


Received August 23, 2020, accepted September 23, 2020, date of publication September 28, 2020, date of current version November 13, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3027234

# Spatio-Temporal Investigations of Dengue Fever in Pakistan Through an HL7 Based Public Health Framework for Hotspot Analysis

FATIMA KHALIQUE<sup>1</sup>, RUBINA SHAHEEN<sup>1,2</sup>, (Member, IEEE), AND SHOAB AHMED KHAN<sup>1</sup>

<sup>1</sup>Department of Computer and Software Engineering, College of E&ME, National University of Sciences and Technology, Islamabad 46000, Pakistan

<sup>2</sup>Department of Software Engineering, Balochistan University of Information Technology, Engineering, and Management Sciences, Quetta 87300, Pakistan

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


**ABSTRACT** The importance of a standard based interoperability framework is now widely recognized. In order to understand the challenges, feasibility and impact of outcomes of such architectures for public health improvement, it is important to implement and demonstrate the usefulness of data analytics based on pathway provided by the framework. This work provides an implementation details and results of transferring dengue data from source database to data analytical programs used by public health authorities. Dengue fever, a mosquito-borne disease that occurs in tropical and subtropical areas of the world, is considered to be a significant threat in both developing and the developed countries. This study investigates the spatio-temporal distribution of dengue in Pakistan from the years 2014-2017 and identify the most frequently affected hotspots of dengue across the the 3 provinces in the country using data mining, clustering and GIS based analysis of data. In addition, the identification of most vulnerable dengue locations has led to investigation into the potential environmental or geographical conditions that may have contributed to dengue prevalence in the area. These investigations present interesting results that shows positive correlation for temperature, humidity and population density with the dengue incidences. In addition, the results are able to identify disease hotspots over time windows that can be mapped to any spatio-temporal scales. The outcome of this research is crucial for optimal use of resources for combating dengue fever at a regional or national scale by identifying the hotspots. The hotspots identified can be use to create a sentinel surveillance network for a pandemic disease. In addition, the work serves as a useful reference for a comprehensive national level early warning and rapid disease outbreak detection framework for any disease of public health context.

**INDEX TERMS** Disease analytics, dengue, HL7 trasmission, hotspot analysis, disease patterns.

## I. INTRODUCTION

The surveillance protocols in many developing countries are disease specific and fragmented. There is no standard based data transmission protocol for notifiable diseases and cases are reported through multiple methods. The data obtained through these methods including spreadsheet reporting, app based reporting is controlled based on predefined variable and format. In order to complement research with quality, on ground data, a pathway must be present to transmit data from source to research and analysis sites in a standardized way. Pakistan has been experiencing dengue fever epidemic since 2010 [1]. From 1995-2004 only 699 dengue cases were reported from 3 districts of Pakistan. During 2005-

2011 numbers of confirmed cases and deaths dramatically increased to 55,946 and 539 respectively effecting 105 out of 146 districts and since then, the disease has become widely recognized as one of the major public health problems in Pakistan. The availability of accurate data patient's addresses, start of epidemics, vector density, composition, trend, resistance level etc) is the fundamental factor to prioritize the areas, type of vector control operations and also the time of intervention(s). For this purpose, presence of functional dengue surveillance at all level is the primary requirement. Among many reasons of not achieving health targets is the lack of standard for health information transmission and absence of standard central repository. The hospitals are not maintaining data according to international formats that hinders information exchange and subsequently its meaningful analysis and utilization of all available data at policy level [2].

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

Population surveys are further limited by the self-reported nature of the information, by a lack of information for small geographic areas, and sometimes by substantial delays in data availability [3]. In addition, data often lacks many important demographic and other features useful for public health disease management.

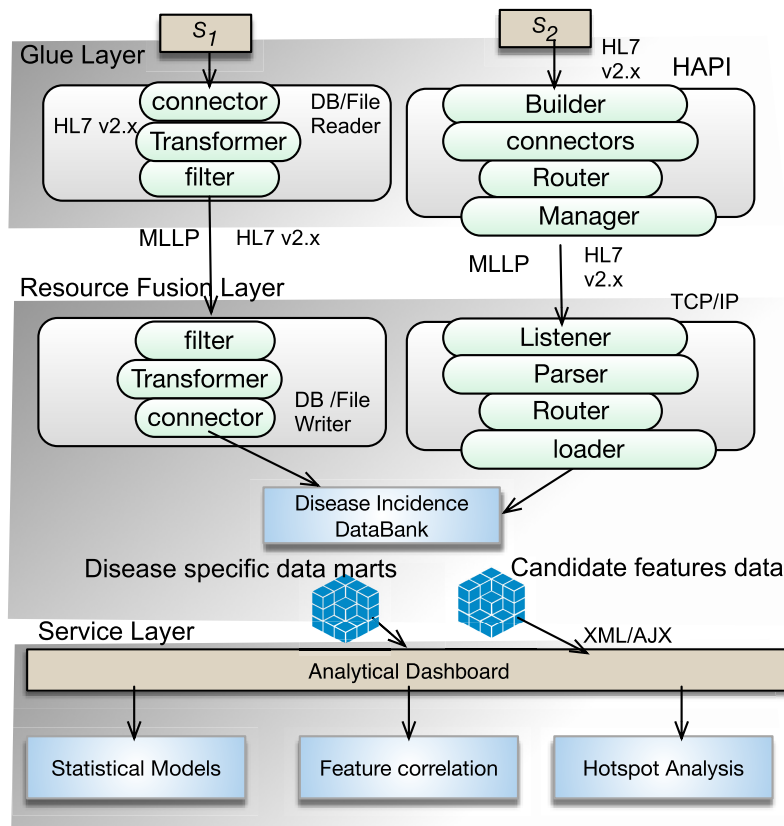
HL7 [4] is a set of standard, formats and definitions for exchanging electronic health records (EHR) and is accredited by American National Standards Institute (ANSI). The 7 seven in HL7 represents the seventh layer in Open System Interconnection (OSI) model where structure formatting and exchange of messages occur. HL7 standard defines a standard format that allows messages to be transmitted between disparate healthcare systems. It recognizes that all health care facilities are not identical and there cannot be a single standard protocol model to represent the data exchange process between these facilities. Therefore, it provides a broad range of messaging standards that can be applied to large scale health facilities, major hospitals as well as to stand alone diagnostic centres, laboratories, and clinics. It is now a widely accepted e-Health standard and is considered efficient for health information exchange. The HL7 V2.x [5] offers messages models, documents and services that cover the health data integration, retrieval and exchange requirements of most of the public health agencies.

The work presented in this article is based on HL7 V2.x standard specification for clinical health information exchange as it is currently, the most widely used standard. Most of the health information exchange work in HL7 has focused on clinical level. However, the alarming growth rate of epidemics, pandemics, infectious disease agents and dynamics around the globe requires that these standards be acquired for population health monitoring and control. Our prior work details the framework as a multi layer architecture for privacy compliant and standard based acquisition, transmission and integration of health data for evidence base disease management [6].

We present an HL7 based disease surveillance and analysis model that is implemented as a disease data exchange model from source such as hospitals and laboratories to destination including public health agencies for case reporting necessary for research in public health decision making and disease prevention and control intervention programs. We present implementation of the transmission protocol of dengue disease data from multiple sources and analysis methods to identify hotspots for dengue incidences based on data from all provinces in Pakistan. The data becomes directly interpretable by the public health agencies as compared to the surveillance data in multiple formats received via email, faxes or csv files. This work presents the dengue case reporting using the presented transmission model as well as statistical and spatial analysis for hotspot identification. Therefore, this study is able to demonstrate the use of information technology tools available in other domains for achieving public health objectives.

The presented work a multi-disciplinary effort in order to achieve public health targets. First, we present a successful implementation for an HL7 standard based transmission of disease data in places where minimum digitization or standard compliance is practiced. We provide a pathway with semantic and implementation details for transmitting data of public health interest from multiple heterogeneous sources and present its availability to public health agencies and research facilities for further analysis and decision making. The presentation of this implementation will help to understand the validity of data sharing ability and models for decision making process. Using HL7 based transmission protocol, we reduce the ambiguity and incompatibility among different data sources. Second, we present a dengue case study in the context of standard based acquisition and transmission framework for 3 provinces of Pakistan and capital city Islamabad, from year 2014-2017. To the best of our knowledge no prior work has performed a spatial analysis of this scale over a standard framework. The implementation shows that active disease surveillance is possible over legacy systems which can result in timely and appropriate allocation of resources as well as implementation of control and preventive programs [7]. In addition, the analytical results provides a reference for selecting facilities based on cluster analysis for implementing sentinel surveillance that can provide detailed data over a population while using lesser public health resources.

