MedCloud: Cloud-Based Disease Surveillance and Information Management System

ADNAN BASHIR [1,3], **ASAD WAQAR MALIK** [1], **ANIS UR RAHMAN** [1],
SOHAIL IQBAL [1], **PAUL ROBERT CLEARY** [2], **AND AAMER IKRAM** [3]

[1] School of Electrical Engineering and Computer Science, National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan
[2] Public Health England, Liverpool L3 1DS 13088, U.K.
[3] National Institute of Health, Islamabad 44000, Pakistan

Corresponding author: Adnan Bashir (adnan@nih.org.pk)

ABSTRACT Outbreaks can overwhelm fragile health systems that lack the tools, infrastructure, policies, and systems to keep communities healthy and safe. Timely detection, preparedness, and appropriate response are essential for limiting both the loss of human life and social economic disaster due to disease outbreaks. Countries must build effective and sustainable disease surveillance and reporting systems that mobilize all levels of the health system–including communities for crisis response. In Pakistan currently, there is no centralize Health Information and Diseases Surveillance System. The current disease surveillance system is a completely manual system, where diseases are being reported through hard copies or print and electronic media. Because of the delayed process, the outbreaks are usually heard through print and electronic media. The ineffective reporting system not only creates problems in managing the countermeasures against a disease outbreak but is likely to cause mass hysteria among the people. In this paper, we propose MedCloud, a cloud-based health management system for disease surveillance and early warning with trend analysis. The proposed system provides a nationwide connected disease surveillance system with the provision of statistical data processing, data validation, and secure backup features. MedCloud is flexibly designed to easily incorporate new plugins. It uses cloud virtual machines to dynamically handle users sessions; thus, can easily and efficiently be accessed using the Internet.

INDEX TERMS Disease surveillance, health management, early warning system, cloud-based, disease outbreaks.

I. INTRODUCTION

Countries around the world areupgrading their public management systems using information and communication technologies (ICTs) in order to streamline workflows. The introduction of ICTs in the public sector has led to the introduction of a new term called as e-government. With e-government, it is expected to not only enhance good management and response but also access to cost-effective, dynamic and user-friendly solutions. The solutions provide a hassle-free data transfer from geographically distributed locations to central government institutions.

One of the most crucial and demanding public sector entity is health management. Moreover, health is one of the top

The associate editor coordinating the review of this manuscript and approving it for publication was Xiang Zhao .

challenges outlined in sustainable development goals (SDGs). The stakeholders of the health sector include the public and government. Thus, a delay in actions can directly affect the public [1], [2]. Unfortunately, in most of the countries, there is no proper health information management system (HIMS). Without such a system, manual operations are performed when information is shared via email or through messengers. However, such systems fail to predict or identify outbreaks. In most cases, outbreaks are identified after a few days [3].

In developing countries like Pakistan, India, and Bangladesh, there is no HIMS system available to support health-related issues. Over the years, in Pakistan, a number of health information systems are deployed under different public initiatives. The systems work in isolation, each focusing on a specific activity or for disease monitoring, with no

mechanisms to detect outbreaks and share information at different levels to spread awareness. In summary, there is no system providing regular, timely, and valid information on disease outbreaks and other health-related issues. The information systems are functioning with varying degree of success as surveillance systems. The systems are gathering information at the first level of care facilities, but with the non-existence of a systematic and continuous data reporting and data sharing the systems become ineffective. Moreover, the systems vary from place-to-place and level-to-level in terms of functionality, data collection, data format, reporting and record keeping methods. Overall, existing systems are highly fragmented, often vertical leading to duplication of efforts and unbalanced resource distribution. A system related to disease prevention can benefit from greater information integration and resource sharing at all levels.

In Pakistan, at present, the existing system is based on DHIS[1] system. The system was developed in the early 90s and lacks features for data dissemination and analytics to predict an outbreak. The system is standalone working at the town level, with no integration with other systems installed at neighboring locations.

In this paper, we proposed a geographically distributed cloud-based system referred to as MedCloud. The objective is to overcome the issues and challenges related to an early version of HMIS. The proposed system is highly scalable compared to the older health management system. The main feature of MedCloud is its web-based interface that provides an online health surveillance system. Users at the district (or health facility) level can access weekly surveillance data through their desktop systems, laptops, or using any handheld devices. Moreover, the visualization module allows its users to view reports, charts, maps on any device. Furthermore, the interface supports local regional languages to facilitate locals. The learning curve of the system is gentle compared to traditional DHIS system. The system is robust handling connectivity interruptions through offline data synchronization module, that supports offline analysis. The system provide a range of validation functionality to ensure completeness and quality of recorded data. The users can also add epidemiological interpretations to their analyses. Thus, with flexible design, MedCloud is easy to adopt at places with low literacy rates.

The next section explains the requirements of ideal disease surveillance system. Section III covers the literature review. Section IV explains the proposed MedCloud framework. The results and related discussion are covered in Section V. Finally, the conclusion is presented in Section VI.

II. REQUIREMENTS OF IDEAL DISEASE SURVEILLANCE SYSTEM

Disease surveillance systems play an important role in early identification, monitoring and elimination of infectious diseases. It is a critical component of any health manage-

---

[1]DHIS: District Health Management System

ment system. A typical surveillance system is a data-driven model where data is entered at dispersed locations. This data drives action within a system and also generate alerts for remedial actions. However, the data must be accurate, good in quality and entered timely to use the system effectively.

Disease surveillance and response for communicable diseases aim to estimate the burden of such diseases in the country and provides means to inform all stakeholders. An ideal integrated disease surveillance and response system should collect and transmit data in real time to all the stakeholders. It should be able to incorporate data from existing surveillance systems in real time and analyze data to devise rapid response strategies. Some of the key features of a disease surveillance system along with objectives are explained as follows:

A. RAPID REPORTING

The reporting system must be designed in such a way that data can flow in both directions and can be entered at any level. For data entry, proper data validation checks need to be installed to avoid human errors. For quick data entry, the form should be filled up using item selections. Thus, requiring less typing effort. In most of the cases, field workers and support staff are unfamiliar with medical terms, therefore, a user-friendly interface can facilitate submission of complete and accurate reports.

Complete and timely reporting is an essential element of a surveillance system. The system should adopt rapid data capturing procedures that are locally appropriate, feasible and sustainable [4]. Moreover, at every level, the experts can decide the key indicators for every disease; as this can vary from region to region [5]. A consensus on case definitions should be reached and notifiable diseases should be short-listed for inclusion in the integrated disease surveillance and response system. Moreover, this consensus should include the minimum essential data captured by the system for a response.

Generally, the systems should use bottom-up data flow, in the case of Pakistan, data is collected at the basic health units (BHUs) at the community level, for instance, at union councils. The data is then transmitted to district health headquarters followed by transmission to the federal and national level.

