IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 20, 2021, accepted January 30, 2022, date of publication February 7, 2022, date of current version February 15, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3149349

Towards Data Driven Spatio-Temporal Threshold Identification Based on Cost Effective Public Health Information Management Framework

MUHAMMAD NAZAKAT[1], FATIMA KHALIQUE[2], SHOAB AHMED KHAN[1], AND NADEEM AHSAN[1]

[1]Sir Syed Centre for Advanced Studies in Engineering, Islamabad 44020, Pakistan
[2]Department of Computer Science, Bahria University, Islamabad 44000, Pakistan

Corresponding author: Fatima Khalique (fathema.khalique@gmail.com)

ABSTRACT Appropriate public health action comes from data driven decision support systems. While sophisticated health information exchange framework may be costly in developing countries, the health care delivery system in place may provide a promising infrastructure that spans all parts in a region. Therefore, while digital and non digital data is constantly being generated from variety of sources including public and private health sectors, the health care delivery systems remain the primary and most fundamental source for data on population health status. For low and middle income countries with minimum digitization and resource constraints, traditional existing health care delivery system can be taken advantage of, for efficient production and timely transmission and utilisation of data to identify thresholds for disease outbreak in order to improve health status and health system performance. Due to lack of appropriate universally agreed criteria for threshold, defining local thresholds for infectious disease is not only crucial but also more appropriate. In this paper, we present a low cost data driven framework called Health Data Driven Framework (HDDF) through which data generated at health care facilities may be used for threshold detection and alarm generation. We also identify a localized method based on spatio-temporal mining of available data for appropriate threshold identification.

INDEX TERMS Data driven public health, public health informatics, threshold identification, infectious diseases, spatio temporal analysis.

I. INTRODUCTION

Health data collection, analysis and transmission is conducted to achieve public health targets of population health improvements. Data collection and analysis cannot be designed to consume resources if action does not follow. Conversely, appropriate action comes from evidence based decision making that relies on delivery of data available through multiple pathways from heterogeneous sources [1]. Public health problems must be well defined before a solution can be identified. The health problem definition is realized through quality health related data. Without meaningful data, public health dynamics are misinterpreted, resulting in policies and programs that do not adequately address the problem leading to an appropriate allocation of resources. The pathway through which data reaches public health authorities defines the data quality. An effective and functional information system provide a sound foundation towards ensuring that programs designed based on received data meet their goals and resources allocated are not biased given the underlying health situation [2]. In developing countries, information systems often fail to achieve the public health objective due to numerous reasons including lack of infrastructure, clearly defined data pathway for disease reporting, limited trained staff and lack of technological resources [3]–[5]. In contrast to developed countries where health information systems are organized, systematic and designed to adapt to rapid technological developments, the developing countries need a cost effective way of ensuring data driven public health policy making that is built on existing available limited infrastructures.

Despite the technological advancements in health care in general and public health in particular, much needs to be done in developing countries. In lower and middle

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

income countries, the context of public health is considerably different from developed countries where health is a basic commodity and systems for evidence based decision making are in place. The availability of data is one of the many differences causing health disparity globally.

A major challenges in existing surveillance methods used in low and middle income countries are their variety, variation in implementation details and fragmentation [6]. The multiple programs per disease are expensive in terms of time, resources and budget. The health care delivery systems are not aligned with surveillance requirements to allow effective flow of health data. Multiple programs run in parallel for disease surveillance engaging resources and increasing cost while providing limited meaningful analysable data. The fragmented programs data either remain un-integrated or consume further time, cost and resources for its integration and analysis [7]. This introduces huge delay in response towards disease outbreaks due to lagged disease threshold picture. In addition, in many countries digitalization of paper based records has not been fully realized causing many high priority disease cases to go unreported. Furthermore, in order to respond timely, threshold based disease alarms need to be generated. These thresholds are derived from baseline incidence of diseases. A challenge is posed by the fact that the baseline calculation requires availability of data over a period of time that requires a framework for data integration, storage and analysis [8].

Therefore, while outdated surveillance mechanisms in developing countries cannot be recreated fully without a huge cost penalty, it is possible to strengthen the existing systems in cost effective ways to ensure it provides an effective evidence base for disease analysis leading to data driven decision making [9]. This can be achieved through taking advantage of existing health care delivery infrastructure in place by transforming it to provide a data driven pathway for achieving public health objective. This pathway available through health care delivery systems can provide a pivotal foundation to implement digital public health information system in a cost effective way [10].

The paper presents Health Data Driven Framework (HDDF), that is based on the existing health care delivery system infrastructure for prompt data delivery to decision makers in order to address the current and emerging healthcare related issues including but not limited to threshold identification for infectious disease outbreak in lower and middle income countries. The HDDF presented provides a pathway for data acquisition, analysis and transmission of medical and health related data. The framework includes multi tiered architecture with components placed strategically to allow seamless transmission of health data from multiple sources to a conceptually centralized tier which maybe physically distributed for analysis and decision support. The application experience of framework is presented in context of Health Care Delivery System (HCDS) in Pakistan, however it is applicable to any low or middle income country with resource constraints through proper modifications.

II. RELATED WORK

One of the major application of data received through surveillance system is early detection for infectious disease for effective intervention [11], [12]. These systems rely on timely availability of data to predict disease epidemic through underlying surveillance architectures [13]. Approaches working towards threshold identification can be classified under two concepts. Early Detection Methods (EDS) and Early Warning Systems (EWS).

EDS detect the periods of epidemics after the epidemics has already started. If epidemics are identified to have occurred at early stages and continue to increase over time, the interventions can be useful and effective [14]. Thus EDS work with persistent epidemics where delay in response may not be a limiting factor. In contrast, methods in EWS attempt to predict the onset of epidemic before the transmission increases to a threshold value. These methods rely on data from other sources such as environmental factors or other social determinants to predict the onset of an epidemic. However, such data may not be readily available for the spatio-temporal region and time to study might not also be clearly defined.