With public health informatics becoming more sophisticated, it is important to provide an objective evidence base for population health analysis instead of aggregated information. The current surveillance techniques in many developing countries, including Pakistan still remains heterogeneous and fragmented with varying data reporting formats due to lack of standardization. A standard based transmission protocol is presented and implemented in this study to provide a transmission channel from data source to data analytics platforms for public health authorities. This pathway allows near real time reporting of diseases cases to public health authorities as early as they are diagnosed with the disease or are admitted with a diagnosis. HL7 is a relatively young specification for health data exchange and its available application is mostly for clinical data exchange among heterogeneous EHR in healthcare entities to maintain a common patient profile and to share patient data. We demonstrate that this data interoperability specification can be applied in public health data acquisition and transmission settings where minimum digitization is present (in low and middle income countries, e.g. Pakistan) where infrastructure for complete interoperability is limited as well as in developed countries where at the time electronic health record systems or hospital information systems were developed and implemented, the standards of health data exchange were still in their infancy. Therefore, this work may also be useful where interoperability middle ware solutions are required to exchange health data for public health informatics.



**FIGURE 1.** The overview of process of acquiring dengue incidences data from two sources and transmitting to server through a multi layer framework for dissemination to concerned health authorities for disease dynamics analysis.

## II. RELATED WORK

In developing countries, there is still need for digitization of health records. The reportable disease data is collected mostly from public hospitals that do not have standard EHR systems installed. The efforts being done in context of public health informatics are localized and non standard in most cases [8]–[10]. In order to establish an effective dengue control and prevention mechanism in a region, it is important to acquire and transmit case reports in real or near real time instead of passive surveillance. The surveillance data acquisition and flow in Pakistan has been previously defined [6]. In context of dengue surveillance, while it is a reportable disease, the data collection mechanism remains fragmented and vertical from regional to national level as is common in many other developing countries [11], [12]. HL7 is an internationally recognized not only as clinical data exchange standard [13]–[16] but also in disease surveillance for public health management [17]–[19]. However, fewer examples in literature discuss the implementation stages for enhancing passive surveillance using HL7 interoperability standard [20]–[22].

Several epidemiology and other factors have been investigated affecting dengue outbreaks [23], [24]. Candidate features of interest or determinants disproportionately affects the dengue disease spread. It has become increasingly important

to investigate these candidate features including climatic and non-climate factors for control and prevention of disease in vulnerable populations. The process involves identification of dengue hotspots, complex relationships patterns between the disease incidences and its transmission that drive the spread of dengue fever globally [25]. Dengue disease dynamics have been studied along with multiple factors including demographic, climatic, topographical, geographical, migration and traveling [26]–[30]. To build an effective model for predictive and AI based analysis, strongly correlated factors with disease incidence rate have been investigated. Many studies have reported strong influence of temperature and rainfall on dengue transmission [31]–[33]. Several potential reasons have been identified for this influence. For example, warmer temperature facilitates the survival of disease causing mosquitoes, increase their reproduction, growth and transmission rates. Similarly, precipitation or rainfall creates potential breeding sites for mosquitoes and is shown to have a positive correlation with disease incidences [34]–[36]. However, some research has also shown negative correlation and have attributed their results to removal of breeding sites due to excessive rain [37], [38]. On the other hand, using time lags based analysis, it has also been shown that hot and dry weather followed by heavy rainfall also increases dengue

incidences [39]–[41]. Therefore, the relationship of climatic factors and dengue outbreak is not as straight forward and needs further investigations by analyzing dengue dynamics in populations with high incidence rates to understand the global transmission dynamics of dengue. Further, the identification of clusters over spatio-temporal scales is crucial to understand the disease dynamics at a location. While some diseases maybe randomly distributed, most communicable diseases tend to form cluster over space or time. The presented analysis and their results serve a useful reference for further investigations into the geographical location for potential causal relationship of dengue cases with other factors. For example, using network analysis such as distance from water bodies, construction sites and other breeding sites for dengue mosquitoes to understand the complex disease dynamics can be undertaken based on the analysis of results of the study.

GIS based spatial analysis is becoming increasingly popular for management of infectious diseases [42]–[44]. Researches have applied mathematical models, AI based analysis and other techniques in order to study dengue disease dynamics and in terms of locations and external factors [45]–[49]. However, the spatial analysis is strongly influenced by the scale. Many studies fine scale the results to city, district or town administrative level [50]–[52], while some use point time series data to analyze the outbreak pattern [53], [54].

Pakistan faces enormous dengue burden each year [55], [56]. The disease outbreaks have been attributed to multiple factors ranging from climatic change to natural disasters such as flooding and earthquakes [57], [58]. Due to its endemic nature in Pakistan, several studies have been conducted to model and spatial temporal dynamics for control and prevention of the disease [59]–[61]. Presently, to the best of our knowledge, no recent study has been conducted correlating temperature, humidity, and population density with dengue case incidences in Pakistan. In addition the data used in previous studies is of a high spatial scale or localized, taken from a single hospital or city.

Studying disease dynamics in multiple varied environments and analyzing their results can lead to new scientific evidences that are an important reference for further scientific studies. The importance of studying features of diseases over different areas cannot be undermined. To make the process more meaningful, health data needs to be exchanged through a standard based pathway.

In this paper, we provide an interoperability framework implementation experience for studying dengue disease dynamics in any given population. This framework can be generalized to include any set of diseases of public health interest, e.g communicable diseases. The presented work is able to demonstrate the use of interoperability tools to inform public health authorities for effective decision making by using methods to provide insights into diseases over a geographical area e.g. using hotspot analysis. This is possible due to implementation of an interoperability framework that carried disease data directly from hospitals to public health

level case to case instead of aggregated passive forms. This is highly required in many low and middle income countries where there is a lack of interoperability infrastructure. In this work we attempt to uncover challenges of placing an interoperability standard over existing infrastructure and present case of dengue disease spread studies through data transmission over the proposed pathway. In addition, the presented work is aimed at finding new evidences in varied environments for diseases spread through exploratory analysis.

### III. METHODS AND MATERIAL

#### A. STUDY AREA

Dengue is endemic in Pakistan with its usual peak in Monsoon season. Despite of government efforts thousands of cases have been reported from all across the country every year. The climate in Pakistan varies from hot summers to cold winters with variations between the two according to locations. In general winter lasts from December to February, spring from March to May, hot summers and rains from June to August and a monsoon season from September to November [62]. We study the dengue incidence distribution patterns for three provinces, Punjab, northern province Khyber Pakhtunkhwa (KPK), south east province Sindh and Islamabad Capital Territory (ICT). Punjab is the highest populated province of Pakistan and has seen highest dengue incidences from the study time period. In 2017, KPK experienced an epidemic of dengue, with more than 18000 cases with high number of deaths [63].

#### B. DENGUE DATASET

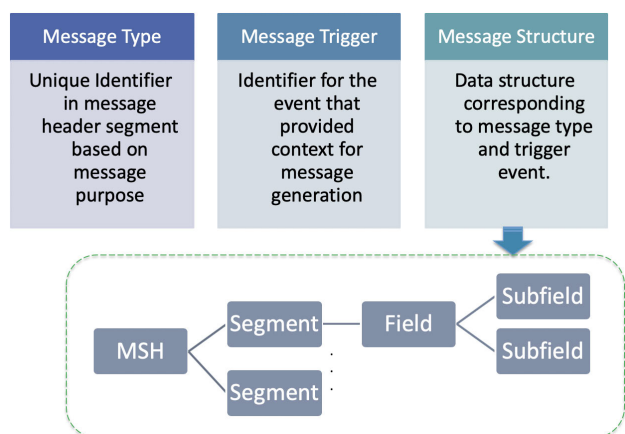
The dataset used in this research process is obtained from National Institute of Health (NIH) [64], Directorate General of Health of Provinces (DSCP) [65]–[67] and Dengue Surveillance Cell Punjab [68]. It includes dengue confirmed cases data from all districts in the 4 provinces of Pakistan from year 2014 to 2017. The cases are confirmed positive from first clinical diagnosis or from laboratory diagnosis. The data includes demographic details and case incidence information for the afore mentioned time frame and locations. The data is mapped to HL7 standard messages. The locations are geocoded to find latitude and longitude for meaningful spatial analysis of data. The disease incidence data is augmented with contextual features of interest. These include weather data attributes and population data. Weather data for affected areas is obtained from Pakistan Meteorological Department (PMD). The dengue incidence data set consists of 42,239 cases reported from the study areas.