B. ZERO REPORTING

Zero Reporting policy should be enforced to help identify non-reporting sites. This policy is in accordance with the World Health Organization (WHO), surveillance guidelines for poliomyelitis and Japanese encephalitis [5]. Thus, improving the overall data quality. Furthermore, the surveillance system should be robust enough to bear the load of the whole country. It should be capable of exporting data in different file formats to ensure sharing of data among multiple stakeholders.

C. SUPPORT OTHER DATA SOURCES

The surveillance system should be flexible to incorporate other data sources for information exchange. Often hospitals deploy some sort of customized patient management system. It is difficult to introduce an entirely new system at country-wide without the unification with existing solutions. Therefore, a surveillance system should allow easy integration to existing systems. This integration results in more sources of information like citizen data, consensus data, climate, and population data, which allows identification of areas at risk of outbreaks.

D. EASY ACCESS TO DATA STORAGE AND MANAGEMENT

The entire surveillance system should be based on data stored in an easily accessible location. The data should be accessible to all stakeholders at all times hence we will be needing a cloud to put the data on it. The staff should be well trained to manage and operate the system effectively. Appropriate measures should be taken to make the data secure and safe. Generally, the data is stored at a central repository whereas local repositories are also maintained. The data should be accessible through a secure internet connection with key members allowed access to use of the surveillance system dashboard. For instance, cloud-based databases are accessible from anywhere and provide inherent backup and restore support [6]. In locations with limited or no internet connectivity, the system ensures data synchronization from hand-held devices so that the system can continue function during outages [7].

E. AUTOMATED DATA ANALYSIS

Automated and expert analysis is one of the core features of any surveillance system. It ensures timely responses to critical situations. With data analyzed automatically, the system should generate outputs in the form of alerts and recommendations. However, it is important to get the system recommendations monitored by a penal of experts. This can help gauge system accuracy leading to confidence among experts over the system. With the availability of data and use of appropriate decision module, accuracy can be improved over time. Notably, this is a future multi-disciplinary research direction in the field of medical health sciences.

F. CUSTOMIZABLE OUTPUT

The system design should be flexible to easily integrate other tools and applications. Moreover, the output format should be flexible, tailored according to the requirement of field officers. Other features include data visualization, work list, reports, and returns. Such a system allows customization to the user interface as per convenience, increasing productivity while collecting data for the multi-tier surveillance system.

G. TARGETED RESPONSE AND FEEDBACK SUPPORT

Based on the medical data entered at different levels, the surveillance system should generate user-centric and timely recommendations. This can be achieved using an intelligent system based on available information, that is historical events can help suggest a remedial plan. The recommendations can be a list of tasks following some protocol. Upon completion of each task, data could be entered into the system. Such feedback mechanisms are lacking in many surveillance systems, resulting in lower monitoring accuracy, and importantly the intelligent system is missing self-learning [8]. The target is to help field workers take quick actions based on the recommendations.

H. MULTI-EXPERT APPROACH

The system should provide experts from health and ICT sectors an opportunity to work together, with the goal to run the system effectively. The system should use maps to interpret and visualize data helping policymakers to avoid disease outbreaks. Moreover, this could help develop a geospatial model of the outbreak, identify the spatial distribution of different diseases, and develop prevention strategies for a particular region.

III. LITERATURE REVIEW

Health is a basic right of every individual and it is the job of the government to ensure its provision to achieve sustainable health goals. Almost every developed country in the world has a health information system (HIS) in place under the government. The system is primarily used for policy making, trend recording, and detection of outbreaks. The developing countries are either using a limited HMS or are in the process of implementing one.

Domeika *et al.* [9] propose an electronic reporting system for surveillance of communicable diseases in Lithuania. The system is designed to eliminate any data duplication, but takes a month to report a notifiable disease at the national level except a few diseases with high priority. Even in the latter case, reporting takes a week to reach at the national level.

Chandrasekar [3] reviews the importance of surveillance systems and elaborates the issues with existing systems. Generally, such systems should focus on accuracy and notification speed. Therefore, the computerized based system has been introduced.

For instance, in 1997, Sweden introduced a computerized reporting system called SmiNet with the purpose to improve the overall disease reporting process in terms of time. In 1998, an improved reporting tool called SmittAdm developed in Lotus Notes is introduced. This is followed in 2004 with an internet-based version called SmiNet to facilitate users through online reporting and monitoring. The solution uses a central repository at the national level complemented with county servers maintained to support daily activities [10].

In 2004, ULISAS project is implemented as a collaborative effort between Lithuania and Sweden [18]. The project used SmittAdm developed earlier in Sweden and started with an early deployment in a few cities of Lithuania. The system changed from a paper-based monthly aggregated data into a timely case-based electronic reporting system [9].

TABLE 1. Comparisons with existing systems.

Authors	GIS support	Open source	Scalable	Support alerts	Interoperability	Real-time data	Cloud-based	Supports data analysis
Domeika et al. [9]	×	×	✓	×	×	×	✓	×
Rolfhamre et al. [10]	×	×	✓	×	✓	×	✓	×
Reinhardt et al. [11]	✓	×	×	✓	×	×	✓	✓
Ward et al. [12]	×	×	×	✓	×	✓	✓	✓
Brownstein et al. [13]	✓	×	×	✓	✓	✓	✓	×
Yang et al. [14]	×	×	×	✓	✓	✓	✓	✓
Zeldenrust et al. [15]	×	×	×	✓	×	✓	✓	✓
Wakamiya et al. [16]	×	×	✓	×	×	×	✓	×
Noaman et al. [17]	×	×	×	×	×	✓	×	×
Proposed (Adnan et al.)	✓	✓	✓	✓	✓	✓	✓	✓

ULISAS was able to integrate systems for physician and laboratory notifications, data-validation, and online access to the system.

Furthermore, epidemic intelligence community increasingly relies on geographic information systems (GIS) to assess outbreaks in real-time and to facilitate decision makers. The system provides visualization of complex spatiotemporal events and helps analyze data comprising several layers of information. In contrast to traditional maps, GIS is updateable, and they can appropriately help target intervention and prevention programs, especially in less developed countries. EpiScanGIS is a visualization cluster-based open source tool [11]. It extracts relevant information such as the number of cases, fine type, county, p-value, and diameter of the abnormal accumulation of cases. It was launched in Germany in 2006 with the main objective to provide real-time monitoring of meningococcal meningitis in connection with demographic information. The disease is detected through computer-assisted clustering performed on maps. The system is built using open source components and supports dynamic map creation. The system has the potential to monitor other diseases, and improve accuracy using other information sources like weather data, and demographic conditions can further improve the system [11].

