy, flexible, and secure way to share information among users, doctors, hospitals, and the governmental agencies.

Figure 2 shows information and process flow for the proposed system. The system starts when a user registers using mobile application or website. After registration, a unique case number is provided by the system, and the data table is truncated using information granulation approach for protecting the identity of the user. BBN is applied to analyze information uploaded by the user to categorize him/her into uninfected and possibly infected user. Possibly infected user is the category where the initial analysis of the proposed system classified the user as infected but not confirmed. Further lab
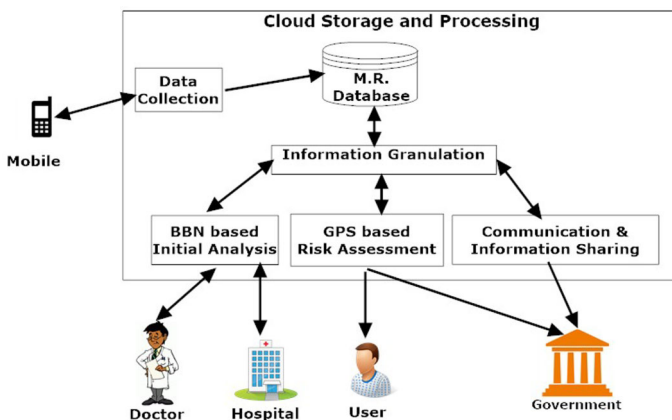
Fig. 1 Cloud system for managing MERS-CoV epidemics
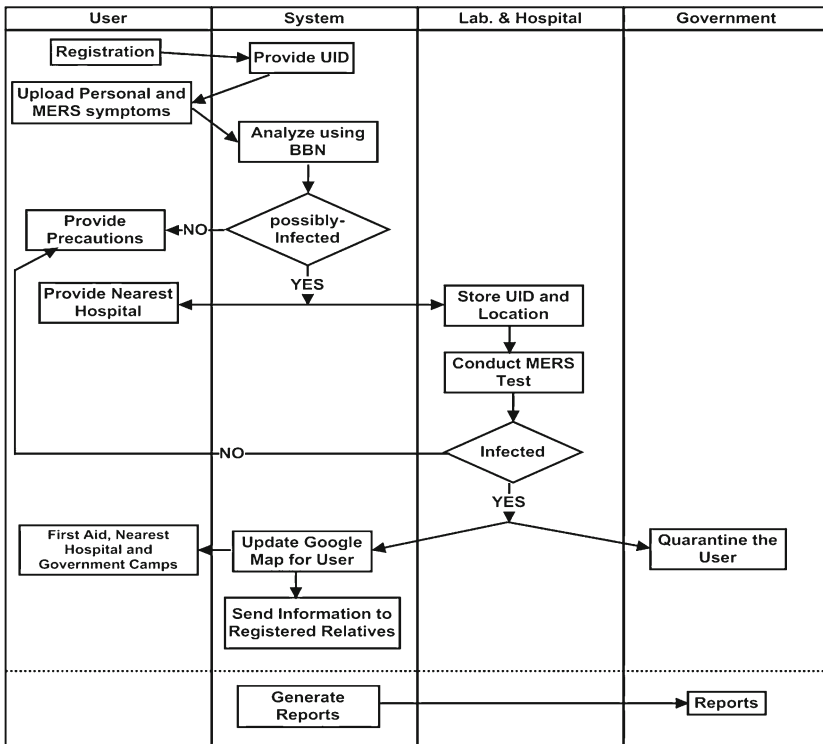
Fig. 2 System information flow for preventing MERS-CoV infections

tests are needed to confirm MERS-CoV infections. If BBN predicts user as possibly infected, the system provides information of the nearest hospital to the user and also shares his/her infection status information with multiple hospitals nearest to the user's geographic location. The hospital performs proper diagnostic test for MERS. If the user is found to be infected, he/she is quarantined at that time. An alert is being sent to the governmental agencies, hospital management treating MERS, close relatives, and Google maps will be updated accordingly. It is very difficult for the governments to approach each user and check all population. However, each citizen can enter his/her symptoms due to curiosity or general awareness. This will help the government to handle MERS more effectively by identifying users in their respective initial stage.

3.1 Data collection

Multiple health-related attributes are stored as shown in Table 2. These attributes are divided into personal- and MERS-CoV-related attributes. Personal attributes remain the same for most of the periods, whereas MERS-CoV attributes can change over time. The values of these parameters are stored in M.R database. BBN is used to categorize users depending on these attributes.

Table 2 Health attributes of user for MERS-CoV

S. no.	Attributes	Description
(a) Personal attributes		
1.	Name	Name of user
2.	Address	Permanent address of user
3.	Telephone numbers	Home/mobile telephone number
4.	Age	Age in years of user
5.	Sex	Male or female (M/F)
6.	Occupation	Occupation of user
7.	GPS	Geographic location of house and work place of user
8.	Relatives	Names and mobile numbers of user's close relatives

S. no.	Attributes	Output	Description
(b) MERS-CoV related attributes			
1.	Shortness of breath	Yes/no	Whether a user has shortness of breath symptom
2.	Cough	Yes/no	Whether a user has cough symptom
3.	Fever	Integer	Body temperature in °C
4.	Acute respiratory distress syndrome (ARDS)	Yes/no	Whether a user has ARDS
5.	Consumptive coagulopathy	Yes/no	Whether a user has consumptive coagulopathy symptom
6.	Food exposure	Yes/no	Whether a user consumed unprocessed, raw food, blood products, drinks, and used smoking apparatus such as hookah
7.	Animal exposure	Yes/no	Whether a user has exposure to animals (e.g., pets, rats, camels, etc), animal products
8.	Infected human exposure	Yes/no	Whether a user has contact with individuals with MERS-CoV symptoms
9.	Risk area exposure	Yes/no	Whether a user lives or works in risk area like Saudi Arabia

3.2 Information granulation

The data collection component receives personal, demographic, and MERS-CoV symptom-related information from the user. This information is sensitive and not

required to be shared with everyone. Disclosure of such information, even unintentionally, to unauthorized user can cause mass paranoid panic among citizens of any country. The proposed system implements secure sockets layer (SSL)-based encryption for all communication among different entities of the system. According to recent attacks on information systems, SSL is not adequate for such a vital information system. Attending this critical issue, proposed system uses information granulation [31] concept to prevent unauthorized access of data. Each data table is fragmented into three matrices of different security levels and stored on different secure servers. Even if anyone is able to retrieve APIs data matrix, he/she will not be able to extract exact personal information about the user. This section explains definitions and algorithms used to perform this procedure. Aside from the benefit of preserving the identity of patients, information granulation also helps the government and health authorities to generate results at a high degree of generalization.