#### C. DATA MAPPING

The HL7 messages are composed of segments that are further categorized into fields where actual data is stored. Each message has a type and a trigger associated with it. (Fig.2). There are more than 183 segments included in HL7 v2.x. When an event occurs such as patient admission, patient transfer, patient discharged etc., it defines a trigger and is represented

**TABLE 1. Segment level mapping profile for HL7 ADT-A01 messages from  $S_1$  and ORU-R01 messages from  $S_2$ .**

HL7 Segment	Description	Data Element	Description	Description	Data Element
SYNDROMIC DATA ELEMENTS OF INTEREST (ADT Messages)			LABORATORY RESULT MESSAGES (ORU Messages)		
PID	Patient Identification	Patient Identifier Sex, Age, Mohallah, UC, District, Province, Zipcode, Contact no	PID	Patient Identification	Patient Identifier, Sex, Age, Mohallah, UC, District, Province, Zipcode, Contact No.
PV1	Patient Visit	Reporting Hospital/ District, Datetime for encounter, Status of Discharge	OBR	Observation Request	(NS1, PCR), Datetime for report
PD1	Patient Demographic	Occupation, Ethnicity	OBX	Observation Segment	Test performed, Results, Datetime for tests, test status, Labortray Name
OBX	Observation Segment	Chief complaint, Onset of symptoms, Height, weight, Smoking status platelet count at admission, Fever, Body aches, Bleeding, Skin rashes, Travel history			
DG	Diagnosis	Status ( Admitting, Underprocess, Final)			



**FIGURE 2. The HL7 message structure.**

by a particular type of message. Messages can have a single trigger and thus a single format (e.g. VXU message) or they can have multiple triggers and formats (e.g. ADT message). Segments in the messages can be optional, repetitive or both. Field are the components of message holding the data content of the message. Field may have none or several sub fields. The glue layer with an HL7 agent and HL7 gateway is implemented at health care facilities including hospitals and laboratories from where the disease cases are reported. The data reported may be in any form including spreadsheets, aggregated sums, and non digitized information. For every new data source added to the framework, a glue layer with HL7 agent need to de designed according to the source specification. Data mapping is a challenging task that involves understanding source data context and how it is being stored in the data base and what it represents. Therefore, when a mapping specification is defined for a source, the types and formats of data involved and their place in HL7 messages must be clearly defined. Consequently, it involves a complete knowledge of HL7 specification as well as the working of source legacy system. The glue layer in the framework that resides at the data source facility, is responsible for extracting

the case based information of public health interest required from the source. The data is queried from HL7 compliant database and mapped to HL7 specification for transmission to a central public health server in form of HL7 messages. In cases where HL7 compliant form of data is not present at the source, an HL7 agent is implemented that creates HL7 messages based on source specific attributes mapping to HL7 specification. In this study the dengue surveillance data is mapped to HL7 segments based on mapping specifications summarized in Table.1. Table.1 gives the segment level specification for mapping of source specific attributes to HL7 messages. We use two types of messages for transmitting data, ADT-A01 (Admission, Discharge, Transfer) messages used to transfer data from  $S_1$  data source and ORU-R01 messages to transfer observations and results from test laboratory at  $S_2$ . At the receiver end, the received HL7 messages are fused in the underlying database. Missing values present in the original data are transmitted as the empty fields in the HL7 messages. For analysis purposes the missing value are dropped from the data set. For example, patient address may not be present, which will lead to unavailability of point location for spatial clustering. Therefore, records with missing or out of range values are removed from the data.

**D. DATA TRANSMISSION PROTOCOL**

Information moves from multiple layers in the framework following an HL7 based protocol. The data from multiple sources and on multiple diseases is made available to concerned authorities in the form of data marts that are spawned from the underlying public health consolidated data bank. This data is combined with other candidate features of interest that are collected and transmitted by other means and serve as context features for a given population. The details of components of the framework and the processes involved are presented in. For the purpose of this study, we present the HL7 mapping done at the dengue data source glue layer end the analytical processes employed for feature selection and hotspot detection. We have implemented the framework using two methods, first, HL7 compliant database quering and MirthConnect v3.4.1.8057, and second, using

HAPI (HL7 application programming interface) for creating client and server programs in Java for HL7 message transmission and receiving. MirthConnect is an open source cross-platform HL7 engine for bi-directional transmission of multiple HL7 messages [69]. HAPI is an open source HL7 library and parser for Java [70]. The overview of the transmission of dengue data from sources to analytical dashboards is presented in Fig.1.

Mirthconnect uses channels for connecting source to destinations. For dengue case study we create channels at two sources, one for DSCP data source ( $S_1$ ) and second for NIH data source ( $S_2$ ) for HL7 message transmission. At ( $S_1$ ) communication channel continuously listens for any incoming HL7 message at a predefined port. When an HL7 message is received, a filter validation process maps all required data attributes of interest based on a predefined mapping. At receiving end, the filter validation checks for the presence of all required data attributes and transformers reads the data received into single attributes in the incoming channel that are then mapped to the server's local database. Thus, mapping at source and destination are performed inside the channel. The difference between the two sources data transmission is the presence of an HL7 compliant database at ( $S_2$ ). In this case the Mirthconnect connectors can directly read the database into the HL7 message based on the transformer mappings. For data source ( $S_1$ ), first, HL7 mapping is done by an HL7 agent and sent to the HAPI connector for transmission by the agent.

The data from provinces is acquired and mapped to HL7 messages using HAPI, an open source java based HL7 API in Eclipse IDE environment. An HL7 message is created and managed using HAPI context that controls the configuration and instantiation of HAPI objects including a server listening to incoming HL7 messages and transmitting over a network as well as a client that receives the HL7 messages. HAPI also allows message encryption and process threads giving more control over data transmission protocol. Further more, HL7Service interface in HAPI allows multiple message types to be received using multiple channels.

## E. DATA ANALYSIS

The data transmitted from all sources is provided to the public health authority dashboard using AJAX that allows data to be updated asynchronously. Data received is filtered for the required attributes. For example, for our study, District, Province, Gender, Reporting date and patient address are extracted from the message. Data was preprocessed for misspelled location names, missing addresses and converting addresses to obtain fixed longitude and latitude values. For this purpose, we used OpenRefine tool. For the received dengue dataset, at the analytical layer, we use multiple linear regression model to find the relationship between candidate features of interest including temperature, humidity and population density with dengue incidence data by fitting linear equation to recorded data. This gives a correlation of disease incidence to candidate features of interest. Our spatial

analysis of dengue is based on two hypothesis, first, any patterns in the disease dynamics can be identified using the near and recent rule; and second, there are multiple factors involved in the disease propagation. The candidate features of interest are selected and analyzed with disease incidence data. We apply spatial clustering to identify vulnerable sites of dengue outbreak during each year and study the effects of external factors on number of cases in these identified hotspots.

Spatial pattern analysis of the disease incidences is conducted to identify disease hotspots. The hotspot identification process involves clustering disease incidences based on incidence location for spatially close incidences. In contrast to traditional aggregated data analysis, we employ spatial point processes for dengue incidence modeling. The spatial point process assumes that an incidence occurs at a random well defined point in space that must be unambiguous with respect to other points [71]. We are able to represent the disease cases as a set of discrete incidences occurring at a random point in space defined by their latitude and longitude and cluster spatially close incidences to determine a hotspot at a given area on map [72]. This helps us to identify locations to study other features of interest for the hotspot areas and serves as reference for predictive analysis.

Formally, we define each disease incidence  $DI$  as given by Equation. 1:

$$DI_{jk} = (x_j, y_j, t_k) \quad (1)$$

where  $DI_j$  is an instance of disease incidence at a location, defined by a pair of coordinates  $(x, y)$ , and  $t$  is the time of the occurrence incident. For hotspot detection mechanism, we employ partition based clustering to our spatio-temporal data. The disease incidences for each year are clustered independently to identify hotspots. Therefore, the data is sliced into 4 windows  $w_1, w_2, w_3$ , and  $w_4$ , each of 12 months period starting from 2014 to 2017. Each cluster represents a dense disease incidence area with a centroid location based on mean distance that we call a hotspot. The disease incidence stream is clustered into  $m$  clusters where each cluster  $C_l$  with  $l = 1, \dots, m$  is defined by a tuple as in Equation 2.

$$C_l = (id_l, HS_l, DI'_l, s_l, w_t) \quad (2)$$

where  $id$  is the cluster identification,  $HS$  is the centroid or hotspot in the cluster,  $DI'$  is the subset of  $DI$  that belongs to cluster  $C$ ,  $s_l$  is the size of the cluster  $C_l$  and  $w_t$  is the time window of the disease incidences where  $t = 1..4$ . Each hotspot is defined by a pair of coordinates  $(x, y)$ , that is,

$$HS = (x, y) \quad (3)$$

where  $x, y$  are longitude and latitude representing location of the hotspot respectively. A population density based threshold is then applied over each window to identify categories of hotspots ranging from most significant hotspot to coldspots over the study area based on cluster density.