The primary objective of such systems is the reporting specific infectious diseases so that an appropriate public health response can be initiated. However, for an effective system, reporting should be timely and accurate. In 2002, an internet-based reporting system called OSIRIS was introduced in Netherland, a replacement to a paper-based reporting system. In traditional systems, medical physicians and microbiological laboratories are required to notify the Gemeentelijke Geneeskundige Dienst (GGD, municipal health services) of patients diagnosed with notifiable infectious diseases as they are responsible for initiating control measures. The GGD are required to send summaries of these reports to the chief medical officer at the Inspectorate of Healthcare and the National Institute for Public Health and the Environment.

Note that the reports being sent to different organizations require different types of data summaries with specific processes followed. The internet-based reporting system supports generating these reports through a unified process, with access to only authorized users. These preliminary reports can be used as an early warning against any outbreak [12].

Brownstein *et al.* [13] presented a data mining approach to collect disease surveillance data through online news sources. The authors propose a system called HealthMap that is freely accessible and automated information system to organize outbreaks according to geography, time, and disease. The platform is a multi-stream real-time surveillance platform continually aggregating reports on new and ongoing infectious outbreaks. The system performs the extraction, categorization, filtering, and integration of reports to facilitate knowledge management and early disease detection. Main sources of information include news reports, expert-curated discussions like ProMED-mail, and validated official reports from reputed organizations. As the system relies on internet sources, it is easier to disrupt. Moreover, it is discouraged to use data without validation and verification in the medical field.

Vlieg *et al.* [19] compare two early warning system implemented in China [14] and the Netherlands [15]. The traditional system deployed in China requires reporting of the diagnosed disease to the Chinese Centre for Disease Control. The health care provider enters case information using a standard form into the notifiable infectious diseases reporting information system (NIDRIS). The web-based system enables healthcare institutions to report any cases of notifiable infectious diseases. To facilitate early warning at different levels, the China infectious disease automated-alerts and response system (CIDARS) is in place. In the Netherlands, when a notifiable infectious disease is suspected and/or laboratory tests confirm it, the case is reported by the attending physician and the laboratory to the regional public health services (PHS). The case information is collected and entered by the PHS into a web-based database for further analyses.

The system uses the 'barometer' algorithm to predict any cases of infectious diseases. Though the countries are located in different geographic regions there are many similarities in the early warning systems.

In [20], Njuguna *et al.* analyze the effectiveness of electronic integrated disease surveillance and response system. The Idea of Integrated Disease Surveillance and Response was partially rolled out in Sierra Leone in 2003 by the Ministry of Health. However, the Ebola Virus Outbreak during 2014-2015 made it obvious for the authorities to implement IDSR properly. Previously, the IDSR system was paper-based, therefore it affected the timeliness and completeness of reporting. Further, the Ministry of Health identified the need for electronic system which was anchored onto the district health information system. This effort was collaborated by the WHO, CDC, and e-Health Africa. The system kick-started the IDSR and according to the statistics collected by the authors, the annual average proportion of on-time weekly reporting was 93% in 2016 which increased to 97% during 2017. The completeness and timeliness of the system helped the government to detect 96% of suspected outbreaks and public health events through this system in 2016 which was further improved to 100% in 2017. However, it was observed that there was also an over-reporting. This caused the data quality to decrease. The reason for the decrease in data quality was that reporting was still being done on paper at some levels. These reports were then uploaded onto the system, however, the paper-based reporting caused the data quality to decrease because it required proper transcription of data from various sources.

Randriamiarana *et al.* [21] introduced the concept of the integrated disease surveillance system in Madagascar during 2007. The data was collected on papers through the health care structures and later transferred to the central level through postal mail or email. As we can see the system was inefficient and lacked in timeliness. The authors identified that only 20% of the data completeness was observed during 2009-10. Further to increase the completeness and timeliness of the data, the data transfer through short messaging services (SMS) through mobile phones was introduced in the two southeast regions of the country. The use of SMS for data is a short term solution. Although the authors concluded that the inclusion of SMS for data reporting has increased the completeness of the data but still it was not effective in case of timeliness and data quality which are the basic requirements of any modern-day disease surveillance system. Moreover, the record-keeping and storage of the data was again a major problem. The system had no real-time data analyzing capabilities and GIS support.

In [16], authors used social media websites for disease surveillance. The work is based on Natural Language Processing, principals of Artificial Intelligence and Machine learning to classify tweets made by different users and tries to extract disease surveillance data. The NLP is used for web documents (MedWeb) designed for extracting health-related information by exploiting data on social media sites like Facebook and Twitter. The MedWeb provides pseudo-Twitter messages in a cross-language and multi-label corpus, covering three languages (Japanese, English, and Chinese) and annotated with eight symptom labels (such as cold, fever, and flu). MedWeb uses a number of tweets that include at least one keyword of the target disease. This system is using Social Media as a platform hence a number of problems also arises and the most important is the quality of data and its completeness. A lot of false positives can trigger the system. The system will also miss a lot of cases as most people do not post about their illness on social media. The system cannot be categorized as real-time and no response activity against an outbreak can be performed due to incomplete and unreliable data.

In [17], the authors propose a system primarily built for predicting Central-line-associated bloodstream infection (CLABSI). Currently, the surveillance of CLABSI is being done manually and is often limited to Intensive Care Units only. The system captures patient data from the hospital databases and predicts whether a patient is likely to develop CLABSI using knowledge discovery rules and CLABSI decision standard algorithms. Applying the proposed approach can decrease CLABSI rates and hence also reduce the patient treatment costs. The system is fully web-based and collects data from hospital databases, scanners, and other IT resources. However, this system is only limited to the CLABSI prediction and hence is not scalable for other diseases. The system is real-time with limited capability and capacity.

Liabsuetrakual *et al.* [22] proposed a web-based epidemiological surveillance system for maternal and newborn. The proposed system includes data analysis and presentation features which are acceptable in many hospitals in Thailand. The system can use on various platforms and accessible through handheld devices. The patients, nurses, and doctors can access the system using their username and password. The system is yet to be installed at the national level.

Simonsen *et al.* [23] discussed the role of big-data toward the disease surveillance system. In other fields of research, big data has proven useful but in health sciences, countries are still relying on a paper-based manual reporting system. In order to improve the health services, new generation systems are required that incorporate the computer intelligence and big data gathered from various sources. Further, the authors have reviewed the efforts required to develop surveillance systems based on social networks, including Google Flu Trends. The authors conclude by advocating the use of a hybrid system based on big-data.