An initial data table containing all information is fragmented into three separate data matrices of different security levels: Level 1 (personal information), Level 2 (demographic information), and Level 3 (MERS-CoV) information. Level 1 is highly sensitive information containing personal attributes such as name, social security number (SSN), address, and mobile number. Level 2 is the mediocre level of information containing demographic attributes such as gender, race, height, and ethnicity. Level 3 is least sensitive information containing MERS-CoV attributes and symptoms. Even if anyone could retrieve Level 3 information because it is maintained at the least secure system, he/she will not be able to find an exact identity of the user. For retrieving exact identity of user, a person requires the knowledge of all three levels of matrices.

To solve the problem of preserving the personal information of user, data should be first pre-processed from data tables to appropriate data matrices. Figure 3 shows the transition of data table to three matrices.

Definition 1 A data table $D^t = (N, A)$ can be defined such that N is a non empty set of individuals who have been registered over the MERS-CoV mobile application, and A is the set of attributes associated with each user of the data table. A mapping function

PI- Personal Information, DI- Demographic Information,
MI- MERS-CoV Information, Sym.- Symptoms
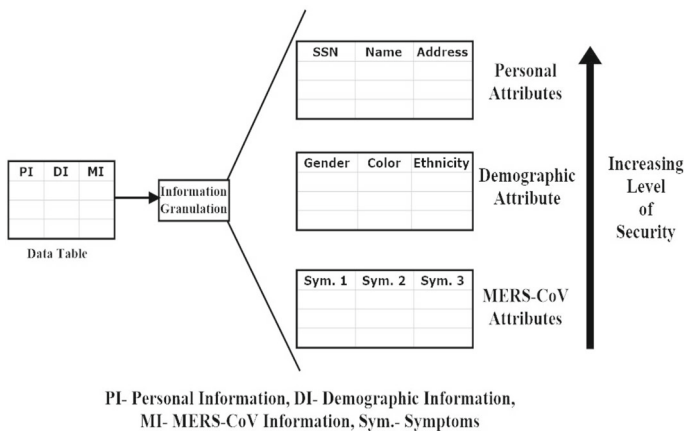
Fig. 3 Information granulation-based data table conversion

F is used to map the value of each attribute $a \in A$ to its specific values $v_a \in V_A$ such that $F : N \rightarrow V_A$ where V_A are the all possible values of attribute A.

Definition 2 Initial stored data table D^t is divided into three matrices of different security levels which are L_1, L_2, and L_3 such that $D^t = (N, L_1 \cup L_2 \cup L_3)$. These matrices are joined using a unique case number.

3.3 BBN-based initial analysis

Data stored in MERS-CoV information matrix are analyzed for initial analysis of the users. To classify users as uninfected or possibly infected, BBN [32] classifier is used. BBN classifier works on the principle of conditional probabilities and Naïve bayes classification. It classifies the user data into MERS-uninfected or possibly infected classes by comparing conditional probabilities. Any user X will be classified into uninfected class (U) and not into possibly infected (I) based on a symptoms S only if
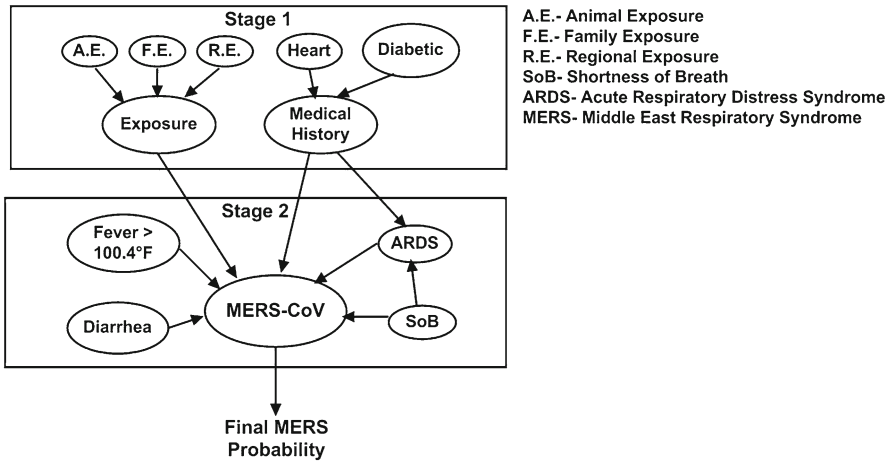
$$P(U|S) > P(I|S),$$

where $P(U|S)$ is the probability of any user X having an infection of MERS-CoV (U) if he/she has symptom (S) which can be derived using Bayes' theorem as follows:

$$P(U|S) = \frac{P(S|U)P(U)}{P(S)},$$

where $P(S|U)$ is the probability of having symptom (S) when any user is infected with MERS infection (U), $P(U)$ is the probability of having MERS infection based on different exposure levels, and $P(S)$ is the probability of having any symptom (S). Naïve bayes [33] classifier assumes conditional independence of all attributes, which is not possible in MERS-CoV because health parameters are always related to each other. A two-stage Bayesian belief network (BBN) is proposed to overcome the issue of conditional independence among MERS-CoV attributes as well as to consider most of the attributes for accurate prediction.