The Early Aberration Reporting System (EARS) by CDC predicts the disease outbreak for syndromic surveillance after the incidents. This system has been popularly used by local and state health departments. However, EARS work with recent data for baselines. Other set of techniques that work with long term data use quasi-Poisson regression and consider seasonality and trends for outbreak detection [15]–[18].

Sensitivity and specificity are the standard ways to evaluate the validity of threshold. Sensitivity can be measured as %age of truly predicted epidemic periods for alert while specificity is measured as %age of non epidemic periods predicted for alert. However, very few studies have been able to apply sensitivity and specificity for threshold validation [19]–[22]. Other non standard and informal methods for threshold validation can be problematic in EDS [23], [24].

In this paper, we present a disease surveillance pathway that can be used a threshold identification method based on spatio-temporal mining, neighbourhood analysis and percentile over time slices in selected regions of data. We use specificity and sensitivity to validate our approach and compare the results with two local and a global threshold standard.

III. MATERIALS AND METHODS

In order to understand the applicability of HDDF to different HCDS, it is crucial to understand the existing infrastructure of HCDS of the country. Therefore, we present here the national health care delivery system in Pakistan and map our framework architecture components to the existing system.

A. HEALTH CARE DELIVERY SYSTEM

Pakistan inherited its HCDS through British government rule in 1947. The existing HCDS of Pakistan is diverse, which consist of public, private, civil society, humanitarian

contributors, and national and international donor agencies. The HCDS consist of both public and private healthcare services. Private sector includes private hospitals, clinics, homeopathic, and Hakeems. In addition to healthcare services provision is obligated through legislation by the provisional government, private sector of healthcare, national and international organizations also have significance contribution. Both vertical and horizontal elements of HCDS exist in Pakistan. Administratively, Pakistan consists of provinces, Punjab, Sindh, Balochistan and Khyber Pakhtunkhua (KPK). In addition, federal includes Islamabad Captital Territory (ICT) as well as Federally Administered Tribal Areas (FATA). Each province is governed by Health minister, health secretary and Director General(DG) health. Figure 1 outlines the health information exchange and governance structure in Pakistan. Vertical public healthcare functions in three layers; primary, secondary and tertiary [25]. Primary healthcare (PHC) facilities consist of Basic Health Units (BHUs) and Rural health Centers (RHCs). The Tehsil Head Quarters (THQ) hospital handles population at sub-district level. The District Head Quarters (DHQs) hospitals provide healthcare facilities to their respective district population. Both THQs and DHQs hospitals form the Secondary Health Care facilities (SHC). In addition, Tertiary Health Care (THC) facilities and public hospitals are situated in towns and cities and also serve as teaching facilities. In addition, several organizations have their own health care facilities for the employees. Table 1. gives the total number of various facilities across the provinces of Pakistan [26].

People of small towns and villages visit PHC facilities on daily basis for various health issues. Moreover, health awareness programs run by government and vaccination programs are carried out by these health facilities [27]. For complex cases, patients are referred to SHC facilities. Medical records of this tier are mostly in manual format by record entry to registers or files. The SHC facilities are relatively more advanced in terms of healthcare resources. Patients are frequently referred form BHUs and RHCs here for their treatment. These facilities also have operation theatres and medical wards for patients. At present, some of these facilities have digital records for daily in/outs of patients and medical treatments but not all THQ and DHQ hospitals are completely digitized [28]. These THC facilities have most advanced health resources and capable of handling hundreds of patients at national level. Additionally, these facilities are responsible for coordination of healthcare issues with ministry. Critical medical cases referred by secondary tier are accommodated and treated here. Most of the federal health facilities are already digitized and some are in process. Figure.2 shows approximate number of people visiting each of the HCDS facility at different layers [26].

B. PROPOSED INTEGRATED DATA DRIVEN FRAMEWORK

In this paper we propose a framework that aligns with the existing HCDS in Pakistan through incorporation of small

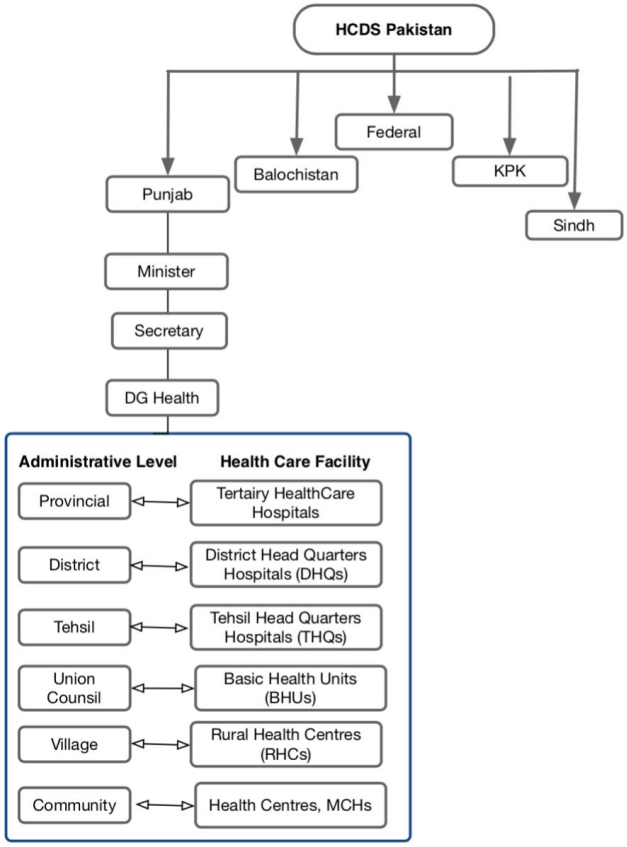

FIGURE 1. Healthcare delivery and health information exchange and governance structure in Pakistan.