In order to study the relationship between the multiple candidate factors, we use Pearson correlation based candidate feature selection. Pearson correlation coefficient works to measure the strength of relationship between candidate features of interest. The coefficient of determination or R-Squared is calculated to determine the variability of candidate features with respect to number of dengue incidences in the MLR analysis. Visually, the strength of correlation is higher if a point representing sum of squared distances lies closer to the regression line. Multiple Linear Regression (MLR) also called Multiple Regression (MR) method has been applied to predict the level of affect candidate features have on dengue incidence. Thus MLR models the linear relationship between independent variables, that is, candidate features and dependent variable, that is, disease incidences. MLR is based on Ordinary Linear Regression (OLR) and is able to model the process using a single regression equation given by equation. 4

$$DI_i = \beta_0 + \beta_1 CF_{i1} + \beta_2 CF_{i2} + \dots + \beta_f CF_{if} + \epsilon \quad (4)$$

where  $i = 1, \dots, n$  with  $n = \text{number of observations}$ ,  $DI$  is the predicted of the expected disease incidence,  $\beta_f$  is the correlation co-efficient for each candidate factor and  $\epsilon$  is the residual error. The model evaluation is based on multiple criterion including, statistical significance of the selected candidate factors, non-redundancy in the candidate features, normal distribution and non-spatial correlation of residuals, and R\_squared value. The closer the value of R-Squared to 1, the better the MLR prediction model will fit. The R-squared value equal to 1 would indicate that all the variability in the number of dengue cases can be explained by temperature and humidity only, which is rarely a logical case.

In addition, for interactive exploratory analysis, use grid-based clustering technique to identify hotspots for the dengue incidences. The visual display of incidences on map allows zooming to different levels. The zoom level defines the number of grids the map will be divided into. The incidences are clustered based on spatial closeness of disease incidences in each square grid. That is, the disease incidence at a particular location and incidences is selected and incidences closest to the selected incidence are mapped to the same cluster. The measure of closeness is the distance between the clusters. The grid size can be customized to different values and The process continues until all disease incidences are allocated to some cluster and the results are displayed on the google maps. Disease incidence can also be customized to be selected as a reference for finding hotspots near it. This exploratory hotspot analysis of a disease helps to investigate the disease dynamics in a given area. For example a dense cluster reflects a very significant hotspot and environmental features of the given area can be compared to analyze potential causes of increased number of cases.

#### IV. RESULTS

The data is extracted from the sources based on preliminary studies for dengue public health features of interest and

selected and transmitted using two strategies, mirth connect channels, with data base readers and writers, HL7 agent based message transmission and HAPI based protocol implementation. Fig.3 shows a sample ADT-A01 message created by mapping source attributes to HL7 fields when a dengue patient is admitted to the hospital. The HL7 messages are transmitted to a central test server and stored in a population data bank. When using MirthConnect channels, we were able to transmit HL7 messages without having to implement network administrative tasks such as connection, queue and thread management or error logging and handling. However, when connecting to multiple heterogeneous sources that are not HL7 compliant, interfacing with mirth connect will require more scripting to create HL7 messages from the source data bank. The source transformers also allow changes in data format. On the other hand, with the HAPI implementation, there was more control over the configuration and management of network, encryption and multiple message transmission. However, both communication servers used were successfully implemented to realize the standard based health data transmission for public health analytics.

The dengue incidence data received at the server end and stored in an SQL relational database is queried for patient address to represent case incidence location stored in PID segment of HL7 message, district stored in PID segment of HL7 message stored in PID segment of HL7 message, province stored in PID segment of HL7 message, patient gender stored in PID segment of HL7 message and case reporting date stored in PV1 segment of HL7 message as shown in Table.1. The PID segment contains patient specific information whereas PV1 segment contains event or visit specific information of a patient. The patient address is geocoded to obtain spatially meaningful location on the map. For point process based analysis the data is fetched from the database into XMLfiles in PHP.

For the purpose of this study, we use the patient address, gender, and date-time of incidence from the incoming data. The statistical analysis of received dengue data provides following results. 22628 cases were reported in 2017. In 2014, 2015 and 2016, the total numbers of cases were 1991, 9899 and 7721 respectively. In years 2014-2017, male and female affected are 73.1% and 26.8% respectively. The initial disease spread statistics are presented in Table 2.

In order to study the impact of temperature changes on dengue incidences, we focus on the rate of change of monthly dengue cases with temperature changes within each province. Fig. 4 shows the number dengue cases in all 12 months of the year for each year bracket in 3 provinces and federal capital of the country. The results show that the months of August, September, October and early November, are favorable for Dengue in Punjab, KPK, and Islamabad. These months form the post monsoon period in these areas. Sindh is seen more affected by the disease in months from October till January.

The spatial hotspot analysis revealed multiple hotspots through out the country. We applied our hotspot detection algorithm for each years data to identify locations with

TABLE 2. Dengue Distribution across the study area from 2014-2017.

	2014			2015			2016			2017		
	Cases	M	F	Cases	M	F	Cases	M	F	Cases	M	F
Islamabad	96	65	31	526	353	173	2813	1828	985	296	192	104
Punjab	1314	972	342	5242	4139	1103	2745	2012	724	418	307	111
Sindh	114	82	32	1818	1309	509	1245	896	349	2692	1952	740
KPK	467	341	126	2277	1662	615	918	670	248	18784	13731	5053
Total	1991	1460	531	9863	7463	2400	7721	5406	2306	22190	16182	6008

```
MSH|^~\&|Our System|Patient Health Data|Their Remote System|Patient Health Data|20190110185401||ADT^A01^ADT_A01||123420190110185402|P|2.4
EVN|A01|20190110185401
PID|||"100"||"Shams ur rehman"^^"Nigat"|||"F"|||"Gujran""Tehkal Payan 1"^^"Peshawar"||"^^^^^^"3005972274"
PD1|""
NK1|^"Shams ur rehman"
PV1|^"KTH "||||||||||||||||||||||||||||||||||||||||"Admitted"
OBX|"Yes"- "No"- "No"- "No"|||"2017-07-26"
```

FIGURE 3. The sample ADT-A01 message for single dengue case transmission based on mapping of source attributes to HL7 fields.

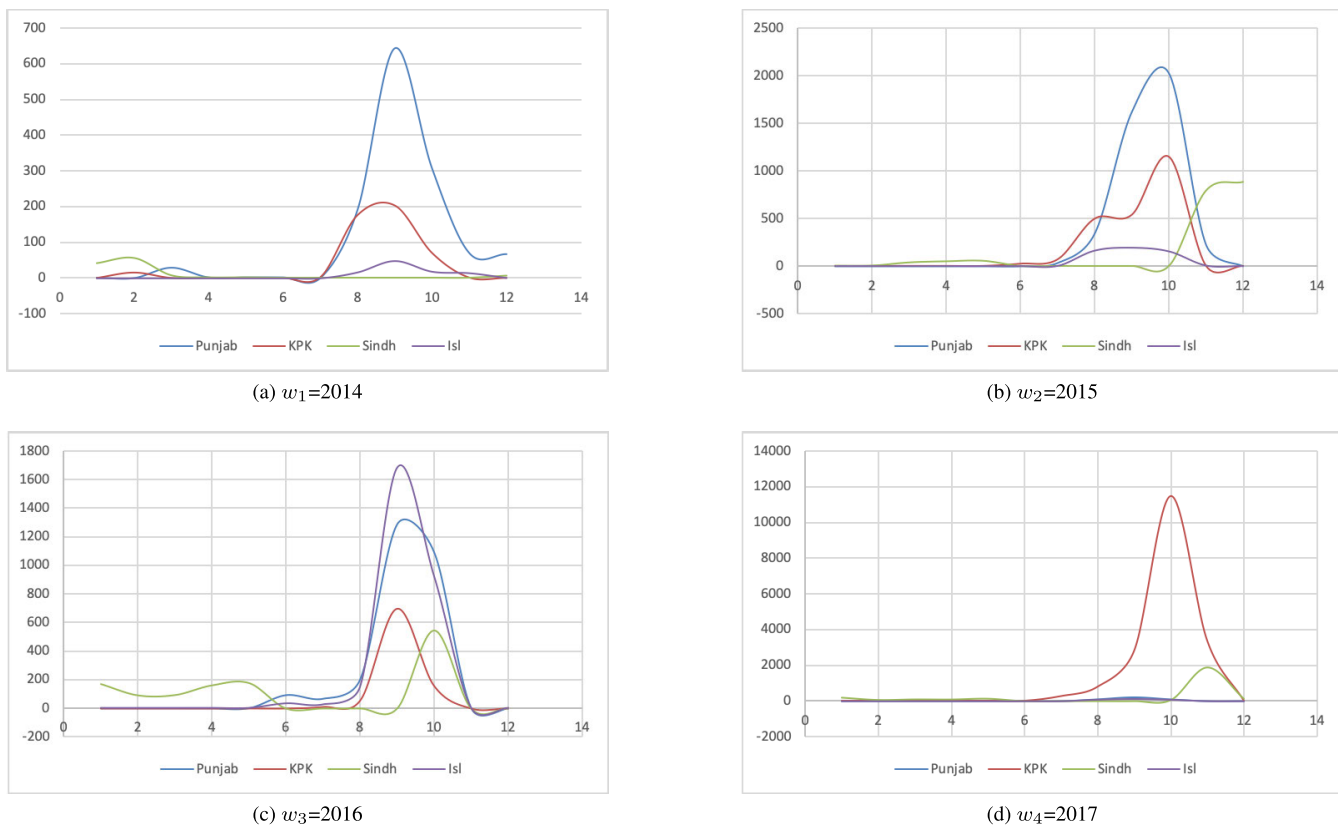


FIGURE 4. Monthly dengue incidences in 3 provinces Punjab, Sindh, KPK and federal capital Islamabad, Pakistan. The x-axis represent the 12 months of the year where as y-axis shows number of dengue incidences. The results show that maximum number of disease incidences are reported between August to October for Punjab, KPK and Islamabad whereas Sindh experiences peak during October to December in all years.