Figure 4 shows two-stage BBN for the proposed system. It contains two stages of BBN prediction based on direct dependency of the multiple attributes. For example, exposure level of any user to MERS-CoV virus depends on his/her animal exposure, family exposure, or regional exposure. Probability of exposure is predicted based on these three exposure attributes using the first stage of BBN classification. Similarly, medical history of any user may depend on multiple attributes such as heart, lungs, and diabetes, so predicting their effects on final MERS-CoV prediction is complex. The first stage of the proposed BBN predicts probability of effect of medical history taking into account all attributes. The second stage of the proposed BBN uses the output of the first stage of BBNs as well as some unique attributes to classify users into infected or possibly infected category, as shown in Fig. 4.

A.E.- Animal Exposure
F.E.- Family Exposure
R.E.- Regional Exposure
SoB- Shortness of Breath
ARDS- Acute Respiratory Distress Syndrome
MERS- Middle East Respiratory Syndrome

Fig. 4 Two-stage Bayesian belief network for detecting MERS-CoV-infected user

3.4 GPS-based risk assessment and re-routing for users

Risk assessment is utilized for continuous monitoring and management of MERS-CoV infections. The objective of risk assessment is to identify risk-prone areas, population at risk and control the epidemic. MERS-CoV is an airborne disease so up-to-date and relevant information of possibly infected users should be securely available to the government healthcare agencies. Location-based risk assessment and healthcare services will be highly beneficial for airborne diseases such as MERS-CoV. Location of possibly infected users can be effectively used to separate an area into different risk levels. If proper up-to-date information regarding risk-prone areas is available online, then people living in risk-prone areas can be effectively alerted in due time, and infection-control suggestions can be issued to them [19]. Figure 5 shows use of the GPS [34] for pinning hexagonal location of possibly infected users identified by the proposed system. However, pinpointing the exact location of the possibly infected user can cause mass paranoid panic. To effectively represent the spread of infection and risk-prone areas, hexagonal structure-based mapping is applied as indicated in Fig. 6. Total geographic area, possibly infected by MERS-CoV, is converted into non-overlapping hexagonal structures of the same size. Size of hexagonal structure can be changed as decided by the authority. A different color scheme is also proposed to effectively represent the different locations based on the level of infection in population density [36] of MERS-CoV possibly infected users. The chances of MERS-CoV infection are very high at the home and workplace, so these two places are used to define different risk-prone areas on the Google maps [35] web service.

Different coloring-based hexagonal mapping helps to recognize the flow of infection as well as extremely sensitive areas. Location-based representation of MERS-CoV infections over Google maps services help the governmental agencies as well as uninfected citizens to prevent an increase in the epidemic. Algorithm 1, Algorithm 2, and Table 3 are used to make hexagonal structure dynamic and adaptive in nature.
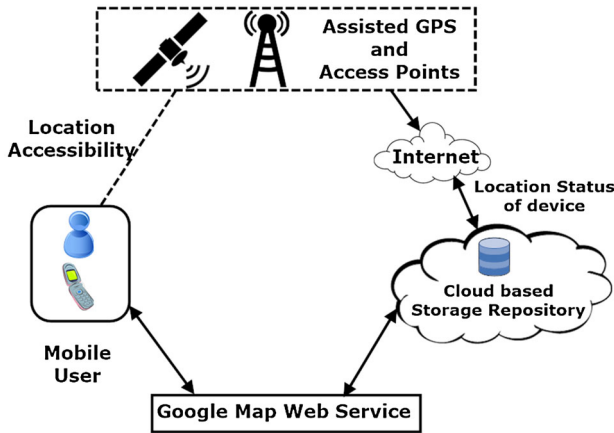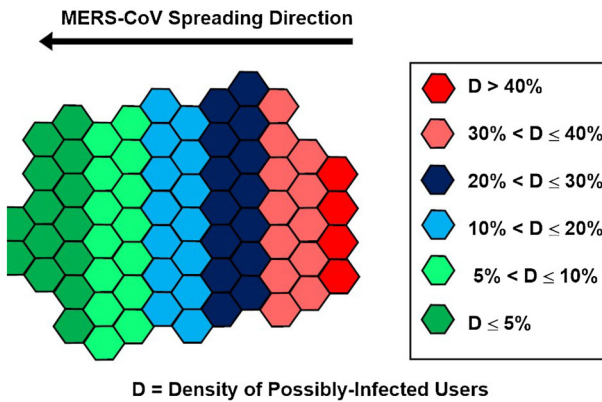
Fig. 5 Location-based healthcare service model

Fig. 6 Hexagonal representation of MERS-CoV-infected areas

Table 3 Different levels of MERS-CoV infection

Infected density (in %)	Risk level (color)
$D > 40$	Low (light green)
$30 < D \leq 40$	High (light red)
$20 < D \leq 30$	Medium (dark blue)
$10 < D \leq 20$	Medium (sky blue)
$5 < D \leq 10$	Low (parrot)
$D \leq 5$	Low (dark green)

Table 3 represents the coloring scheme on Google maps based on the density of possibly infected users in the selected hexagonal. The proposed system is automatically updated with time as new cases arrived and possibly infected users are tested. Algorithm 2 will be triggered by only authorized medical officers using highly secure

biometrics because Algorithm 2 uses data from Level 1 data matrix, explained in Sect. 3.2.

new[ ][ ][ ] is three-dimensional matrix which stores user id (UID), latitude, and altitude values of infection hexagonal of each new possibly infected user's home and workplace.

stored[ ][ ][ ] is three-dimensional matrix which stores UID, latitude, and altitude values of infection hexagonal of all pinned infected user's home and workplace.

Algorithm 1: Mapping risk-prone areas on google maps

Step I: Identify location of home and work hexagonals of the newly possibly infected user.
Step II: Store the newly identified hexagonal in three-dimensional matrix new[ ][ ][ ].
Step III: for every hexagonal in new[ ][ ][ ]
       Step III.1: Find total number of population and possibly infected users in hexagonal.
       Step III.2: Increase the density factor of hexagonal.
       Step III.2.1: Update the color of hexagonal based on calculated density of hexagonal, represented in Table 3.
Step IV: Move the user hexagonal location to the matrix stored[ ][ ][ ].
Step V: Exit

Algorithm 2: Updating hexagonal mapping after possibly infected user is medically tested for MERS-CoV

Step I: Select location of possibly infected user from highly secure level 1 data matrix.
Step II: Find the user and perform blood as well as cough test.
Step III: If blood test is positive,
       Step III.1: Quarantine the user from society.
       Step III.2: Check all family members and quarantine if anyone is found positive.
       Step III.3: Update the Level 1 data matrix.
Step IV: Else
       Step IV.1: Update user status to uninfected.
Step V: Update the color of user's hexagonal based on calculated density of hexagonal, represented in Table 3.
Step VI: Exit

Proposed map technique acts as an effective platform to represent the spread of MERS-CoV infection and population at risk to the governmental agencies, healthcare departments, and general population. Different colors represent different intensities of MERS-CoV infection in any particular area.