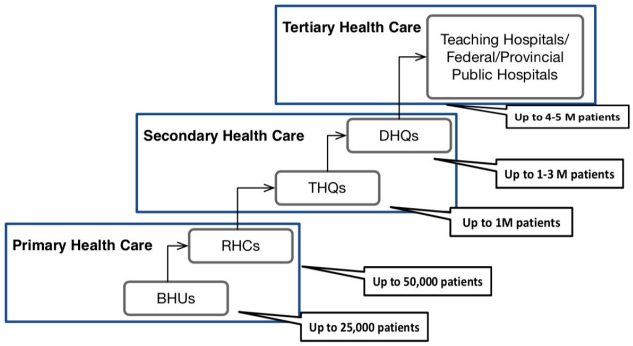

FIGURE 2. Healthcare at national and provincial level.

software components at all tiers in system. These components form a multi layered framework through primary, secondary and tertiary health levels in Pakistan and can be adopted by other developing countries through proper modifications. The proposed health data driven framework (HDDF) consists of tier that can b mainly categorized as source layer and server layer. The source layer consists of source adapters, that are software components designed to extract attributes of public health interest from digitized electronic records in the respective tiers. In addition source layers consists of gateways, that are software components designed to transmit

TABLE 1. Number of health care facilities at different hierarchies across all provinces in Pakistan.

	Tertiary	Secondary		Primary				
	Public Hosp.	DHQ Hosp.	THQ Hosp.	RHC	BHU	Dispensary	MCH Centres	Sub-h. Centre
Punjab	23	34	88	293	2461	499	289	443
Sindh	7	11	56	130	774	643	90	15
KPK	9	21	77	90	822	307	49	30
Balochistan	4	27	10	82	549	575	90	24
Federal	2	12	14	12	179	11	22	211
Other	0	11	39	36	223	277	235	484
Total	46	108	280	638	5002	22318	775	1207

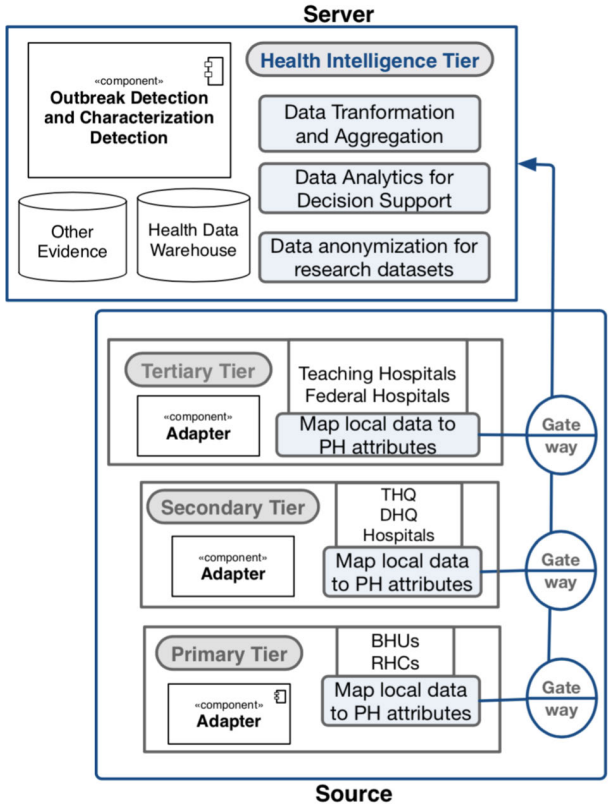

FIGURE 3. Schematic overview of proposed framework with components placed strategically at each HCDS tier.

and receive standardized health data to and from source layer. The server layer consists of business intelligence components for public health in order to support evidence based decision making driven by the data from multiple underlying sources. In order to align the proposed framework with existing HCDS in Pakistan, figure.3 gives placement of components at appropriate levels of HCDS discussed above. The schematic overview presented in figure,3 shows that a complete automated health data pathway can be created incrementally through gradual addition of BHUs and RHUs in source layer. Similarly, the framework can be scaled to include more primary, secondary and tertiary sources.

1) SOURCE LAYER AND HCDS TIERS

The source layers in the primary tier consists of software components responsible for handling health data generated through patients visiting BHUs and RHCs. In proposed framework the medical record generated daily is digitized and stored with the help of source adapters at local level. The role of these adapters is to standardize the health data representation through basic transformation strategies and transmit this data to the tertiary tier. This transmission can be done on periodic basis defined by the public health authorities without waiting for threshold levels to be achieved. However, not all data need to be transmitted through the digital gateways. Health data attributes specific to public health interest are identified and mapped to a standard health representation. This means that similar multiple light source adapters can be configured and installed at every BHU and RHC.

The source adapters at secondary tier of HCDS consists of software components responsible for handling data generated at THQ and DHQ hospitals. The adapters at this level are responsible for the medical records generated daily at each facility and transmitting it to provincial level. The source adapters at this level follow the same configuration as done at primary tier, however, more health data attributes are available at this level for transmission to tertiary tier.

The tertiary tier consists of software components responsible for handling data generated at federal public hospitals, medical research centres and teaching hospitals in provincial and federal region. The source adapters at this level are more technically sophisticated with complex data mappings due to scale of patient handling and coordination responsibilities undertaken by the facilities at tertiary tier.

2) HEALTH INTELLIGENCE TIER

This tier forms the data integration and analysis platform with a public health data warehouse as a foundation of health intelligence. In addition to data storage, this tier

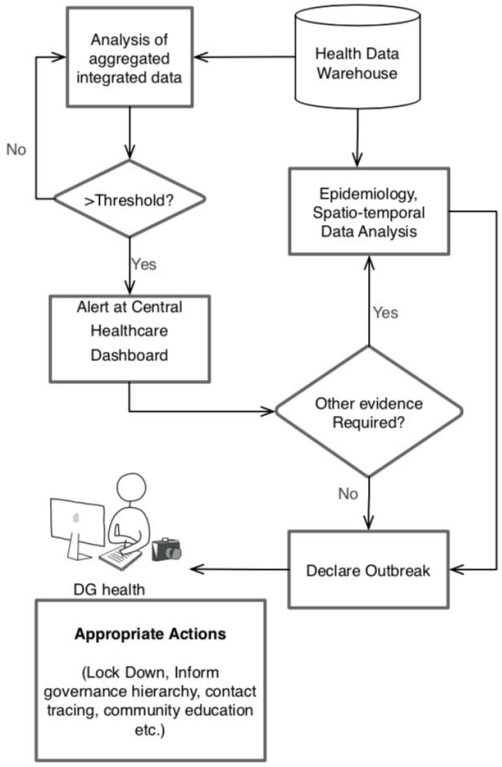

FIGURE 4. Information flow for threshold based disease outbreak alerts at health intelligence tier.

is responsible for generating alerts and alarms based on pre defined thresholds and disseminating the information to concerned health authority. The tier incorporates aspects of business intelligence for health such as descriptive, predictive and prescriptive analytics to solve public health related problems including disease alerts, contact tracing, supplies and equipment allocation and tracing, policy evaluation and analysis. This information becomes available to public health authorities where necessary actions can be taken for prevention, response and recovery from major public health incidents.