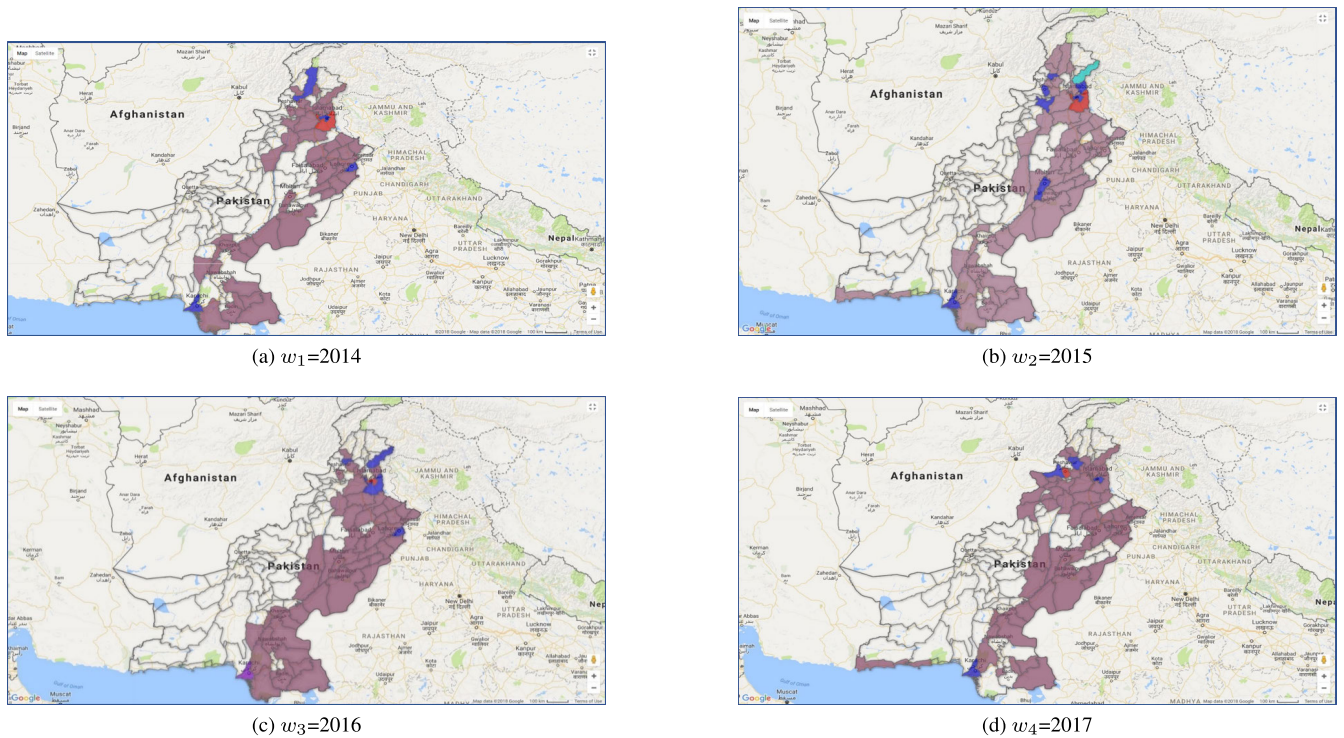
dengue incidences spatially closer to each other forming a cluster. The clusters are formed based on geocoding of street addresses of patients. The districts of these hotspots are identified. The identified hotspots districts and cluster size for each is presented in Table 3. The hotspot analysis

shows districts Malakand, Peshawar, Abbottabad, Mansehra and Mardan of KPK province to be most affected over the selected time period. These districts are more developed and urbanized as compared to other less affected ones. The threshold is applied to find districts for most significant



**TABLE 3.** Administrative districts for identified hotspots over the study areas for each time window  $w$ . The size represents the cluster members  $s$  at each hotspot representing dengue incidences.

Districts	Size	Districts	Size	Districts	Size	Districts	Size
$w_1$		$w_2$		$w_3$		$w_4$	
Karachi	72	Karachi	1741	Karachi	955	Karachi	2661
Rawalpindi	1111	Rawalpindi	3303	Rawalpindi	1160	Rawalpindi	213
Lahore	97	Lahore	146	Lahore	1110	Mardan	322
Swat	78	Multan	365	Mansehra	295	Peshawar	17670
Sheikhupura	60	Mansehra	589	Islamabad	2813	Islamabad	296
Malakand	130	Malakand	397	Malakand	263	Malakand	236
Abbotabad	65	Abbotabad	493			Khyber Agency	330
Islamabad	96	Islamabad	526				
		Peshawar	383				
		Kohat	310				



**FIGURE 5.** Most significant hotspots identified in red, significant hotspots identified in blue and aqua blue, cold spots in brown for each window  $w_1 \dots w_4$  over the geographical map of the identified districts for the hotspots.

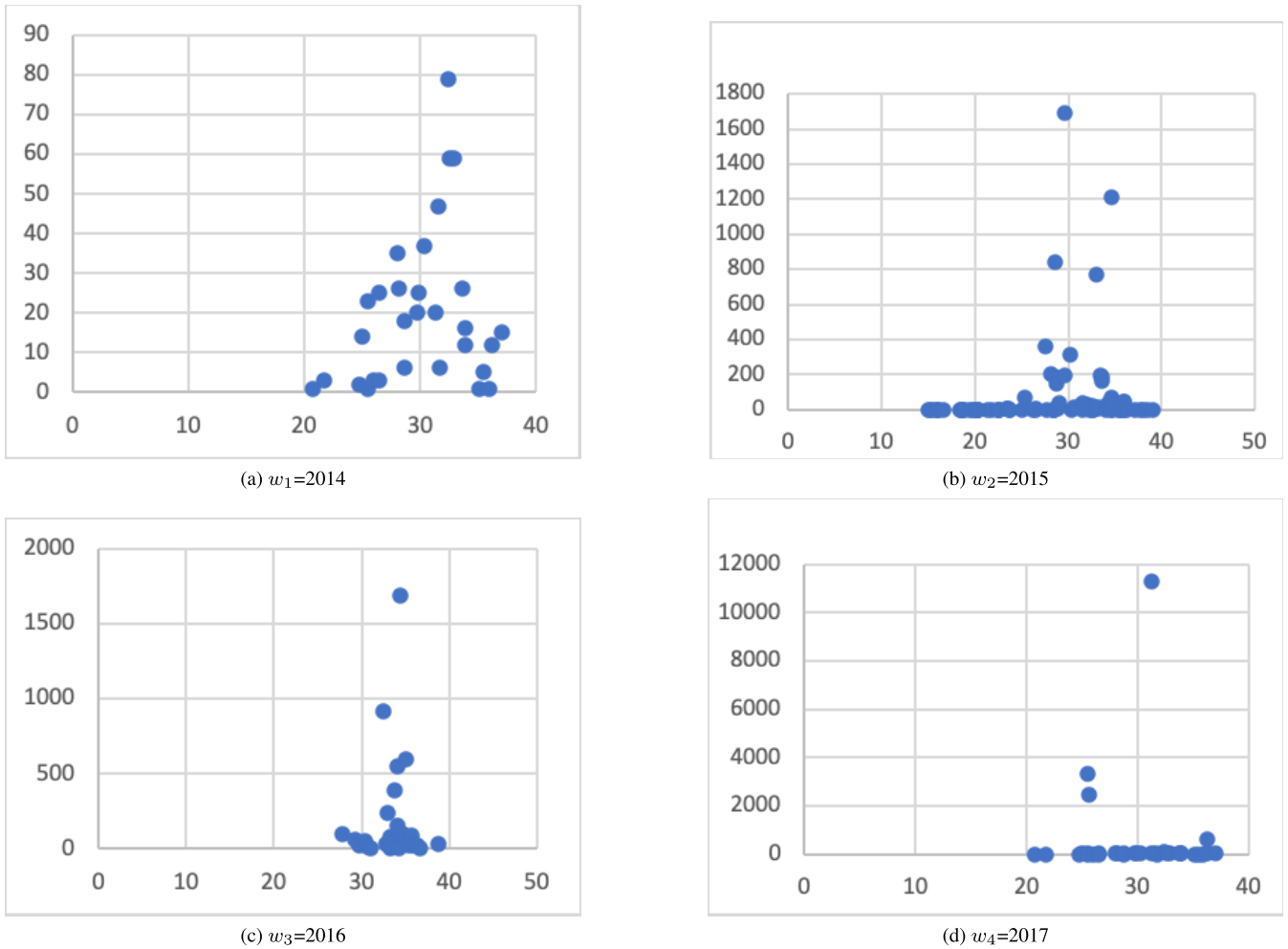
**TABLE 4.** Regression statistics for MLR.