3.5 Communication and information sharing

Information regarding patients, medicines, and reports of each user is stored in the database on the cloud known as M.R database. It can be accessed securely by hospitals, concerned doctors, and various governmental agencies using SSL and secure shell (SSH) technologies [37]. SSL is a security protocol, which encrypts the data flowing

from server to user, and vice versa. SSL certificate will be provided for each website that uses specified technology to send or receive data. However, SSH provides a secure remote connection with the system. Both SSL and SSH will be used in the proposed system to encrypt data as well to provide a remote server connections. Cloud-based information sharing provides easy and effective communication so that first aid will be provided to user by experienced physicians, and patient does not need to provide again and again his/her medical record during his/her visit to different hospitals. The latest information about the disease can be shared with people residing in different geographic areas very easily. The time needed to provide service is effectively reduced and better services can be provided to the user using effective cloud information sharing methods like Facebook, Twitter, and instant messaging.

4 Experimental setup and performance analysis

Even after an in-depth search on the Internet and email correspondences with some of the doctors of Middle East countries, we were not able to receive any MERS-infection database to test the proposed system. For experimental setup and performance analysis of the proposed system, synthetic data are generated in collaboration with Dr. Pankaj Sood. He is an infectious disease specialist. Experimental setup of the proposed system is divided into four parts for proper understanding:

- Creation of synthetic data.
- Training and testing of BBN.
- Performance analysis of the proposed system on Amazon EC2 cloud.
- Google map-based risk assessment.

Table 4 Sample combination of MERS-CoV infections

S. no	Diarrhea	Fever	SoB	A.E.	F.E.	R.E.	Heart	Diabetic	Lung	MERS
1	Y	Y	N	N	N	N	Y	N	Y	Y
2	Y	Y	Y	N	N	N	Y	Y	Y	Y
3	N	Y	Y	Y	Y	N	Y	Y	Y	Y
4	N	Y	N	N	N	N	Y	N	N	N
5	N	Y	Y	N	N	N	N	N	Y	Y
6	Y	Y	N	N	N	N	N	Y	N	N
7	N	Y	N	N	N	Y	N	N	N	N
8	N	N	Y	N	N	N	N	N	N	N
9	Y	N	Y	N	N	N	N	Y	N	N
10	N	N	N	N	N	N	N	N	N	N

SoB shortness of breath, *A.E.* animal exposure, *F.E.* family exposure, *R.E.* regional exposure, *Y* yes, *N* no

Table 5 Division of all possible MERS-CoV combinations

Category	Probability	Description	Count
A	$p > 0.40$	Most possible combination	297
B	$0.25 < p \leq 0.40$	Possible combination	459
C	$0.10 < p \leq 0.25$	Rarely possible combination	172
D	$p \leq 0.10$	Very rare possible combination	96

**Table 6** Attributes used for personal and demographic data

S. no.	Attribute	Category	S. no.	Attribute	Category
1	Age	P	11	Family position	P
2	Class of work	D	12	Migration	P
3	Education	P	13	No. of family member	P
4	Wage per week	P	14	Birth country of father	D
5	Marital status	P	15	Birth country of mother	D
6	Occupation industry	D	16	Self birth country	D
7	Occupation field	D	17	Citizenship	D
8	Race	D	18	Social Security Number	P
9	Ethnicity	D	19	Name	P
10	Gender	D	20	Address	P

P personal information, *D* demographic information

4.1 Creation of synthetic data

Generation of synthetic data for testing of the proposed system is done systematically by consulting with Dr. Pankaj Sood. The proposed system uses ten different MERS-CoV attributes to predict whether any user is infected or not with each attribute's output as either Yes or No. A sample of 10 such combinations is listed in Table 4. Hence, the total number of permutations of these cases will result in $2^{10} = 1024$ number of combinations. All health attributes are related to each other so possibility of occurrence of any one combination is different from another. Hence, all combinations are divided into four categories based on their probability of occurrence, as listed in Table 5.

Personal and demographic information of around 0.2 million citizens is obtained from census data [38] which is integrated with combinations of MERS-CoV. Data used from census has 20 attributes which are listed in Table 6. These attributes are divided into personal and demographic attributes. Figure 7 shows the distribution of age and race in census data used in the creation of synthetic MERS-CoV data. Age is distributed using normal curve with mean of 34.49 and race has almost all types of cases. All possible combinations of MERS-CoV attributes generated are mapped randomly to these 0.2 million user data based on probabilities as shown in Table 5. Figure 8 shows the procedure followed to generate synthetic data of 0.2 million users with MERS-CoV attributes randomly distributed among them.
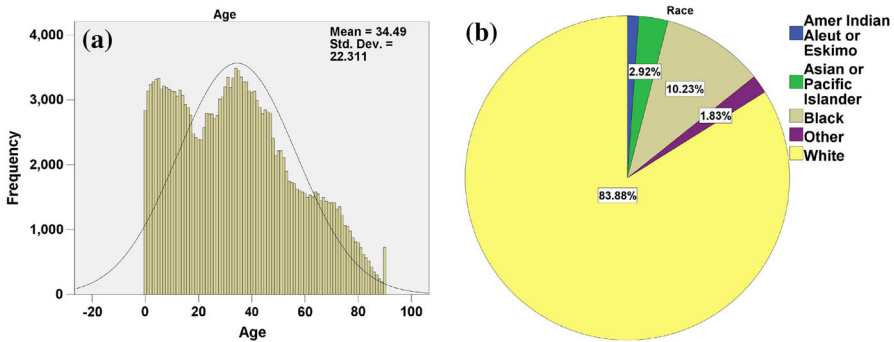
Fig. 7 Distribution of **a** age and **b** race attribute in synthetic data collected from the census