IV. PROPOSED SYSTEM – MedCloud

Undoubtedly, the health sector is a data-intensive industry. It is crucial that the data is available to all stakeholders and tiers within a bounded time. Currently, in Pakistan, the DHIS system in use to some extent. DHIS was developed in the 1990s. It lacks many of the features required for modern infectious disease surveillance. The system is somewhat

paper-based where diseases are reported by sending hard copies or using an email-based system. In many cases, disease outbreaks are discovered through print and electronic media with the government having no or very limited information about the reported outbreak. This ineffective reporting system not only creates problems when managing countermeasures against such outbreaks but it is likely to cause mass hysteria among the population due to late detection of disease outbreaks.

The proposed MedCloud framework is an extended version of DHIS-2, an open source, web-based framework developed at the University of Oslo. The framework is designed to meet the requirements of Pakistan for disease surveillance and health information management. Note that MedCloud is not only a disease surveillance tool but a web-based platform to process, validate, analyze and present statistical data. It is worthwhile to mention that the platform supports analytics and visualization using streaming data.

MedCloud is hosted on a central server accessible through the internet. Users at the district level (basic health facilities) enter weekly surveillance data using devices like desktop computers, laptops, and smartphones. The dashboard is flexible and adaptable and provides access from different devices to view reports, charts, and maps. The underlying interface is reconfigurable and supports regional languages like Urdu and Pashtu. The data entry is easy requiring minimal prior training. It provides a range of validation functionality to ensure that the collected data is adequately complete and of high quality. Moreover, the platform supports surveillance officers to send messages to users or supplement data with epidemiological interpretations based on their assessments.

MedCloud deployment is expected to scale well handling large volumes of data at the country level. A central server is set up as a data warehouse to integrate disparate datasets. The server can easily interface with other information systems providing interoperability, so that provinces can have their preferred information system. The deployment is extensible, that is adding applications for additional functionality is simple. It is configured to allow appropriate access to data at each user level. For instance, users at the province level can access all data from their province while users at the federal level can access the entire data. Furthermore, with limited internet access and cellular coverage, MedCloud is designed to cope with any interruptions caused due to internet connectivity, as it provides offline support for data collection and analysis.

V. MedCloud – FRAMEWORK IMPLEMENTATION

MedCloud is implemented to execute across all platforms using Java. The implementation supports easy integration with other relational DB systems. However, it is implemented on Ubuntu 16.04, with PostgreSQL DB and Tomcat Servlet container. MedCloud server is hosted centrally and accessible from provinces and districts. The main modules of MedCloud implementation (Figure 2) are discussed below.
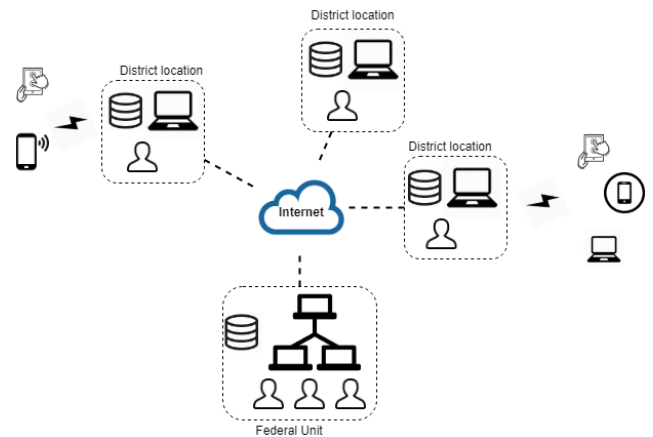
FIGURE 1. MedCloud deployment design at federal and remote units – the remote units gather data from mobile field officers. This data is analyzed locally and/or shared with federal unit for further assessments and recommendations.

A. BACKUP AND INFORMATION SECURITY MANAGEMENT

MedCloud provides data backup and security features available in DHIS-2. The backup facility is based on triggers that automatically creates a backup after a set time interval. This allows users to define a backup plan with the underlying DHIS-2 managing the backup activity. The failover is provided through a secondary server. In case both primary and secondary servers fail, another server is also kept as a backup. The hard disks are cloned at the end of each day to ensure no data loss in case of failures. Note that providing medical services at large-scale can result in system failures leading up to a chain reaction at different levels. Therefore, in MedCloud design, the entire data is stored on network access storage (NAS) placed inside the data center, which is cloned to another NAS placed off-site. Furthermore, MedCloud can also be configured to use virtual machines (VM), that is providing failover with other VMs. The main purpose of using VMs is due to easy scaling and fast recovery.

Due to the sensitive nature of the health data, security management is critical at the MedCloud server. In addition to supported encryption features in DHIS-2, we add cascading encryption to further enhance the security of the stored data. Furthermore, MedCloud is hosted behind an appropriately configured firewall providing an additional security layer. In contrast to DHIS-2, MedCloud uses the HTTPS protocol to connect with users.

B. HIERARCHY MANAGEMENT MODULE

This module is responsible for maintaining the overall hierarchy of the system. To implement the system at the country level, the first step is to define the data flow among nodes and top management. A well-defined organizational structure is required to identify information access at various levels. MedCloud has the provision to add/remove organization units that are used to control hierarchy. In MedCloud, we use these
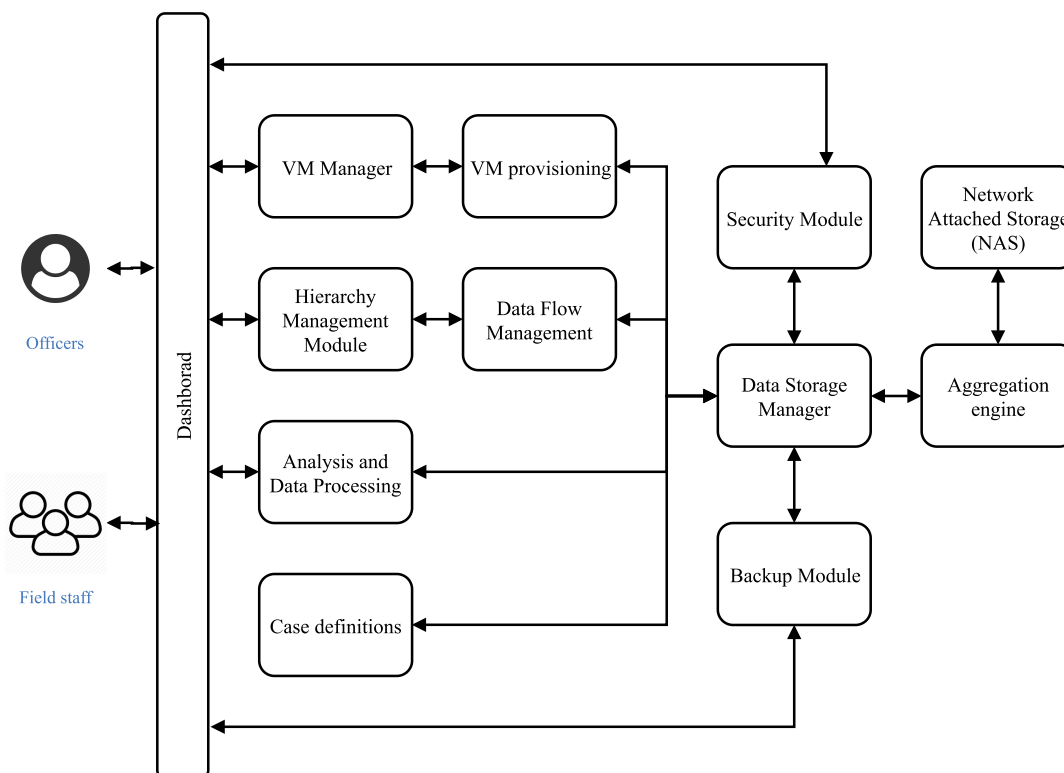
FIGURE 2. MedCloud flow architecture – Officers and field staff workers access the services using a dashboard. All backend modules are hosted on a separate system and accessed through web services. Data is stored at a shared NAS location, accessible to the services. The data is replicated to another NAS locally and at a remote location to reduce the system failure probability.