In addition, the framework is scalable. For new primary, secondary and tertitary healthcare facilities, new source adapters can be implemented. More evidence data can also be used to augment the source pool for public health analytics such as data from population census, epidemiology data, socio-economic data as well as spatio-temporal features of the regions under study. Figure.4 shows the flow of information for threshold based disease outbreak alerts at health intelligence tier. The selection of threshold is an important criteria for early outbreak detection. The availability of pathway to acquire and store data from source to analytics makes it possible to apply threshold that are applicable a certain geographical region based on it the disease historical outbreak. The threshold applied can be a fixed recommended value or can be identified based on data analysis. For data based threshold value temporal identification methods are suitable for infectious diseases that require historical data for threshold calculation for a certain region.

3) THRESHOLD IDENTIFICATION BASED ON SPATIO-TEMPORAL MINING

In order to determine the trend of disease outbreak for threshold identification, we first identify the spatial hotspots for the disease. For this purpose, we cluster the patient point data using k-means clustering. This clustering is done based on patient addresses. Thus point in close spatial promixity are clustered together. The centre for these clusters is identified and corresponding districts for the centres are identified. This means the number of points in each cluster can be different from the number of patients reported from that district administratively. In order to introduce the temporal feature towards threshold identification, we use Getis-Ord method for each weekly time slice of data. The common districts from clustering results and Getis-Ord method are identified as hotspot locations for threshold determination. These hotspot locations are again determined from single case based point data. However, in this step the point data is the patient addresses that are part of each cluster. Getis-Ord Gi* statistic is used to determine spatially significant hotspots. This statistic takes into account the state of adjacent and neighboring districts in terms of disease count.

$$G_i^* = \frac{\sum_{j=1}^{D} \omega_{i,j} c_j^d - \bar{C} \sum_{j=1}^{D} \omega_{i,j}}{S \sqrt{\frac{\left[D \sum_{j=1}^{D} \omega_{i,j}^2 - \left(\sum_{j=1}^{D} \omega_{i,j} \right)^2 \right]}{D-1}}} \tag{1}$$

where c_j^d is the malaria case count at district j and $\omega_{i,j}$ is the spatial weight between districts i and j and D is the total number of districts and

$$\bar{C} = \frac{\sum_{j=1}^{D} c_j}{D} \tag{2}$$

$$S = \sqrt{\frac{\sum_{j=1}^{D} c_j^2}{D} - \left(\bar{C}\right)^2} \tag{3}$$

For each district, a higher value of Gi* statistic implies that the district is a hotspot with respect to its neighbouring districts. This allows us to identify districts that can be used for data driven threshold identification. However, the granularity of spatial hotspots can be reduced to tehsil or union counsil level for more localized analysis.

In this work, we employ a moving percentile temporal identification method for early detection of disease outbreak. 12 centiles are computed using moving percentile method. These percentiles are calculated using temporal information for diseases cases in the identified hotspot locations. We use week as single unit for diseases outbreak threshold detection. The disease events are sliced into weekly counts for each

district identified through spatial analysis. Within each hotspot, the time slices are separately analyzed for moving percentile threshold identification based on previous 3 time slices data. For each time slice t, 12 threshold values are predicted,i.e. ρ_0-ρ_{11}. Each location is then classified as a hotspot or a cold spot for a particular time based on each of the ρ_p where $p = \{0.40, 0.45, .., 0.95\}$. i.e.

$$\begin{aligned} if \sum c_t^d > \rho_p \\ H_t^d = 1 \\ else \\ H_t^d = 0 \\ \forall t \in \{0, .., T\} \\ \forall p \in \{0, .., 11\} \end{aligned} \tag{4}$$

Multiple ρ_p, may classify a location as a hotspot for a given time slice. In order to evaluate the optimal threshold for a given location and time slice, we compare this value with the already established threshold criteria. For this purpose, we determine the WHO recommended threshold and threshold values generated by C1 and C2 algorithm defined in the Early Aberration Reporting System (EARS) that has been used for large scale event monitoring. The appropriate threshold for a certain spatial region is selected based on specificity and sensitivity analysis. We use specificity or True Negative Rate (TNR), sensitivity or True Positive Rate (TPR) and Cohen's kappa coefficient(κ) as evaluation statistic for threshold selection in each hotspot. The Cohen's Kappa for binary classification of a hotspot as outbreak is given as

$$TPR = \frac{TP}{TP} + FN = 1 - FNR \tag{5}$$

$$TNR = \frac{TN}{TN + FP} = 1 - FPR \tag{6}$$

$$\kappa = \frac{2 \times (TP \times TN - FN \times FP)}{(TP+FP) \times (FP+TN)+(TP+FN) \times (FN+TN)} \tag{7}$$

where TP is the true positives, FNR is the false negative rate or miss rate, TN is true negatives, FP is false positives and FPR is the false positive rate. When threshold is low, TPR increases which means more true epidemics will be predicted leading to lesser missed epidemics. On the other hand, with low threshold values, TNR also decreases that will lead to more false alarms. In contrast, high values of threshold will decrease TPR leading to more missed epidemics but false alarms will be reduced. Therefore, the threshold values need to be carefully calculated to balance TPR and TNR.