Fig. 8 Procedure followed to create synthetic MERS-CoV data of 0.2 million users
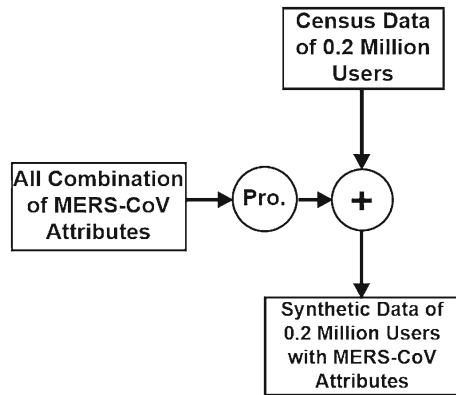
Table 7 Experiment results of different algorithms of "bnlearn" package in R studio

Algorithm	Independence tests	Learned arcs	Execution time (s)
gs	1548	26	9.023
iamb	2349	23	11.457
fast.iamb	1785	26	9.984
inter.iamb	2413	23	12.547
hc	2294	33	20.415

4.2 Training and testing of BBN

Synthetic data generated are used to train and test the BBN which will be used for classification of users into uninfected or possibly infected users. To develop a Bayesian network, data of 5000 users are used in R Studio using "bnlearn" package. Table 7 shows the results of different learning algorithms. This package is run on Amazon EC2 virtual machine with four virtual CPUs and 2 GB of RAM. As the results show, "gs" and "fast.iamb" suited the proposed system better because of their low execution time and more learned arcs. For further experiments, "fast.iamb" has been used in this

paper. Trained BBN is tested in Weka 3.7 [39] for various statistical measures. BBN learned is divided into two separate BBNs which are the first stage and the second stage, as explained in Sect. 3.3. Experimental evaluation is done for three BBNs as follows:

- The first stage which calculates probability of exposure and medical history.
- The second stage which calculates MERS-CoV when the first-stage probabilities are statistically provided.
- Complete BBN which calculates MERS-CoV-infection probability when the first- and the second-stage BBNs work collaboratively.

Summary of the models (the first stage, the second stage, and the complete BBN) tested in Weka 3.7 is shown in Table 8 and Table 9. All BBNs classify the users with an accuracy of more than 80 % in all the stages of the proposed BBN. The detailed accuracy of each class parameter as classified by BBN are also listed in Table 9. True-positive (TP) rates—also known as sensitivity and false-positive (FP) rates, also known as specificity—are statistical measure to rate quality of any classification algorithm. TP rate defines the percentage of possibly infected MERS instances which are correctly classified by classification algorithm. FP rate defines the percentage of uninfected MERS instances which classification algorithm classifies as uninfected [40]. A classification algorithm with high TP and FP rate is desirable. The proposed classification algorithm provides a high TP rate that is 0.800, and a high FP rate that is 0.880. Precision and recall provide the relevancy of the classified instances. The higher the value of precision and recall, the better will be the classification because only relevant instances will be classified [40]. The proposed algorithm provides very high precision and recall rate values which are 0.809 and 0.800, respectively. F-Measure is a statistical value for classification accuracy which lies between 0 and 1. Algorithm with a high F-Measure value will be more accurate, and the proposed two-stage BBN provides an F-Measure of 0.826. Matthews correlation coefficient (MCC) is a statistical measure used to find quality of any binary classification. As the proposed algorithm is for binary classification, so MCC measure is very important for it. It lies between -1 and $+1$, and the higher the value, the better is the classification. The proposed algorithm provides high value of MCC which is 0.720. Similar is the case with receiver-operating characteristic (ROC); high value of ROC area represents the accuracy of the proposed algorithm [40]. With the level of accuracy conceived in the above-listed statistical terms, it justifies the use of BBN for the proposed system.

4.3 Information granulation

Data table developed using the synthetic data creation procedure shown in Fig. 8 is converted to three separate matrices containing only the desired information. Table 10 shows the snapshot of three separate matrices for eight attributes and seven users. All attributes are not shown here due to space constraint.

Table 10 shows the snapshots of three matrices obtained after information granulation step. Any level matrix is obtained by Cartesian product of available matrices. Table 10b matrix is generated by Cartesian product of MERS-CoV attribute matrix and Demographic information matrix. Table 10a, b, respectively, has star markings

Table 8 Summary of tenfold cross validation of BBN in Weka 3.7

	Second-stage BBN	First-stage BBN		Comp. BBN
		Exp.	Dis.	
Correctly classified instances	43,509 (87 %)	45,510 (91.0 %)	45,486 (90.0 %)	41,578 (83.1 %)
Incorrectly classified instances	6491 (13 %)	4490 (9 %)	4514 (10 %)	8422 (16.8 %)
Kappa statistic	0.8375	0.904	0.89	0.8214
Mean absolute error	0.3109	0.247	0.23	0.3310
Coverage of cases (0.95 level) (%)	100	100	100	100
Mean rel. region size (0.95 level) (%)	91.12	93.14	93.47	90.84
Total number of instances	50,000	50,000	50,000	50,000

Table 9 Detailed accuracy by stage for BBN in Weka 3.7

| Model	TP rate	FP rate	Precision	Recall	F-measure	MCC	ROC area
First stage	0.884	0.920	0.894	0.890	0.856	0.752	0.984
Second stage	0.851	0.900	0.861	0.860	0.837	0.735	0.980
Complete	0.800	0.880	0.809	0.800	0.826	0.720	0.970

among some of attributes because these attributes are not available while calculating Cartesian product due to security constraints. This procedure will help to keep the identity of user confidential, whereas his/her MERS-CoV data can be easily obtained using the case number.