Regression Statistics	
Multiple R	0.164106271
R Square	0.026930868
Adjusted R Square	0.019115052
Observations	252

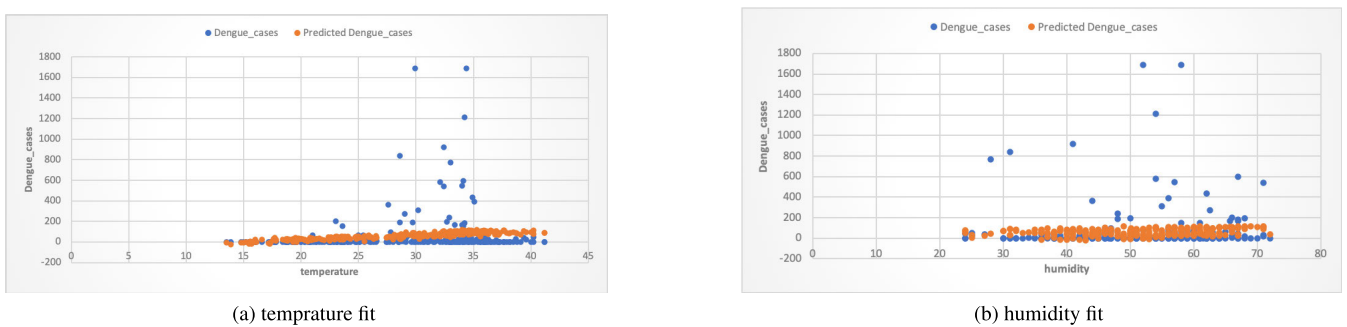
hotspots, significant hotspots and cold spots based on the the cluster density for each window size  $w$ . Fig.5 shows the geographical locations for the hotspot districts on the map. Rawalpindi is identified as the most significant hotspot for  $w_1$ , where as Karachi,Lahore,Swat,Malakand and Islamabad are identified as significant hotspots for  $w_1$  and shown in blue (Fig.5a). The districts shown in brown color are identified as cold spots based on cluster density. For  $w_2$ , Rawalpindi

is the most significant hotspot, and Karachi, Multan, Lahore, Peshawar, Mansehra, Malakand, Abbottabad and Islamabad as significant hotspots (Fig.5b). Similarly, for  $w_3$ , Islamabad is found as most significant hotspot. Rawalpindi, Lahore, Mansehra and Malakand identified as significant hotspots shown in blue and Karachi as high significant hotspot in purple color (Fig.5c). For  $w_4$ , Peshawar forms the most significant hotspot and Karachi, Rawalpindi, Mardan, Malakand, Khyber Agency and Islamabad are categorized as significant hotspots (Fig.5d).

The hotspot districts identified are further analyzed for candidate factor correlation. We use temperature and humidity as candidate features and find correlation with the dengue disease incidence in significant hotspots districts. The R-squared value calculated is 0.41 that shows a positive relation between temperature, humidity and dengue incidences.



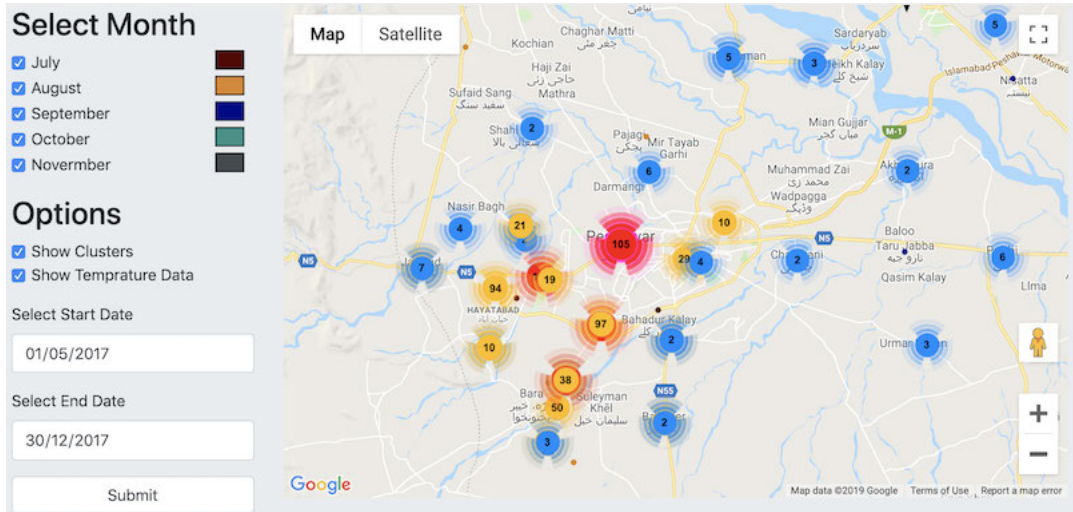
**FIGURE 6.** Relationship between dengue incidences and temperature values. The x-axis represent the temperature values in °C; y-axis gives dengue incidences. It can be seen that highest number of dengue cases were recorded for temperatures between 30-37 °C.



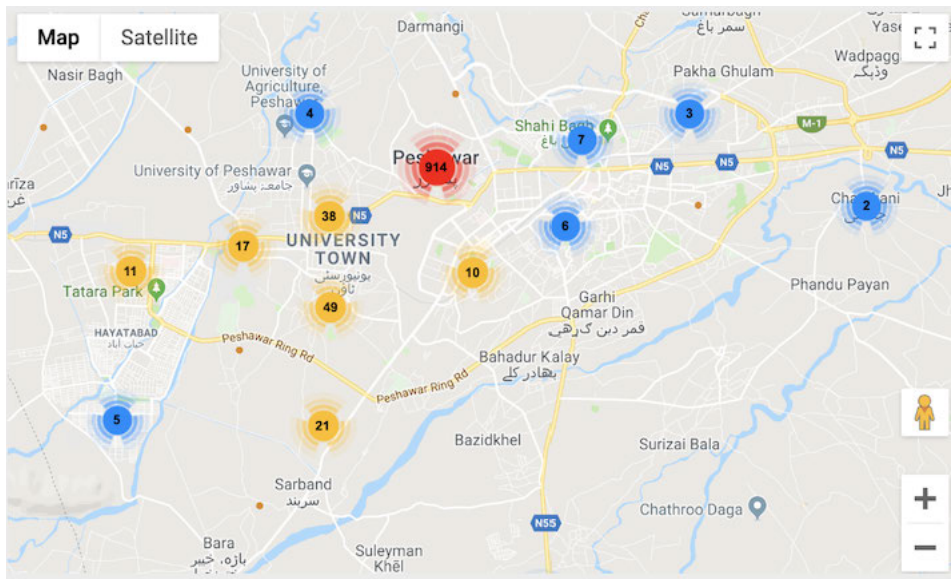
**FIGURE 7.** Monthly dengue incidences vs climatic factors fit plot for predicted vs actual number of dengue cases.

A very strong relation can be analyzed between temperature and the dengue cases. Multiple Linear Regression is performed to see the relation of temperature and humidity on dengue cases and results shown in Table. 4. Mean monthly maximum temperature and humidity are taken as independent variables while corresponding Dengue cases as dependent variable. The results show positive correlation val-

ues for temperature (0.445) and humidity (0.245). In addition, dengue incidences also show positive correlation with the population density in the hotspots. The correlation analysis gives 0.48,0.39,0.48 and 0.41 values from year 2014-2017. Therefore, the densely populated areas are more prone to the disease. We find the f-statistic to determine if the regression relationship exist. Table 6 and p-values in Table. 5 shows



**FIGURE 8.** The exploratory graph showing spatio-temporal clusters of dengue incidences in KPK for selected months in year 2017. The cluster size shows number of neighbouring incidence in the selected spatial grid scale. Colors represent clusters in different selected months.



**FIGURE 9.** Fine scaled hotspot identification in KPK for identification of other potential factors affecting high disease incidence rates. The Metro-Bus project in Peshawar was identified as a major factor through this analysis since the hotspots lie close to the project areas.

a significant relationship between dengue cases exist as p-value is less than 0.05. The impact of temperature on dengue cases is investigated by acquiring mean monthly maximum temperature for the identified hotspots. The hotspot districts were analyzed against the temperature data to study the temperature and dengue cases relation. Fig. 6 shows the relationship between temperature and dengue incidences. It can be seen by observing results presented in Fig. 6 that for all the four years, highest number of the dengue cases are reported where temperature ranges from 30-37 °C in the study areas in Pakistan. Fig. 7 give the predicted vs actual cases for humidity and temperature variations. Table 5 gives the descriptive statistics based on MLR analysis. The large value of MS shows that there more unexplained variation

in the data. This leads to employing an exploratory analysis for identification of potential variables in the dengue prevalent populations to be added as independent variables in the regression model.