C. STUDY AREA AND DATASET

In this study we use data for malaria cases from 38 districts of Punjab, Pakistan from year 2015 to 2019. This data is collected through passive surveillance over the multiple tiers in Pakistan HCDS as presented in Fig.2. The data is obtained in patient line list form from Punjab Information Technology Board (PITB) [29]. We use patient location and reporting date from the data transmitted from primary, secondary and tertiary layers. In order to build the model, we used data from 2015-2018 to identify the districts which must be used for prediction model building. 26358 cases of malaria were reported during the study period in Punjab. After preprocessing for missing values for location or date time, we obtained 26254 patient records.

IV. RESULTS

As a first step towards threshold identification, we performed spatial clusteing of Malaria disease based on patient address using k-means. In order to determine the value of number of clusters, we plotted the variance in data as function of k and identified the value that flattens the curve through Elbow Method. We identified 22 clusters.In addition we identified the cluster centres and centroid and total number of patients in each cluster. This information is presented in Table.2. In order to determine the temporal variations among these hotspots, we then proceed to determine the GI* statistic. For this purpose, each cluster is partitioned into weekly time slices with its corresponding disease case members. For each time slice we determine the hotspot based on GI* statistic. For example, for week x, C1 is designated as a hotspot but for week y C2 may become the hotspot with even lesser number of points due to the fact that the neighbouring clusters are also hotspots. This allows the temporal variations in disease events distribution to be represented in the proposed scheme. We identify the corresponding districts for each hotspot location and select the centres with frequently assigned hotspot status over multiple weeks. Based on this method, we identify 12 districts over Punjab for that participitate for threshold selection as shown in Fig.5.

The WHO recommended threshold is currently being employed in Disease Early Warning System (DEWS), in Pakistan. However, in order to improve disease alert system efficiency, the data collected from multiple sources can be analysed to identify evidence and spatio-temporal based thresholds for the disease. As a next step towards threshold identification we applied the centile method and compared the results with the C1, C2 algorithms that provide our two local threshold standards L1 and L2 repectively and WHO recommended method used in DEWS as a global threshold referred to as G.

Table 3 gives the selected thresholds for each hotspot location when using L1, L2 and G methods for malaria threshold for outbreak. For some hotspot locations, multiple techniques identified the same ρ_p For example, at HS_1, ρ_5 is selected as the threshold based on all algorithms where as at H_2, ρ_9 is identified by L1 and G while L2 identifies $\rho_1 1$. Similarly for each hotspot location, the percentile threshold identified by each selected algorithm is given.

V. DISCUSSION

The existing HCDS in Pakistan provides a comprehensive interrelated network through which multiple health facilities are connected. When data is driven through same pathway of existing infrastructure, instead of new models and

TABLE 2. Results of spatial clustering with total number of points in each cluster.

Cluster	No. Of Member Points
C1	1661
C2	2028
C3	896
C4	1526
C5	1027
C6	2694
C7	1507
C8	254
C9	671
C10	1033
C11	1728
C12	1145
C13	1422
C14	2140
C15	517
C16	1369
C17	1888
C18	375
C19	268
C20	947
C21	777
C22	381

surveillance programs, it would greatly reduce the costs. Reliable health information system components can be added at appropriate tiers that are pivotal for efficient production and timely utilisation of information on determinants of health, health status and health system performance. The proposed framework is cost effective due to its use of health care delivery infrastructure already in place where patients visit and some form of paper or digital record keeping of these visits are already being carried out. Installation of small software components makes it possible to extract attributes of public health interests from these records and transmit it to health intelligence tier. Health analytical tier employs analytical and decision support capabilities to achieve public

FIGURE 5. Identified hotspot districts through spatio-temporal mining of malaria data using clustering and GI* statistic in Punjab province of Pakistan.

health targets. The disease outbreak intensity and available resources drive the disease detection and response plan. The outbreak intensity varies from disease to disease which could be measured for a timely response through the availability of health data transmission pathway. In addition, the separate program per disease does not consider the dynamics of diseases in relation to other diseases. For example, while COVID 19 is epidemic in a region, there may be other diseases within the same region. The dynamics and relationship of these diseases can only be understood if data is available in an integrated manner with appropriate determinants. This information is specially useful when these disease share common resources such as ventilators, medications, hospital spaces and physician expertise [31].

In developing countries with scarce resource and high disease incidents, simple thresholds are preferred. A costly information and surveillance system can further escalate the problem by unintentionally adding time delays in data collection, analysis and response. Response thresholds in such regions, therefore, should satisfy the immediate need to identify and contain outbreaks. Same principle can applied to other public health related alerts, for example,

- Physician to cases deficiency alert
- Low vaccination or medicine inventory alert
- Communicable disease endemic alert
- Non-communicable disease alert
- Medical Equipment and testing facilities deficiency alert

According to WHO regulations, an established ''alert threshold'' when reached should trigger an immediate response. However, disease threshold vary depending on elements of transmission and its potential to cause an epidemic. Table 4 shows some thresholds recommended by WHO for disease outbreak alert [30]. For some diseases a single case the threshold is achieved and alert must be generated, for others some complex threshold computation maybe required.The response activities in general include control and prevention, for example, improving water sources to avoid cholera, mass vaccination to prevent COVID 19 and measles, breaking transmission routes or immediate treatment and isolation for dengue or COVID 19. While the

TABLE 3. Threshold comparisons with two local and one global standard and validation using TPR and TNR.