units to define a flexible hierarchy structure. Since MedCloud is planned for deployment in Pakistan, we create custom organization units with user roles to ensure no irrelevant transfers of information happens among nodes. The result is a hierarchy with nodes capable to view only its own data and data from nodes falling under its area of responsibility. In MedCloud, the organization units can be assigned particular levels within the geographical hierarchy and organized into non-geographical groupings. At every level, there are some required attributes including the name, code, contact person information, and geographical coordinates. These operational units later on support data uploads. Prior to this offloading of data to the data warehouse, it goes through various stages to clean any anomalies and to transform it into the desired structure. Once at the data warehouse, the data is made available for analysis using traditional and online analytical data processing techniques. To further improve performance, the data is also stored in an aggregate form.

C. ANALYSIS AND DATA PROCESSING MODULE

As MedCloud is based on DHIS-2, therefore, it can be used for data gathering, perform user-defined validations and present the results in a customizable form. Moreover, the framework supports easy integration of other tools like user-defined reports, charts, geographical maps, and tables. Further, the aggregated data is stored in data-marts which are

designed to support data analysis. The data aggregated in the form of spatial dimension based on user-defined organization structure. In order to access data, models can be used based on data-specification. As per the DHIS2, data-marts are designed to meet the performance under a high-concurrency situation. Thus, the data analysis request can be efficient and produce results in a single query.st of the data analysis requests can be served using a single query.

A data processing module is referred to as the aggregation engine. It is designed to process a low-level significant amount of data; thus, suitable for the national level management system. Another significant feature of the aggregation engine is to provide quick access to aggregate data. Moreover, it can also manage the task schedule and refresh mechanism of data-marts. Furthermore, MedCloud is customized to manage user rights to view visualisations and to export and process data for use in other analysis tools. For instance, a preprocessed data visualization is available for viewing and download but only for users with an officer role to avoid data leaking.

D. DATA STORAGE IN MedCloud

In large-scale data management system, the data warehousing techniques are usually adopted is either normalized or dimensional. MedCloud is focused on the dimensional approach. The dimensional technique manages the data in the form of

dimensions and real-life events. However, the real-life events provide the numeric data; whereas, dimension provides the location information that gives more meaning to the stored data. In MedCloud, the real-life event/factual data refer to the data object in the user-defined model. The data object can hold data in the form of numbers or string. However, the data become more meaningful when associated with organizational structure defined while configuring the system. Thus, MedCloud flexible structure can allow easy implementation of various dimensions. However, like another system, this organizational model needs to define at the time of system configuration. Moreover, the MedCloud extend these features by allowing the addition of groupset for the user to add custom dimensions even after the data capture process.

1) DATA ELEMENTS AND INDICATORS

The metadata required for configuration is the data elements. The data elements are single items of data that are used to capture the system; for example, the number of measles cases reported. Data elements can be of different type for example text, numbers, dates, times, email addresses or any other type of data which needs to be collected. For each data element, the required metadata are name, code, description, type, permitted values, and data aggregation procedures. The data elements used for the calculation of indicators in MedCloud such as measles cases per year can be divided by the mid-year population to create an indicator for annual measles incidence. The metadata for an indicator also specifies a formula to indicate how different data elements can be combined. The entire metadata can be defined using Web application. However, the metadata can also be prepared in a spreadsheet and uploaded in the MedCloud system as a CSV file. The data integrity checks identify anomalies in the metadata. It is important that metadata is kept up to date with changes such as closure, merging, renaming of organization units, or personal details.

2) DATA FLOW MODULE

The data flow module in MedCloud can provide data movement in both the directions i.e. from the bottom to top level and top to bottom. However, to support the data flow, at the time of infrastructure deployment, the flow needs to be defined. Usually, the flows are defined in terms of data, recommendations, and alerts. Moreover, the type-specific flows can also be implemented in MedCloud.

As the resources are limited, countries usually develop a list of priority diseases for surveillance. The prioritization of communicable diseases in MedCloud can be set up based on features such as disease mortality rate, potential outbreaks, national or international relevancy, and amenability to intervention. However, as per our literature review, the critical diseases in developing countries are hemorrhagic fever, respiratory infection, diarrhea, diphtheria, measles, and tuberculosis. Therefore, the system has been developed to provide surveillance to these diseases initially, but it can be scaled up to provide surveillance for the communicable diseases.

3) CASE DEFINITIONS

The use of country-wide standardized case definitions, methods and tools facilitate the timely use of data at different levels. The commonly used case definitions are shown in Table 2.

4) AGGREGATED DATA

The new cases and death count for different diseases are routinely collected by the health facility and reported to district health offices. This is referred to as aggregated data and it is the simplest form of surveillance indicator data to collect. Aggregated data is adequate for assessing the burden of disease, trend monitoring, identify geographical variation, and outbreaks. An example on-line form for collecting aggregated data is shown in Fig. 3.

TABLE 2. MedCloud case definitions.

Indicator	Abbreviation	Case definition
Acute haemorrhagic fever	AHF	Acute onset of fever of less than three weeks duration in a severely ill patient with any two of the following: haemorrhagic or purpuric rash, nose bleed, blood in vomit/sputum/stool, other haemorrhagic symptoms; and no known predisposing factors for haemorrhagic manifestations.
Acute respiratory infection	ARI	Any person presenting with severe pneumonia characterised by fever (temperature >= 38 degrees C) and one or more of the following: cough, sore throat or shortness of breath.
Severe acute respiratory infection	SARI	An acute respiratory infection,with history of fever or measured fever of >=38 degrees C + cough + onset, within the last 10 days, which requires hospitalisation.
Acute watery diarrhoea	AWD	An illness characterised by three or more watery (non-bloody) stools within a 24 hour period and severe dehydration, in a person aged under five years.
Diphtheria	Diphtheria	An upper respiratory tract illness with an adherent membrane of the nose, pharynx, tonsils or larynx.
Influenza-like illness	ILI	An acute respiratory infection with measured fever of >=38°C + cough + onset within the last 10 days
Measles	Measles	Any person in whom a clinician suspects measles infection, or any person with fever and maculopapular (i.e. non-vesicular) rash and cough, coryza or conjunctivitis

FIGURE 3. Online data collection form.