We further performed an exploratory cluster analysis for the hotspot districts throughout the country. The sample snapshots for the KPK province is shown for 2017 dengue incidences in Fig.8 and Fig.9. Since KPK was the most hard hit province for dengue outbreak in 2017 from all the available time frame, with more than 18000 cases in a single year, we explored the locations with hotspots of significant sizes and found the following results. Areas adjacent to Tehkal, Pishtakhara, University road, Mithra and Kohat road were identified as the significant sub hotspot locations. These

TABLE 5. Variable section for MLR.

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-151.4967459	91.027625	-1.664294173	0.097311518	-330.7790093	27.78551754	-330.7790093	27.78551754
temperature	4.76831784	1.891652458	2.520715589	0.012337371	1.042638582	8.493997097	1.042638582	8.493997097
humidity	1.47601692	1.227030131	1.202918235	0.230150495	-0.940664162	3.892698002	-0.940664162	3.892698002

TABLE 6. ANOVA.

	df	SS	MS	F	SF
Regression	2	290066.3772	145033.1886	3.44568849	0.03341105
Residual	249	10480710.62	42091.20732		
Total	251	10770777			

are the undeveloped areas lacking basic municipal amenities such as sewerage and drainage systems and cleanliness that may have eventually lead to provision of breeding places for mosquitoes. Therefore, the exploratory hotspot mining reveals that there are other potential factors than temperature and population density, affecting disease transmission that need to be included in the model to explain the dengue incidence rate differences in different hotspots each year. One strong speculation for increased number of cases may be the initiation and progress of major construction projects including Metro Bus project in the main cities of the provinces. These projects provide breeding sites for the breeding of dengue mosquitoes in humid and rainy season.

## V. DISCUSSION

When using channel for ( $S_2$ ), we connect with the source database and selected attributes that were required at the public health analysis layer. However, MirthConnect did not allow filtering of the data. This meant that disease case data already transmitted will again be sent along with the new data. This challenge can be overcome by adding a flag column in the source database that represents if the particular disease incidence has already been reported. This means the glue layer at the source need to be customized according to the source data model. The model remains scalable in terms of implementing glue layer only for every new data source added to the framework. The implementation process, in general, shows that the data transmitted through secure communication channels provides a standardized way of filtering, transmitting and transforming the disease incidence data, that is transmitted in real time and maybe less prone to errors.

Dengue and its impact in Pakistan is devastating. Dengue spatial pattern is highly intense in main cities of country like Lahore, Islamabad, Karachi, Rawalpindi, Peshawar, Abbottabad and Malakand. Dengue increasing trend is seen from August to mid-November in Punjab, Khyber Pakhtunkhwa and Capital Territory of Pakistan while from December to March in Sindh. The temperature favorable for the breed of dengue mosquito is noted to be in range 30-37 °C. Densely populated slums and unplanned urban areas are major hotspots of dengue disease. The hotspots identified in Lahore, Rawalpindi, Karachi, Peshawar, Abbottabad and Islamabad augment the findings of the study. Further, in KPK, it is seen that the urban centers with large population are more affected including Malakand, Peshawar, Abbottabad, Mansehra and Mardan. Urban environment is more favorable

to dengue spread as most of the population of the province resides in urban centers, hence they are naturally the most affected ones. In addition, due to the non-existence of diagnostic facilities, the cases in other districts largely remain unreported. It may also be concluded that a combination other factors is responsible for this trend.

For example, massive construction projects have harvested an enormous toll of the disease. Metro Bus project in Peshawar in 2017 can also be seen as a major reason for the surge in dengue cases. Similarly, Metro Bus project in Multan in 2015 can be attributed towards the high 361 cases for city of Multan that had an average of 3 cases per year. 2017 KPK dengue epidemic is attributed to a different lineage of dengue virus serotype 2 that entered in Pakistan from China [73]. Construction on smaller scale is equally dangerous as there are diggings involved and water accumulation results in spread of disease.

The increasing death count due to dengue in recent years has emerged as one of the most challenging task for health care authorities at provincial and national level. Every year, with approach of high rate incidence months in the country, increased financial, human and other health care resources need to be allocated for prevention and control of dengue. The research presented can be utilized well at multiple administrative level to identify and explore hotspots for prevention and control of different diseases. The resource allocation and preventive efforts can be well aligned with the more affected hotspots identified in the spatial analysis results of this work.

The presented work comes with some inherent data related limitations. For example, the spatial analysis can only be as fine scaled as the recorded data allows. In addition, as with other disease data, dengue incidence data also has unreported cases. In addition, the incidences with missing location information had to be removed in order to present a meaningful spatial analysis. Another limitation of the presented work is the non standardized acquisition of candidate features that can be taken up as a potential future investigation. For example, METeorological Aerodrome Reports (METARs) can be used to acquire, transmit and process weather data information. Standards need to be formulated and adopted by the public health agencies that support data transmission and archiving of public health interest. Additionally, the work can be augmented by introducing data from other sources such as construction work projects data, water sources data and urbanization data to further understand the disease dynamics in context with other candidate features of interest.

## VI. CONCLUSION

The methodology and results presented can help to identify significant hotspots across a region to implement sentinel surveillance. The work provides a reference for identifying and addressing challenges in implementing standard based data acquisition and transmission from heterogeneous sources to a public health management server. Identifying disease hotspots through a evidence based analysis provided through presented pathway can ensure appropriate allocation of resources. In addition, the spatio-temporal analysis presented in this work identifies dengue hotspots with variation in spatial and temporal scales. This helps in understanding the dynamics and relationship of dengue among populations and sub populations at a fine scale and having different densities with varying patterns of candidate features. Therefore, the presenting work provides a reference for moving from passive to active surveillance and presenting a direction towards adopting sentinel surveillance where required resources can optimally be allocated.