Hotspot	Thresholds											
	L1	TNR %	TPR %	κ	L2	TNR %	TPR %	κ	G	TNR %	TPR %	κ
H_1	ρ_5	100	85	0.86	ρ_5	87.12	82.45	0.64	ρ_5	83.21	87.5	0.76
H_2	ρ_9	50.00	64.8	0.26	ρ_{11}	85	100	1.00	ρ_9	76.5	65.25	0.26
H_3	ρ_8	64.8	45.16	0.48	ρ_{11}	87.6	84.62	0.77	ρ_{11}	81.35	100.00	0.94
H_4	ρ_9	54.8	65.5	0.08	ρ_8	100.00	85	0.63	ρ_{11}	56.5	16.67	0.26
H_5	ρ_9	78.5	45.00	0.32	ρ_9	58.42	53.85	0.46	ρ_8	76.5	83.85	0.71
H_6	ρ_8	75.00	56.3	0.16	ρ_8	64	32.31	0.13	ρ_{10}	76.32	66.67	0.05
H_7	ρ_{11}	100.00	65.5	0.57	ρ_{11}	62.79	46.67	0.31	ρ_{11}	92.31	53.85	0.54
H_8	ρ_8	100.00	95.31	0.89	ρ_7	100.00	80.85	0.56	ρ_{10}	62.55	54.55	0.74
H_9	ρ_6	87.18	87.18	0.56	ρ_6	43.68	54.29	0.06	ρ_6	94.44	20.59	0.13
H_{10}	ρ_{11}	0.63	53.85	0.38	ρ_{11}	85.71	25.5	0.21	ρ_{11}	68.72	0.69	0.18
H_{11}	ρ_9	84.62	53.85	0.61	ρ_{10}	74.8	58.35	0.44	ρ_9	92.31	100.00	0.84
H_{12}	ρ_{11}	100.00	14.29	0.49	ρ_{11}	91.67	45.16	0.51	ρ_{11}	80.45	93.20	0.32

TABLE 4. WHO recommendation for thresholds in case of infectious diseases.

Disease	Threshold Cases
COVID	1
Cholera	
Measles	
Polio	
Yellow Fever	
Dengue and other Viral Haemorrhagic Fevers	
Malaria	1.5 x baseline over 3 weeks
Watery diarrhoea	
Bloody diarrhoea	5
Meningitis	

response are generally known, in order to calculate the thresholds, it is critical to have an effective information system in place that records the disease cases and generate alarms for a region based on data from multiple sources. Through the automated transmission of data from primary, secondary and tertiary resources, the alerts can be automatically generated and presented to DG Health as shown in Figure.4. The figure shows that some disease outbreaks can be declared based on simple threshold values which can be applied through extracting data from health data warehouse and following actions may be applied as a response:

- Assessment of current situation and recommend way forward to health minister and prime minister.
- Analysis for long term measures.
- Selection of appropriate actions for HCDS Management
- Identify the intervention to mitigate the negative impacts.

Since data is available from multiple spatio-temporal regions, sources and for multiple diagnosis, the dynamics of disease can be studied in the presence of other evidences. Disease outbreak can be declared based on analysis of data from all the integrated sources available as shown in Fig.4.

The proposed health data exchange framework for public health to create a data driven decision support system through integrated sharing of health data minimizes the costs associated with the separate surveillance programs by use of existing pathways available through HCDS in a region. Thus, parallel and discrete programs that carry data from health care facilities to public health information hierarchy can be augmented or replaced with a data driven near real time health exchange by creating an interface between health care facilities and public health through software agents and gateways. The health data exchange may also be driven by the regulations and health interoperability standards in order to drive this synergy [32]. In addition, with the availability of the framework, information can flow in both direction, top-down and bottom-up. For example, disseminating response information to concerned departments [33], clinical decision making through results obtained public health analytics [34] and ensuring quality and effectiveness of data acquisition and transmission systems. Other possible implementation may include missing person admission record location during a mass casualty event [35]. Appropriate use of interoperability standards such as Health Level 7 [36] allows transmission of detailed health data including attributes that may not have been part of disease reporting systems but may prove useful for advanced public health analytics.

Therefore, the proposed framework can augment the existing health care delivery system infrastructure for comprehensive and near real time data delivery to decision makers in order to address the current and emerging public health care related issues in lower and middle income countries.

DISCLOSURE STATEMENT

No potential conflict of interest is reported by the authors.

VI. CONCLUSION

With the availability of large volumes of health care data generated at multiple tiers in HCDS, and strengthening of data analysis techniques, more attention has been paid to infectious diseases and their spread among population. Surveillance system have a long hisotry in being an effective tool for delay inevitably occurs between date of onset of infection and the reporting date. However, due to vertical multi tiered HCDS, delays occur between date of reporting and the date it is received at the analytical end. This delay depends on many underlying parameters such as availability of infrastructure, reporting frequency and quality of data.In this work we have presented a framework that can seamlessly be integrated with an existing infrastructure for near real time transmission of data for surveillance purposes.