FIGURE 4. View of district dashboard showing different reported cases.

FIGURE 5. View of district dashboard showing reported cases of one indicator alongside the trend line.

VI. BENEFITS OF THE METHOD FOR THE STAKEHOLDERS

The proposed system provides a number of advantages to research communities, policy makers, health authorities, public health experts and the general public. The system extracts useful information that can be used for monitoring, policy-making, trend analysis, etc. The system is expected to help in the identification of patterns in the distribution of diseases across the country. The GIS mapping of the disease spread and its real-time interpretation is the key factor required for the disease surveillance and response. These geospatial figures and their analysis are very useful for research communities and especially to policymakers and health authorities.

The collected data can be further processed to help the public health experts in evaluating the influential factors in disease spread. The data can be exported into different formats which can be used with other systems like climate and weather to better understand the spread of diseases.

The GIS mapping and location feature of the system helps to pinpoint the high-risk areas and population (for fine-grained confirmatory analysis). An example is shown in 6. This provides opportunities for research communities as well as for the policy makers and health authorities. The Data Visualizer tool of the system is a very powerful feature. It automatically processes the data and then provides the outputs as shown in 4. The automated data analysis and interpretations help in getting the outputs in time, which in return helps in generating appropriate and timely responses as shown in 5. This automation will be very helpful in detecting the outbreaks on time and generating an appropriate response for example alarms at emergency operations control at the national level.

If we analyze the system for the implementation point of view, the system has been designed in a way that can handle a huge amount of data collected from all over the country in real-time. The system has been made secure by not only providing the best state of the art firewalls and VPN hardware but also securing the Java-based software. The system is highly dynamic and flexible. The admins can even change the
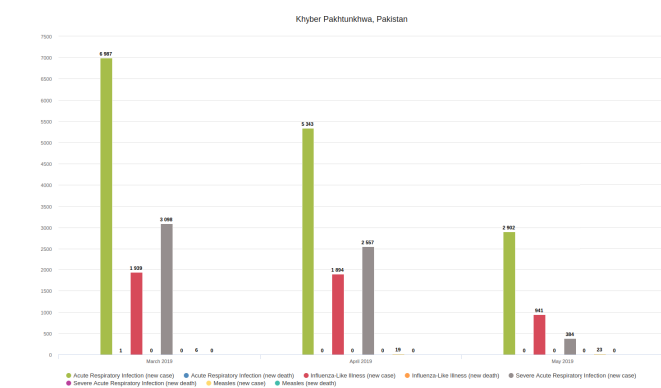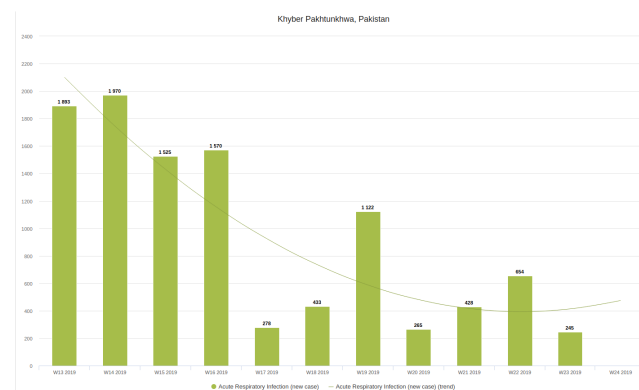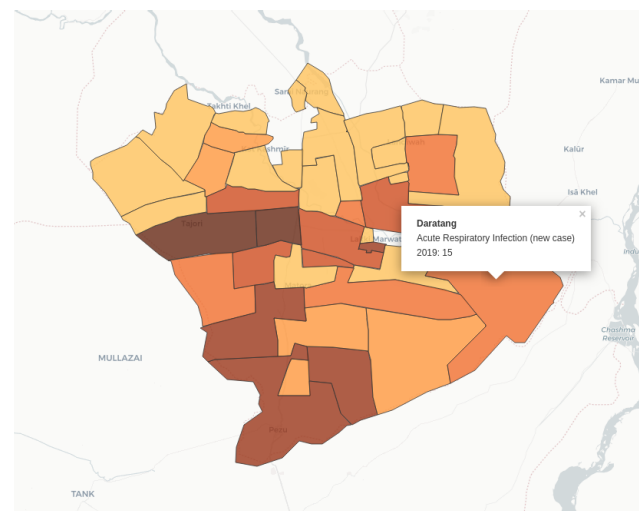

FIGURE 6. GIS mapping to pinpoint high-risk areas and population.

flow of data without any user noticing. In addition, the implemented system can deliver data to external analytical software for further analysis. The system is highly interoperable and also can provide the data in excel and pdf formats hence making it easier to analyze the data on other power software like R or SPSS.

VII. DISCUSSION

A hierarchically structured data flow linking the community, health facilities, and public health practitioners at district, province and federal levels, with appropriate analysis at each level allows the public health response to be delivered rapidly while enabling system-wide situational awareness. In the proposed framework design for the underdeveloped countries, the district health offices have key responsibilities for maintaining the surveillance process. At the beginning of each week, surveillance data collection forms need to be completed at all reporting health facilities, each containing aggregated numbers of cases and death count for every priority disease for the preceding week. The completed forms are submitted to district offices, where the data is checked for quality and entered into the online system. At this stage, the health facilities which have not reported are also identified and can be approached for clarification. Moreover, the number of cases for all health facilities are examined. After the data entry phase, the proposed system can generate alerts. Moreover, based on the alerts, further investigation and actions can be defined with in a system. However, the system also provides the appropriate feedback to the health facilities. As per the flexible design, district health offices can coordinate with other institutions or agencies at the district level to create awareness of priority diseases and corresponding case definitions.

In the proposed system, the standard paper-based data collection forms, are adapted from existing manual systems used by WHO [4]. The proposed MedCloud is based on DHIS2, also used to capture the data for diseases at the health facility level. It is important that the health facilities submit routine data collection forms even when there are no notifiable cases or deaths to report, as it is otherwise difficult to distinguish the absence of disease from a failure in the proposed system. The proposed system also support late data entry to facilities district officers. However, the late reporting may lead to late recognition of outbreaks and incidents. Further, there are few priority-based diseases defined in a system that need immediate reporting such as diarrhea, hemorrhagic fever, Measles, Pertussis, and Diphtheria. However, the district offices can also take immediate action for other diseases where the number of cases triggers an automated alert in the proposed system. The surveillance staff have access to the online data collection and analysis module available in the proposed system. The history of MedCloud auto-generated alerts can be seen on the dashboard. Fig. 7 shows the MedCloud dashboard where officers and field staff can perform the required functionality.