4.4 Performance analysis

MERS-CoV application user data are stored over shared cloud storage provided by Amazon. Storage-optimized i2. xlarge [41] cluster is used to deploy the proposed system over the cloud. It can be accessed by doctors, users, healthcare departments, and the governmental agencies. 0.2 million user data are used to study the performances of classification algorithms and the Amazon cloud. Different classification algorithms such as k-nearest neighbor [42], linear regression [43], and neural network [44] are also implemented to compare them with the proposed two-stage BBN, so that its use can be experimentally justified. One of the vital tasks of the proposed system is the classification of users into infected and possibly infected, so it requires a high performance level. Figure 9 shows the comparison of various statistical measures of used classification algorithms on different size of user data. Figure 9a represents the accuracy of the classification algorithms. The proposed two-stage BBN performed better than all algorithms. However, Neural Networks are also close to BBN in classifica-

Table 10 Snapshot of (a) level 3 data matrix containing only MERS-CoV attributes (b) data matrix containing demographic information and MERS attributes (c) data matrix containing all information

Case no.	SSN	Address	Race	Gender	Occupation	Fever	Respiratory	Exposure
(a)								
1001	****	****	****	****	****	Yes	Yes	Yes
1002	****	****	****	****	****	No	Yes	No
1003	****	****	****	****	****	Yes	Yes	No
1004	****	****	****	****	****	No	No	Yes
1005	****	****	****	****	****	No	Yes	No
1006	****	****	****	****	****	Yes	No	Yes
1007	****	****	****	****	****	Yes	No	No
(b)								
1001	****	****	White	Male	Precision production craft and repair	Yes	Yes	Yes
1002	****	****	Asian	Female	Professional specialty	No	Yes	No
1003	****	****	Black	Male	Not working	Yes	Yes	No
1004	****	****	White	Male	Not working	No	No	Yes
1005	****	****	Other	Female	Machine operators assmblrs and inspctrs	No	Yes	No
1006	****	****	American Indian	Male	Adm support including clerical	Yes	No	Yes
1007	****	****	Black	Male	Not working	Yes	No	No
(c)								
1001	346713859	241219	White	Male	Precision production craft and repair	Yes	Yes	Yes
1002	761482438	241206	Asian	Female	Professional specialty	No	Yes	No
1003	347168492	241213	Black	Male	Not working	Yes	Yes	No
1004	768143983	241219	White	Male	Not working	No	No	Yes
1005	248137951	241207	Other	Female	Machine operators assmblrs and inspctrs	No	Yes	No
1006	875612458	241219	American Indian	Male	Adm support including clerical	Yes	No	Yes
1007	498446348	241206	Black	Male	Not working	Yes	No	No
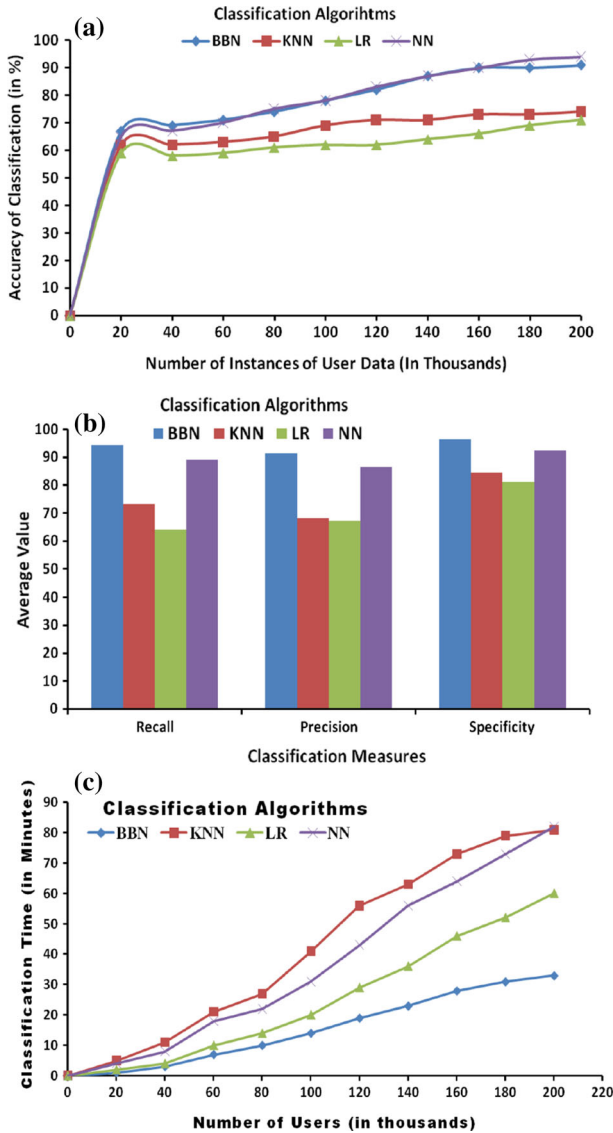
Fig. 9 Performance analysis of classification tool over the Amazon cloud, **a** accuracy of classification, **b** classification measures of algorithms, **c** classification time of algorithms