REFERENCES

[1] W. H. Foege, R. C. Hogan, and L. H. Newton, "Surveillance projects for selected diseases," *Int. J. Epidemiol.*, vol. 5, no. 1, pp. 29–37, 1976.
[2] S. M. Teutsch and R. E. Churchill, *Principles and Practice of Public Health Surveillance*. London, U.K.: Oxford Univ. Press, 2000.
[3] R. Yip and U. Ramakrishnan, "Experiences and challenges in developing countries," *J. Nutrition*, vol. 132, no. 4, p. 827S–830S, 2002.
[4] G. Soto, R. V. Araujo-Castillo, J. Neyra, M. Fernandez, C. Leturia, C. C. Mundaca, and D. L. Blazes, "Challenges in the implementation of an electronic surveillance system in a resource-limited setting: Alerta, in Peru," in *BMC Proceedings*, vol. 2, no. 3. Springer, Nov. 2008, pp. 1–7, doi: 10.1186/1753-6561-2-s3-s4.
[5] L. May, J.-P. Chretien, and J. A. Pavlin, "Beyond traditional surveillance: Applying syndromic surveillance to developing settings—Opportunities and challenges," *BMC Public Health*, vol. 9, no. 1, pp. 1–11, Dec. 2009.
[6] A. Khalid and S. Ali, "COVID-19 and its challenges for the healthcare system in Pakistan," *Asian Bioethics Rev.*, vol. 12, no. 4, pp. 551–564, Dec. 2020.
[7] D. T. Jamison, J. G. Breman, A. R. Measham, G. Alleyne, M. Claeson, D. B. Evans, P. Jha, A. Mills, and P. Musgrove, *Disease Control Priorities in Developing Countries*, 2nd ed. Washington, DC, USA: World Bank, Dec. 2006. [Online]. Available: https://ideas.repec.org/b/wbk/wbpubs/7242.html
[8] R. Wang, Y. Jiang, E. Michael, and G. Zhao, "How to select a proper early warning threshold to detect infectious disease outbreaks based on the China infectious disease automated alert and response system (CIDARS)," *BMC Public Health*, vol. 17, no. 1, p. 570, Dec. 2017, doi: 10.1186/s12889-017-4488-0.
[9] P. Nsubuga, O. Nwanyanwu, J. N. Nkengasong, D. Mukanga, and M. Trostle, "Strengthening public health surveillance and response using the health systems strengthening agenda in developing countries," *BMC Public Health*, vol. 10, no. 1, pp. 1–5, 2010.
[10] D. L. Buckeridge, A. Okhmatovskaia, S. Tu, M. O'Connor, C. Nyulas, and M. A. Musen, "Understanding detection performance in public health surveillance: Modeling aberrancy-detection algorithms," *J. Amer. Med. Inform. Assoc.*, vol. 15, no. 6, pp. 760–769, Nov. 2008.
[11] H. Zhou, H. Burkom, C. A. Winston, A. Dey, and U. Ajani, "Practical comparison of aberration detection algorithms for biosurveillance systems," *J. Biomed. Informat.*, vol. 57, pp. 446–455, Oct. 2015.
[12] J. Xing, H. Burkom, and J. Tokars, "Method selection and adaptation for distributed monitoring of infectious diseases for syndromic surveillance," *J. Biomed. Informat.*, vol. 44, no. 6, pp. 1093–1101, Dec. 2011.
[13] D. C. Hadorn and K. D. C. Stärk, "Evaluation and optimization of surveillance systems for rare and emerging infectious diseases," *Vet. Res.*, vol. 39, no. 6, p. 57, Nov. 2008.
[14] S. Unkel, C. P. Farrington, P. H. Garthwaite, C. Robertson, and N. Andrews, "Statistical methods for the prospective detection of infectious disease outbreaks: A review," *J. Roy. Stat. Soc., Ser. A, Statist. Soc.*, vol. 175, no. 1, pp. 49–82, Jan. 2012.
[15] A. Hulth, N. Andrews, S. Ethelberg, J. Dreesman, D. Faensen, W. van Pelt, and J. Schnitzler, "Practical usage of computer-supported outbreak detection in five European countries," *Eurosurveillance*, vol. 15, no. 36, Sep. 2010, Art. no. 19658.
[16] M.-A. Widdowson, A. Bosman, E. van Straten, M. Tinga, S. Chaves, L. van Eerden, and W. van Pelt, "Automated, laboratory-based system using the internet for disease outbreak detection, The Netherlands," *Emerg. Infectious Diseases*, vol. 9, no. 9, p. 1046, 2003.
[17] S. Unkel, C. P. Farrington, P. H. Garthwaite, C. Robertson, and N. Andrews, "Statistical methods for the prospective detection of infectious disease outbreaks: A review," *J. Roy. Stat. Soc., Ser. A, Statist. Soc.*, vol. 175, no. 1, pp. 49–82, 2012, doi: 10.1111/j.1467-985X.2011.00714.x.
[18] C. Farrington, N. J. Andrews, A. Beale, and M. Catchpole, "A statistical algorithm for the early detection of outbreaks of infectious disease," *J. Roy. Stat. Soc., Ser. A, Statist. Soc.*, vol. 159, no. 3, pp. 547–563, 1996.
[19] M. Albonico, F. De Giorgi, J. Razanakolona, A. Raveloson, G. Sabatinelli, V. Pietra, and D. Modiano, "Control of epidemic malaria on the highlands of madagascar," *Parassitologia*, vol. 41, nos. 1–3, pp. 373–376, 1999.
[20] W. R. McKelvie, A. A. Haghdoost, and A. Raeisi, "Defining and detecting malaria epidemics in south-east Iran," *Malaria J.*, vol. 11, no. 1, p. 81, Mar. 2012.
[21] R.-P. Wang, Y.-G. Jiang, G.-M. Zhao, X.-Q. Guo, and E. Michael, "'Outbreak gold standard' selection to provide optimized threshold for infectious diseases early-alert based on China infectious disease automated-alert and response system," *Current Med. Sci.*, vol. 37, no. 6, pp. 833–841, Dec. 2017.
[22] D. M. Nekorchuk, T. Gebrehiwot, M. Lake, W. Awoke, A. Mihretie, and M. C. Wimberly, "Comparing malaria early detection methods in a declining transmission setting in northwestern Ethiopia," *BMC Public Health*, vol. 21, no. 1, p. 788, Apr. 2021.
[23] J. Cox and T. A. Abeku, "Early warning systems for malaria in Africa: From blueprint to practice," *Trends Parasitol.*, vol. 23, no. 6, pp. 243–246, Jun. 2007.
[24] H. D. Teklehaimanot, J. Schwartz, A. Teklehaimanot, and M. Lipsitch, "Alert threshold algorithms and malaria epidemic detection," *Emerg. Infectious Diseases*, vol. 10, no. 7, p. 1220, 2004.
[25] *Statistics Division, Government of Pakistan Federal Bureau of Statistics. Pakistan*. Accessed: Nov. 17, 2021. [Online]. Available: http://www.statpak.gov.pk
[26] "Data collection survey on health facilites and equipment in the islamic republic of Pakistan," Japan Int. Techno Center Co., Ltd., Jpn. Int. Cooperation Agency, Tokyo, Japan, Tech. Rep. 4RJR18-053, 2018. [Online]. Available: https://openjicareport.jica.go.jp/pdf/12322293.pdf
[27] M. S. Qazi and M. Ali, "Pakistan's health management information system: Health managers' perspectives," *JPMA. The J. Pakistan Med. Assoc.*, vol. 59, no. 1, p. 10, 2009.
[28] R. M. Ansari, Y. Ansari, and S. Y. Ansari, "Status of general practice and challenges to healthcare system of Pakistan," *Open J. Preventive Med.*, vol. 5, no. 12, p. 463, 2015.
[29] G. O. T. P. Punjab Information Technology Board. *Digital Punjab: Disease Surveillance System*. Accessed: Nov. 17, 2021. [Online]. Available: https://pitb.gov.pk/dss
[30] "Comprehensive assessment of Pakistan health information system 2017," WHO Regional Office Eastern Medit., Cairo, Egypt, Tech. Rep. CC BY-NC-SA 3.0 IGO, 2019. [Online]. Available: https://applications.emro.who.int/docs/9789290222651-eng.pdf?ua=1#:~:text=Building%20on%20the%20scoping%20mission,its%20reporting%20obligations%20on%20core
[31] N. Madhav, B. Oppenheim, M. Gallivan, P. Mulembakani, E. Rubin, and N. Wolfe, "Pandemics: Risks, impacts, and mitigation,"in *Disease Control Priorities: Improving Health and Reducing Poverty*, D. T. Jamison *et al.*, Eds., 3rd ed. Washington, DC, USA: The International Bank for Reconstruction and Development/The World Bank, Nov. 2017, ch. 17. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK525302/, doi: 10.1596/978-1-4648-0527-1_ch17.
[32] T. A. Kass-Hout, S. K. Gray, B. L. Massoudi, G.-Y. Immanuel, M. Dollacker, and R. Cothren, "NHIN, RHIOs, and public health," *J. Public Health Manage. Pract.*, vol. 13, no. 1, pp. 31–34, Jan. 2007.
[33] M. E. J. Woolhouse, A. Rambaut, and P. Kellam, "Lessons from ebola: Improving infectious disease surveillance to inform outbreak management," *Sci. Transl. Med.*, vol. 7, no. 307, p. 307, Sep. 2015.
[34] N. Peiffer-Smadja, T. M. Rawson, R. Ahmad, A. Buchard, P. Georgiou, F.-X. Lescure, G. Birgand, and A. H. Holmes, "Machine learning for clinical decision support in infectious diseases: A narrative review of current applications," *Clin. Microbiol. Infection*, vol. 26, no. 5, pp. 584–595, May 2020.
[35] X.-M. Fu, L. Yuan, and Q.-J. Liu, "System and capability of public health response to nuclear or radiological emergencies in China," *J. Radiat. Res.*, vol. 62, no. 5, pp. 744–751, Jun. 2021.
[36] Health Leven Seven International. *Introduction to HL7 Standards. Health Leven Seven International*. Accessed: Nov. 17, 2021. [Online]. Available: http://www.hl7.org/implement/standards/index.cfm?ref=nav