Moreover, in the proposed system, information security of patient data is the most important at every level of the system. The proposed system provides user login to access the data and maintains the history of every user. Further, secure communication mechanism is used to transfer patient data.

In the proposed system, the provincial surveillance and response units have key responsibilities for overseeing and maintaining the surveillance and response process in their
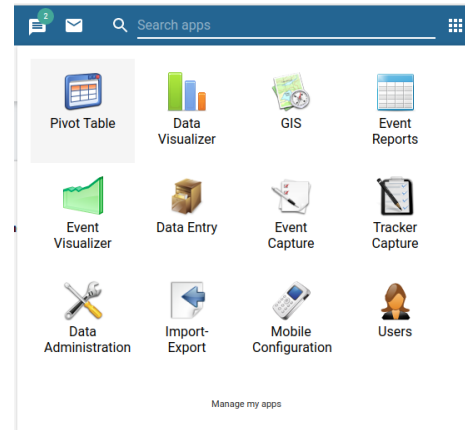
FIGURE 7. Dashboard options – Users can add/change and perform other technical functions.

province that includes, data monitoring, provide advice to districts, provide outbreak alerts, give feedback to district offices with respect to data quality and surveillance analysis/interpretation

VIII. CASE STUDY

The proposed system is designed to work in under developing country i.e. Pakistan. The system is configured with multiple districts, and head office at National Institute of Health (NIH); located in federal territory. The surveillance team at the NIH has key responsibilities at the federal level for overseeing and maintaining the surveillance and response process. Their responsibilities include monitor surveillance system, data quality, alerts, provide advice to activities, outbreak investigation capacity, feedback to provincial teams, training, technical advice, supplies, surge capacity, and technical support. Moreover, the centralized data repository is also placed at the federal level. However, the districts also have their own limited storage to store the data of the local population to timely management of outbreaks.

IX. CONCLUSION

Digital surveillance systems are the need of the era, without proper deployed systems, it is difficult to handle the sudden outbreaks. Moreover, the MedCloud is a step towards the digitization of the disease surveillance system in Pakistan. Previously, isolated systems are in use, which cannot exchange information with other system nor provide data analytics framework. The MedCloud is flexible design based on DHIS-2 framework. This allows for easy integration of new services. Moreover, the hierarchical deployment allows centralizing units to monitor the entire systems placed at district levels and can be accessed from any connected device. It has improved the timeliness of the data flow and maintains the data quality; as both of these factors are very crucial when it comes to disease surveillance and response in any country. If the disease if not reported timely and with correct data, no appropriate and in-time actions can be taken in response to an outbreak. In the future, we would like to extend this work

include artificial intelligence and machine learning modules to predict the outbreaks. This can help early planning and give ample time to transfer the medical equipment to a remote location well before time.

REFERENCES

[1] D. Griggs, M. Stafford-Smith, O. Gaffney, J. Rockström, M. C. Öhman, P. Shyamsundar, W. Steffen, G. Glaser, N. Kanie, and I. Noble, "Policy: Sustainable development goals for people and planet," *Nature*, vol. 495, no. 7441, p. 305, 2013.

[2] J. D. Sachs, "From millennium development goals to sustainable development goals," *Lancet*, vol. 379, no. 9832, pp. 2206–2211, Jun. 2012.

[3] K. Chandrasekar, "Use of information communication technology (ICT) in communicable disease surveillance—A review of literature," *J. Bio-Med. Informat.*, pp. 41–52, 2012.

[4] *Disease Surveillance for Malaria Control: An Operational Manual*, World Health Org., Geneva, Switzerland, 2012.

[5] *Who-Recommended Standards for Surveillance of Selected Vaccine Preventable Diseases*, World Health Org., Geneva, Switzerland, 1999.

[6] N. Sahal, R. Reintjes, and A. R. Aro, "Review article: Communicable diseases surveillance lessons learned from developed and developing countries: Literature review," *Scandin. J. Public Health*, vol. 37, no. 2, pp. 187–200, Mar. 2009.

[7] X.-N. Zhou, R. Bergquist, and M. Tanner, "Elimination of tropical disease through surveillance and response," *Infectious Diseases Poverty*, vol. 2, no. 1, p. 1, Jan. 2013.

[8] H. Bagherian, M. Farahbaksh, R. Rabiei, H. Moghaddasi, and F. Asadi, "National communicable disease surveillance system: A review on information and organizational structures in developed countries," *Acta Inf. Medica*, vol. 25, no. 4, pp. 271–276, 2017.

[9] M. Domeika, G. Kligys, O. Ivanauskiene, J. Mereckiene, V. Bakasenas, B. Morkunas, D. Berescianskis, T. Wahl, and K. Stenqvist, "Implementation of a national electronic reporting system in lithuania," *Eurosurveillance*, vol. 14, no. 13, 2009, Art. no. 19165.

[10] P. Rolfhamre, A. Janson, M. Arneborn, and K. Ekdahl, "SmiNet-2: Description of an Internet-based surveillance system for communicable diseases in Sweden," *Eurosurveillance*, vol. 11, no. 5, pp. 15–16, May 2006.

[11] M. Reinhardt, J. Elias, J. Albert, M. Frosch, D. Harmsen, and U. Vogel, "EpiScanGIS: An online geographic surveillance system for meningococcal disease," *Int. J. Health Geographics*, vol. 7, no. 1, p. 33, 2008.

[12] M. Ward, P. Brandsema, E. van Straten, and A. Bosman, "Electronic reporting improves timeliness and completeness of infectious disease notification, The Netherlands, 2003," *Eurosurveillance*, vol. 10, no. 1, pp. 7–8, Jan. 2005.

[13] J. S. Brownstein, C. C. Freifeld, B. Y. Reis, and K. D. Mandl, "Surveillance sans Frontieres: Internet-based emerging infectious disease intelligence and the HealthMap project," *PLoS Med.*, vol. 5, no. 7, p. e151, 2008.

[14] W. Yang, Z. Li, Y. Lan, J. Wang, J. Ma, L. Jin, Q. Sun, W. Lv, S. Lai, Y. Liao, and W. Hu, "A nationwide Web-based automated system for early outbreak dectection and rapid response in China," *Western Pacific Surveill. Response J.*, vol. 2, no. 1, p. 1, 2011.