tion accuracy, but they have high classification time as shown in Fig. 9c. Figure 9b shows three classification measures: Recall, Precision, and Specificity, as discussed in Sect. 4.2. The proposed two-stage BBN performed better than any other comparable algorithms in all statical measures which justified its use in the proposed system.
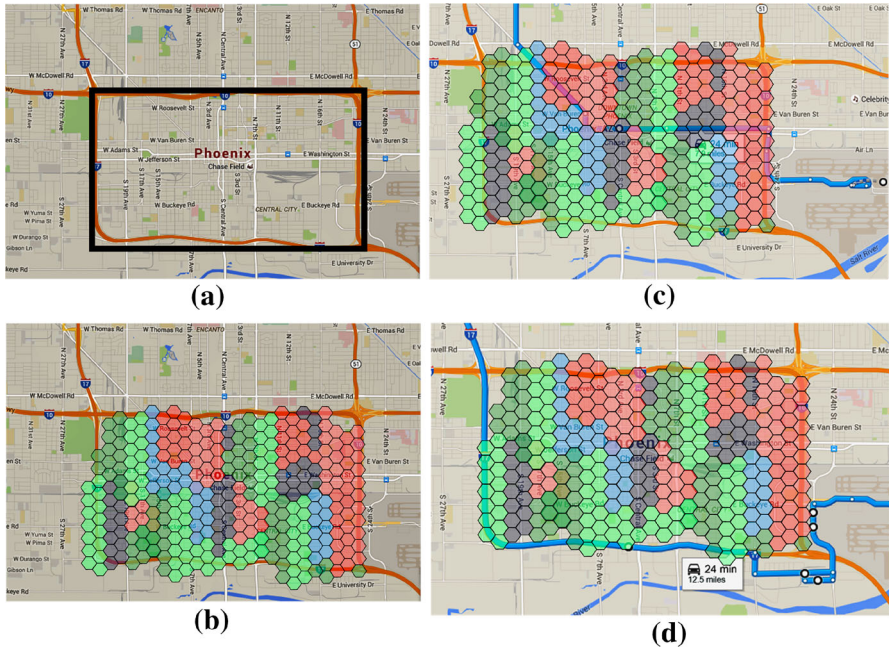
Fig. 10 GPS-based re-routing of user **a** targeted location that is central city block of Phoenix city, **b** hexagonal based mapping of color base risk areas, **c** default routing of user using Google services, **d** new routing of user based on infected areas

4.5 Risk assessment

To study the geographic-based risk assessment, special cases are generated over Phoenix city. Data of fifty thousand users were registered for Phoenix central city block as shown in Figure 10a. GPS location of possibly infected user is also pinned on the Google map. Using the generated data and color mapping algorithm, a net of hexagonal structures is mapped on this particular block as shown in Fig. 10b. Figure 10 c shows the routing of a user form Phoenix airport to Metro High Tech School without using any re-routing algorithm. The blue line indicates the proposed route by Google services. In this case, the user passes through the red zone area for MERS-CoV-infected user. However, when appropriate re-routing is used, as shown in Fig. 10d, the user has been re-routed to the safer side of the block.

4.6 Discussion of results

The proposed system is complex and constituents of multiple components which makes its experimental evaluation even more complex. Thus, experimental evaluation had systematically carried on and is broadly divided into four parts. No availability of data for MERS forced us to create a synthetic data that are generated for 0.2 million users and explained in detail in Sect. 4.1. Then, the training of the proposed two-stage BBN using synthetic data in Sect. 4.1 is done in Sect. 4.2. All the experimental

generated statistical measures justified the use of the proposed two-stage BBN, as listed in Tables 7, 8, and 9. Data table fragmentation process divided into three separate data matrices is also experimentally verified in Sect. 4.3. Table 10 provides the snapshot of data table and three data matrices which indicated the successful conversion. In Sect. 4.4, whole classification process is tested for 0.2 million users on Amazon EC2 cloud using multiple classification algorithms. The proposed two-stage BBN performed better in all cases than its fellow competitors. However, NN is also close to BBN in classification accuracy, but its classification time is much larger than the proposed two-stage BBN. Last, the system is tested for re-routing of citizens, which is successfully completed as shown in Fig. 10.

5 Conclusion

Infectious transmitted diseases using common mediums are one of the major concerns of any nation's government and healthcare departments. With the advancements in information technologies, it is possible to control many infections in the most effective and efficient way. In this paper, a system is proposed for predicting and preventing an airborne disease known as MERS-CoV using cloud computing and GPS. Bayesian belief network (BBN) is utilized to classify users, and cloud computing is used for effective information analysis and sharing. The key point of the paper is the use of geographic positioning system to represent each MERS-CoV possibly infected users on Google maps and provide re-routing to uninfected users. Proposed system implemented on Amazon EC2 cloud provides 80 % accuracy in classification and appropriate GPS re-routing. It will help the citizens to avoid regional exposure and the government authorities to manage the problem more effectively. Future study will include the use of the internet of things (IoT)-based smart devices for more accuracy and effectiveness in controlling infectious epidemic.

References

1. de Groot RJ, Baker SC, Baric RS, Brown CS, Drosten C, Enjuanes L, Ziebuhr J (2013) Middle East respiratory syndrome coronavirus (MERS-CoV): announcement of the Coronavirus Study Group. J Virol 87(14):7790–7792
2. MERS Outbreak in South Korea (2015). http://www.healthmap.org/site/diseasedaily/article/mers-outbreak-south-korea-6715. Last Accessed on 14th June
3. WHO Guidelines for Investigation of Cases of Human Infection with Middle East Respiratory Syndrome Coronavirus (MERS-CoV) (2013). http://www.who.int/csr/disease/coronavirus_infections/MERS_CoV_investigation_guideline_Jul13.pdf. Last Accessed on 5th June 2015
4. Lu L, Liu Q, Du L, Jiang S (2013) Middle East respiratory syndrome coronavirus (MERS-CoV): challenges in identifying its source and controlling its spread. Microbes Infect 15(8):625–629
5. Fighting Infectious Disease with Cloud Computing (2015). http://www.infectioncontroltoday.com/news/2014/09/fighting-infectious-disease-with-cloud-computing.aspx. Last Accessed on 29th May 2015

6. Sood SK, Sandhu R (2014) Matrix based proactive resource provisioning in mobile cloud environment. Simul Modell Pract Theory 50:83–95

7. Park SC, Ryoo SY (2013) An empirical investigation of end-users' switching toward cloud computing: a two factor theory perspective. Comput Hum Behav 29(1):160–170

8. Biswas S, Akhter T, Kaiser MS, Mamun SA (2014) Cloud based healthcare application architecture and electronic medical record mining: an integrated approach to improve healthcare system. In: 17th international conference on computer and information technology (ICCIT). IEEE, pp 286–291

9. Javadi B, Abawajy J, Buyya R (2012) Failure-aware resource provisioning for hybrid Cloud infrastructure. J Parallel Distribut Comput 72(10):1318–1331

10. Gautret P (2013) Middle East Respiratory syndrome (MERS) coronavirus. What travel health advice should be given to Hajj pilgrims? Travel Med Infect Dis 11(5):263–265