MUHAMMAD NAZAKAT received the bachelor's degree in computer software engineering from the National University of Sciences and Technology (NUST), Islamabad, Pakistan, and the M.S. degree in engineering management from the Sir Syed Centre for Advanced Studies in Engineering (CASE), Islamabad, where he is currently pursuing the Ph.D. degree in engineering management. He is also a Ph.D. Scholar in engineering management with the University of Engineering and Technology (UET), Taxila, Pakistan. He is also working as an Information Communication Technology Officer (ICTO) in a Public Sector ICT Organization of Pakistan. He has worked as a System Analyst and a Database Administrator in IT-based Public Sector Organizations. He has also implemented smart card technology in a public sector ICT organization and completed numerous C& IT projects. His research interests include health informatics, AI, the IoT, and project management.

FATIMA KHALIQUE received the M.S. degree in computer science from Uppsala University, Sweden, and the Ph.D. degree in computer software engineering from the National University of Sciences and Technology (NUST), Islamabad, Pakistan. She is currently an Assistant Professor with the Computer Science Department, Bahria University, Islamabad. She is also an Oracle certified Professional and has worked in industry and academia. She has worked as a Lecturer at NUST and the National University of Modern Languages (NUML), Islamabad. She has also worked as a Software Developer at Zhonxing Telecom Engineering (ZTE), Islamabad. Her research interests include data mining, health informatics, artificial intelligence, data analytics, and machine learning algorithms.

SHOAB AHMED KHAN received the Ph.D. degree in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, GA, USA. He is currently a Professor in computer and software engineering (C&SE). He is also a Professor in computer and software engineering with the NUST College, EME, and the Chancellor of Sir Syed Centre for Advanced Studies in Engineering. He is an Inventor of five awarded U.S. patents and has more than 260 international publications. His book on Digital Design is published by John Wiley & Sons and is being followed in national and international universities. He has more than 22 years of industrial experience in companies in USA and Pakistan. He has been awarded Tamgh-e-Imtiaz (Civil), the Highest National Civil Award in Pakistan, National Education Award 2001 and NCR National Excellence Award in Engineering Education. He has also founded the Sir Syed Center for Advanced Studies in Engineering (CASE) and the Center for Advanced Research in Engineering (CARE). CASE is a primer engineering institution that runs one of the largest post graduate engineering programs in the country and has already graduated 50 Ph.D.'s and more than 1800 M.S. students in different disciplines in engineering, whereas CARE, under his leadership, has risen to be one of the most profound high technology engineering organizations in Pakistan developing critical technologies worth millions of dollars for organizations in Pakistan. CARE has made history by winning 13 PASHA ICT awards and 11 Asia Pacific ICT Alliance Silver and Gold Merit Awards while competing with the best products from advanced countries, such as Australia, Singapore, Hong Kong, and Malaysia. He has served as the Chairperson of Pakistan Association of Software Houses (P@SHA) and as a member of Board of Governance of many entities in the Ministry of IT and Commerce. He has also served as a member of National Computing Council and National Curriculum Review Committee.

NADEEM AHSAN is currently the Dean of the Sir Syed CASE Institute of Technology, Islamabad, Pakistan. He has vast experience of engineering & project management assignments at national and international level. He has worked as the Project Director for NASA on a Project for "Intelligent Unmanned Aerial Vehicle for Biosphere Monitoring," from 1989 to 1992. He has also worked with the University of Michigan Transportation Institute (UMTRI), from 1989 to 1993, on research for the evaluation of damage caused by the trucks and other heavy vehicles on roads. He is a leading figure in the field of engineering & project management. In recognition to his services at various national level projects, he received Presidential Pride of Performance Award, in 2003. He has also been awarded Sitara-i-Imtiaz, in 2008.