[15] M. Zeldenrust, J. Rahamat-Langendoen, M. Postma, and V. J. van Vliet, "The value of ProMED-mail for the Early Warning Committee in The Netherlands: More specific approach recommended," *Euro Surveill.*, vol. 6, no. 13, p. 1, 2008.

[16] S. Wakamiya, M. Morita, Y. Kano, T. Ohkuma, and E. Aramaki, "Tweet classification toward Twitter-based disease surveillance: New data, methods, and evaluations," *J. Med. Internet Res.*, vol. 21, no. 2, 2019, Art. no. e12783.

[17] A. Y. Noaman, A. H. M. Ragab, N. Al-Abdullah, A. Jamjoom, F. Nadeem, and A. G. Ali, "WMSS: A Web-based multitiered surveillance system for predicting CLABSI," *BioMed Res. Int.*, vol. 2018, pp. 1–11, Jul. 2018.

[18] M. Domeika, G. Kligys, O. Ivanauskiene, J. Mereckiene, V. Bakasenas, B. Morkunas, D. Berescianskis, T. Wahl, and K. Stenqvist, "Implementation of a national electronic reporting system in Lithuania," *Eurosurveillance*, vol. 14, no. 13, 2009, Art. no. 19165.

[19] W. L. Vlieg, E. B. Fanoy, L. van Asten, X. Liu, J. Yang, E. Pilot, P. Bijkerk, W. van der Hoek, T. Krafft, M. A. van der Sande, and Q.-Y. Liu, "Comparing national infectious disease surveillance systems: China and The Netherlands," *BMC Public Health*, vol. 17, no. 1, p. 415, Dec. 2017.

[20] C. Njuguna, A. Jambai, A. Chimbaru, A. Nordstrom, R. Conteh, A. Latt, S. O-Tipo, R. Musoke, J. Githuku, Z. Yoti, A. Yahaya, A. Talisuna, S. Rajatonirina, and I. S. Fall, "Revitalization of integrated disease surveillance and response in Sierra leone post Ebola virus disease outbreak," *BMC Public Health*, vol. 19, no. 1, p. 364, Dec. 2019.

[21] R. Randriamiarana, G. Raminosoa, N. Vonjitsara, R. Randrianasolo, H. Rasamoelina, H. Razafimandimby, A. L. Rakotonjanabelo, R. Lepec, L. Flachet, and A. Halm, "Evaluation of the reinforced integrated disease surveillance and response strategy using short message service data transmission in two southern regions of madagascar, 2014–15," *BMC Health Services Res.*, vol. 18, no. 1, p. 265, Dec. 2018.

[22] T. Liabsuetrakul, T. Prappre, P. Pairot, N. Oumudee, and M. Islam, "Development of a Web-based epidemiological surveillance system with health system response for improving maternal and newborn health: Field-testing in thailand," *Health Informat. J.*, vol. 23, no. 2, pp. 109–123, Jun. 2017.

[23] L. Simonsen, J. R. Gog, D. Olson, and C. Viboud, "Infectious disease surveillance in the big data era: Towards faster and locally relevant systems," *J. Infectious Diseases*, vol. 214, no. 4, pp. S380–S385, Dec. 2016.

ADNAN BASHIR received the bachelor's degree in computer engineering from Bahria University, Islamabad, Pakistan, in 2010. He is currently pursuing the M.S. degree in computer science with the National University of Sciences and Technology, Islamabad. He has served as the IT Manager with Pakistan Academy of Sciences, from 2011 till 2017. He is also working with the Centre for Disease Control, Atlanta, GA, USA, in their project in the National Institute of Health, Islamabad. He is also working to establish a country wide disease surveillance system and is building up a data center, NIH, which is strong enough to receive data from whole of the Pakistan. His main research interests include health information management systems, hospital information management systems, integration of health, and IT.

ASAD WAQAR MALIK received the Ph.D. degree with majors in parallel and distributed simulation/systems from the National University of Science and Technology (NUST), Pakistan, in 2012. He is currently an Assistant Professor with the Department of Computing (DOC), School of Electrical Engineering and Computer Science (SEECS), NUST. His primary areas of interest include distributed simulation, cloud/fog computing, and the Internet of Things.

ANIS UR RAHMAN received the master's degree in parallel and distributed systems from Joseph Fourier University, France, and the Ph.D. degree in computer science from Grenoble University, France, in 2013. He is currently an Assistant Professor with the School of Electrical Engineering and Computer Science, National University of Sciences and Technology, Pakistan. His main research interests include modeling of visual attention by assessing the different mechanisms guiding it, salient multiobject image and video segmentation and tracking, and efficient implementations of large-scale scientific problems on commodity graphical processing units.

SOHAIL IQBAL received the M.Sc. and M.Phil. degrees in mathematics from Quaid-i-Azam University, Pakistan, in 2002 and 2004, respectively, and the Ph.D. degree in surgical robotics from the University of Paris-Est, France, in 2010. He has served as a Scientific Advisor under the French Ministry of Higher Education, from 2008 to 2010. He is currently an Assistant Professor with the National University of Sciences and Technology, Islamabad. His main research interests include apply mathematical techniques to different areas of computer science and engineering with special focus on robot modeling, energy management, and fuzzy controllers.

PAUL ROBERT CLEARY has been a Consultant Epidemiologist with Public Health England (and a predecessor organization, the Health Protection Agency), since 2009. After over a decade as a Hospital Doctor in a variety of specialties, including infectious diseases, and a period of research in the U.K., and Malawi, Dr. Cleary undertook public health specialist training. He has considerable experience of working with infectious disease surveillance systems and has investigated numerous local, national, and international infectious disease outbreaks. He has extensive teaching experience relating to epidemiology and statistics for the U.K., Field Epidemiology Training Programme (FETP), and the European Intervention Epidemiology Training Programme (EPIET).
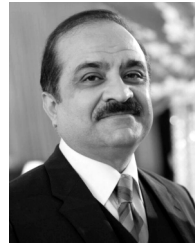
AAMER IKRAM graduated in 1987. He received the Diploma degree in pathology, in 1990, the M.C.P.S. degree, in 1991, and the Ph.D. degree, in 2014. He is currently a Clinical and Environmental Microbiologist and a Biosafety Professional. He is also a Professor and a Consultant of Pathology. He is also an attained Registered Biosafety Professional (RBP) from the American Biological Safety Association, a Biosafety Professional (BSP) from the Institute of Safety in Technology and Research, U.K., and an IFBA Certified Professional. He was awarded FRCP by Royal College Edinburgh, in 2012, the FRCPath by Royal College of Pathologists London, in 2014, and a Fellowship in Public Health from Royal Colleges, U.K., in 2018. He is also the Executive Director of the National Institute of Health, Islamabad, Pakistan. He is also acting as a National Coordinator for global fund as well. He received a Fellowship of microbiology, in 1998.

● ● ●