11. Hon KL (2013) Severe respiratory syndromes: travel history matters. Travel Med Infect Dis 11(5):285–287

12. Omrani AS, Matin MA, Haddad Q, Al-Nakhli D, Memish ZA, Albarrak AM (2013) A family cluster of middle east respiratory syndrome coronavirus infections related to a likely unrecognized asymptomatic or mild case. Int J Infect Dis 17(9):e668–e672

13. Corman VM, Ölschläger S, Wendtner C-M, Drexler JF, Hess M, Drosten C (2014) Performance and clinical validation of the RealStar® MERS-CoV Kit for detection of Middle East respiratory syndrome coronavirus RNA. J Clin Virol 60(2):168–171

14. Al-Tawfiq JA, Zumla A, Memish ZA (2014) Travel implications of emerging coronaviruses: SARS and MERS-CoV. Travel Med Infect Dis 12(5):422–428

15. Bhadra S, Jiang YS, Kumar MR, Johnson RF, Hensley LE, Ellington AD (2015) Real-time sequence-validated loop-mediated isothermal amplification assays for detection of Middle East Respiratory Syndrome Coronavirus (MERS-CoV). PLoS ONE 10(4):1–21

16. Kupferschmidt K (2015) MERS surges again, but pandemic jitters ease: the camel virus appears to be less lethal and more prevalent in humans than assumed. Science 347(6228):1296–1297

17. Soliman T, Cook AR, Coker RJ (2015) Pilgrims and MERS-CoV: What's the risk? Emerg Themes Epidemiol, 12(1):1–3

18. Yusof MF, Eltahir YM, Serhan WS, Hashem FM, Elsayed EA, Marzoug BA, Abdelazim AS, Bensalah OKA, Al Muhairi SS (2015) Prevalence of Middle East Respiratory Syndrome Coronavirus (MERS-CoV) in dromedary camels in Abu Dhabi Emirate United Arab Emirates. Virus Genes 50(3):509–513

19. Arabi YM, Harthi A, Hussein J, Bouchama A, Johani S, Hajeer AH, Saeed BT, Wahbi A, Saedy A, AlDabbagh T, Okaili R, Sadat M, Balkhy H (2015) Severe neurologic syndrome associated with Middle East Respiratory Syndrome Coronavirus (MERS-CoV). Infection. doi:10.1007/s15010-015-0720-y

20. Shalhoub S, AlZahrani A, Simhairi R, Mushtaq A (2015) Successful recovery of MERS CoV pneumonia in a patient with acquired immunodeficiency syndrome: a case report. J Clin Virol 62:69–71

21. Alghamdi M, Mushtaq F, Awn N, Shalhoub S (2015) MERS-CoV infection in two renal transplant recipients: case report. Am J Transplant 15(4):1101–1104

22. Banik GR, Khandaker G, Rashid H (2015) Middle East Respiratory Syndrome coronavirus "MERS-CoV": Current Knowledge Gaps. Paediatr Respir Rev. doi:10.1016/j.prrv.2015.04.002

23. Al Shehri AM (2015) A lesson learned from Middle East respiratory syndrome (MERS) in Saudi Arabia. Med Teach 37(S1):S88–S93

24. Youde J (2015) MERS and global health governance. Int J 70(1):119–136

25. Doukas C, Pliakas T, Maglogiannis I (2010) Mobile healthcare information management utilizing Cloud Computing and Android OS. In: Annual international conference of the IEEE engineering in medicine and biology society, EMBC'10, pp 1037–1040

26. Rao VSH, Kumar MN (2012) A new intelligence-based approach for computer-aided diagnosis of dengue fever. IEEE Trans Inf Technol Biomed 16(1):112–118

27. Thilakanathan D, Chen S, Nepal S, Calvo R, Alem L (2014) A platform for secure monitoring and sharing of generic health data in the Cloud. Future Gener Comput Syst 35:102–113

28. Kaur PD, Chana I (2014) Cloud based intelligent system for delivering health care as a service. Comput Methods Programs Biomed 113(1):346–359

29. Ali ST, Sivaraman V, Ostry D (2014) Authentication of lossy data in body-sensor networks for cloud-based healthcare monitoring. Future Gener Comput Syst 35:80–90

30. Sultan N (2014) Discovering the potential of cloud computing in accelerating the search for curing serious illnesses. Int J Inf Manag 34(2):221–225

31. Yao Y (2007) The art of granular computing. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 4585 LNAI, pp 101–112
32. Liu KFR, Lu C-F (2009) BBN-based decision support for health risk analysis. In: 5th international joint conference on INC, IMS, and IDC, pp 696–702
33. Tütüncü GY, Kayaalp N (2015) An aggregated fuzzy naive bayes data classifier. J Comput Appl Math 286:17–27
34. Taylor MA, Woolley JE, Zito R (2000) Integration of the global positioning system and geographical information systems for traffic congestion studies. Trans Res Part C Emerg Technol 8(1):257–285
35. Google Maps (2015). https://www.google.co.in/maps. Last Accessed on 10th June 2015
36. Calhoun JB (1962) Population density and social pathology. Scientific American, New York
37. Metke AR, Ekl RL (2010) Security technology for smart grid networks. IEEE Trans Smart Grid 1(1):99–107
38. Adult Dataset (2015). https://archive.ics.uci.edu/ml/datasets/Adult. Last Accessed on 10th June (2015)
39. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH (2009) The WEKA data mining software: an update. ACM SIGKDD Explor Newslett 11(1):10–18
40. Baldi P, Brunak S, Chauvin Y, Andersen CA, Nielsen H (2000) Assessing the accuracy of prediction algorithms for classification: an overview. Bioinformatics 16(5):412–424
41. Amazon EC2 Instance Types (2014). http://aws.amazon.com/ec2/instance-types/. Last Accessed on 2nd September, 2014
42. Peterson LE (2009) K-nearest neighbor. Scholarpedia 4(2):1883
43. Bates DM, Watts DG (1988) Nonlinear regression: iterative estimation and linear approximations. Wiley, New York
44. Hagan MT, Demuth HB, Beale MH (1996) Neural network design. Pws Pub, Boston