## REFERENCES

- [1] (Jan. 2019). *WHO-Pakistan: Health Management Information Systems*. [Online]. Available: <http://www.emro.who.int/pak/programmes/health-managment-information-sys%tem.html>
- [2] M. S. Qazi and M. Ali, "Pakistan's health management information system: health managers' perspectives," *J. Pakistan Med. Assoc.*, vol. 59, no. 1, pp. 10–14, Jan. 2009.
- [3] G. S. Birkhead, M. Klompas, and N. R. Shah, "Uses of electronic health records for public health surveillance to advance public health," *Annu. Rev. Public Health*, vol. 36, no. 1, pp. 345–359, Mar. 2015.
- [4] H. L. S. International. *Introduction to H17 Standards. Health Level Seven International*. Accessed: Sep. 2019. [Online]. Available: <http://www.hl7.org/implement/standards/index.cfm?ref=nav>
- [5] H. L. S. International. *H17 Version 2 Product Suite. Health Level Seven International*. Accessed: Sep. 2019. [Online]. Available: [https://www.hl7.org/implement/standards/product\\_brief.cfm?product\\_id=185](https://www.hl7.org/implement/standards/product_brief.cfm?product_id=185)
- [6] F. Khalique, S. A. Khan, and I. Nosheen, "A framework for public health monitoring, analytics and research," *IEEE Access*, vol. 7, pp. 101309–101326, 2019.
- [7] M. E. Beatty, H. S. Margolis, L. Coudeville, R. Hutubessy, J. Hombach, D. Dessis, B. Dervaux, O. Wichmann, M. I. Meltzer, J. N. Kuritsky, D. S. Shepard, and P. Beutels, "Health economics of dengue: A systematic literature review and expert Panel's assessment," *Amer. J. Tropical Med. Hygiene*, vol. 84, no. 3, pp. 473–488, Mar. 2011.
- [8] R. Alexander and M. Alexander, "An ICT-based real-time surveillance system for controlling dengue in Sri Lanka," May 2014, *arXiv:1405.4092*. [Online]. Available: <https://arxiv.org/abs/1405.4092>
- [9] E. Isere, A. Fatiregun, and I. Ajayi, "An overview of disease surveillance and notification system in nigeria and the roles of clinicians in disease outbreak prevention and control," *Nigerian Med. J., J. Nigeria Med. Assoc.*, vol. 56, pp. 161–168, May 2015.
- [10] N. Ibrahim and T. W. Quan, "The development of multi-platforms application for dengue-entomological surveillance system," in *Proc. 6th ICT Int. Student Project Conf. (ICT-ISPC)*, May 2017, pp. 1–4.
- [11] K. Minhas, M. Tabassam, R. Rasheed, A. Abbas, H. A. Khattak, and S. U. Khan, "A framework for dengue surveillance and data collection in pakistan," in *Proc. IEEE 43rd Annu. Comput. Softw. Appl. Conf. (COMPSAC)*, Jul. 2019, pp. 275–280.
- [12] G. E. Coelho, P. L. Leal, M. D. P. Cerroni, A. C. R. Simplicio, and J. B. Siqueira, "Sensitivity of the dengue surveillance system in brazil for detecting hospitalized cases," *PLOS Neglected Tropical Diseases*, vol. 10, no. 5, pp. 1–12, May 2016, doi: [10.1371/journal.pntd.0004705](https://doi.org/10.1371/journal.pntd.0004705).
- [13] R. Chen, P. Mongkolwat, and D. Channin, "Radiology data mining applications using imaging informatics," in *Data Mining in Medical and Biological Research*, E. G. Giannopoulou, Ed. London, U.K.: IntechOpen, Nov. 2008. [Online]. Available: [https://www.intechopen.com/books/data\\_mining\\_in\\_medical\\_and\\_biological\\_research/radiology\\_data\\_mining\\_applications\\_using\\_imaging\\_informatics](https://www.intechopen.com/books/data_mining_in_medical_and_biological_research/radiology_data_mining_applications_using_imaging_informatics), doi: [10.5772/6409](https://doi.org/10.5772/6409).
- [14] M. Yuksel and A. Dogac, "Interoperability of medical device information and the clinical applications: An HL7 RMIM based on the ISO/IEEE 11073 DIM," *IEEE Trans. Inf. Technol. Biomed.*, vol. 15, no. 4, pp. 557–566, Jul. 2011.
- [15] T. J. Eggebraaten, J. W. Tenner, and J. C. Dubbels, "A health-care data model based on the HL7 reference information model," *IBM Syst. J.*, vol. 46, no. 1, pp. 5–18, 2007.
- [16] R. Calamai and L. Giarre, "Enabling primary and specialist care interoperability through HL7 CDA release 2 and the chronic care model: An italian case study," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 42, no. 6, pp. 1364–1384, Nov. 2012.
- [17] I. S. for Disease Surveillance. (2012). *Electronic Syndromic Surveillance Using Hospital Inpatient and Ambulatory Clinical Care Electronic Health Record Data: Recommendations From the ISDS Meaningful Use Workgroup*. [Online]. Available: <https://www.surveillancerepository.org/electronic-syndromic-surveillance-using-hospital-inpatient-and-ambulatory-clinical-care-electronic>
- [18] I. S. for Disease Surveillance. (Jan. 2011). *Final Recommendation: Core Processes and Ehr Requirements for public health Syndromic Surveillance*. [Online]. Available: [https://healthsurveillance.siteym.com?page=sys\\_MU\\_2011\\_recom](https://healthsurveillance.siteym.com?page=sys_MU_2011_recom)
- [19] N. Angula and N. Dlodlo, "A standard approach to enabling the semantic interoperability of disease surveillance data in health information systems: A case of namibia," in *Proc. Int. Conf. Adv. Big Data, Comput. Data Commun. Syst. (icABCD)*, Aug. 2018, pp. 1–8.
- [20] R. A. Hoffman, H. Wu, J. Venugopalan, P. Braun, and M. D. Wang, "Intelligent mortality reporting with FHIR," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 5, pp. 1583–1588, Sep. 2018.
- [21] D. Rajeev, C. Staes, R. S. Evans, A. Price, M. Hill, S. Mottice, I. Risk, and R. Rolfs, "Evaluation of hl7 v2.5.1 electronic case reports transmitted from a healthcare enterprise to public health," in *Proc. AMIA Annu. Symp. AMIA Symp.*, Jan. 2011, pp. 1144–1152.
- [22] E.-W. Huang and D.-M. Liou, "Performance analysis of a medical record exchanges model," *IEEE Trans. Inf. Technol. Biomed.*, vol. 11, no. 2, pp. 153–160, Mar. 2007.
- [23] S. Runge-Ranzinger, P. J. McCall, A. Kroeger, and O. Horstick, "Dengue disease surveillance: An updated systematic literature review," *Tropical Med. Int. Health*, vol. 19, no. 9, pp. 1116–1160, Sep. 2014.
- [24] V. Raclot, R. Ramsey, S. Tong, and W. Hu, "Surveillance of dengue fever virus: A review of epidemiological models and early warning systems," *PLoS Neglected Tropical Diseases*, vol. 6, no. 5, May 2012, Art. no. e1648.
- [25] A. Chang, D. Fuller, O. Carrasquillo, and J. Beier, "Social justice, climate change, and dengue," *Health Hum. Rights*, vol. 16, pp. E93–E104, Dec. 2014.
- [26] M. Rabaa, C. Klungthong, I.-K. Yoon, E. C. Holmes, P. Chinnawitropisan, B. Thaisomboonsuk, A. L. Rothman, D. Tannitisupawong, J. Aldstadt, A. Nisalak, M. P. Mammen, S. Thammapalo, R. V. Gibbons, T. P. Endy, T. Fansiri, C. Pimgate, T. W. Scott, and R. Jarman, "Frequent in-migration and highly focal transmission of dengue viruses among children in Kamphaeng Phet, Thailand," *Int. J. Infectious Diseases*, vol. 16, pp. e90–e91, Jun. 2012.
- [27] N. S. Roslan, Z. A. Latif, and N. C. Dom, "Dengue cases distribution based on Land Surface Temperature and elevation," in *Proc. 7th IEEE Control Syst. Graduate Res. Colloq. (ICSGRC)*, Aug. 2016, pp. 87–91.
- [28] E. Massaro, D. Kondor, and C. Ratti, "Assessing the interplay between human mobility and mosquito borne diseases in urban environments," *Sci. Rep.*, vol. 9, no. 1, pp. 1–13, Dec. 2019.
- [29] A. L. Nevai and E. Soewono, "A model for the spatial transmission of dengue with daily movement between villages and a city," *Math. Med. Biol.*, vol. 31, no. 2, pp. 150–178, Jun. 2014.
- [30] A. Wesolowski, T. Qureshi, M. F. Boni, P. R. Sundsøy, M. A. Johansson, S. B. Rasheed, K. Engø-Monsen, and C. O. Buckee, "Impact of human mobility on the emergence of dengue epidemics in pakistan," *Proc. Nat. Acad. Sci. USA*, vol. 112, no. 38, pp. 11887–11892, Sep. 2015. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/26351662>
- [31] C. W. Morin, A. C. Comrie, and K. Ernst, "Climate and dengue transmission: Evidence and implications," *Environ. Health Perspect.*, vol. 121, nos. 11–12, pp. 1264–1272, Nov. 2013.
- [32] Y.-H. Lai, "The climatic factors affecting dengue fever outbreaks in southern taiwan: An application of symbolic data analysis," *Biomed. Eng. OnLine*, vol. 17, no. S2, p. 148, Nov. 2018.

- [33] Y. Choi, C. S. Tang, L. McIver, M. Hashizume, V. Chan, R. R. Abeyasinghe, S. Iddings, and R. Huy, "Effects of weather factors on dengue fever incidence and implications for interventions in cambodia," *BMC Public Health*, vol. 16, no. 1, pp. 1–7, Dec. 2016.
- [34] N. D. B. Ehelepola, K. Ariyaratne, W. M. N. P. Buddhadasa, S. Ratnayake, and M. Wickramasinghe, "A study of the correlation between dengue and weather in Kandy city, Sri Lanka (2003–2012) and lessons learned," *Infectious Diseases Poverty*, vol. 4, no. 1, p. 42, Dec. 2015, doi: 10.1186/s40249-015-0075-8.
- [35] M. Hashizume, A. M. Dewan, T. Sunahara, M. Z. Rahman, and T. Yamamoto, "Hydroclimatological variability and dengue transmission in Dhaka, Bangladesh: A time-series study," *BMC Infectious Diseases*, vol. 12, no. 1, p. 98, Dec. 2012.
- [36] N. D. B. Ehelepola, K. Ariyaratne, W. M. N. P. Buddhadasa, S. Ratnayake, and M. Wickramasinghe, "A study of the correlation between dengue and weather in kandy city, sri lanka (2003–2012) and lessons learned," *Infectious Diseases Poverty*, vol. 4, no. 1, p. 42, Dec. 2015.
- [37] C. M. Benedum, O. M. E. Seidahmed, E. A. B. Eltahir, and N. Markuzon, "Statistical modeling of the effect of rainfall flushing on dengue transmission in singapore," *PLOS Neglected Tropical Diseases*, vol. 12, no. 12, pp. 1–18, 12 2018, doi: 10.1371/journal.pntd.0006935.
- [38] H.-Y. Yuan, J. Liang, P.-S. Lin, K. Sucipto, M. M. Tsegaye, T.-H. Wen, S. Pfeiffer, and D. Pfeiffer, "The effects of seasonal climate variability on dengue annual incidence in hong kong: A modelling study," *Sci. Rep.*, vol. 10, no. 1, pp. 1–10, Dec. 2020.
- [39] S. G. Kakarla, C. Caminade, S. R. Mutheneni, A. P. Morse, S. M. Upadhyayula, M. R. Kadiri, and S. Kumaraswamy, "Lag effect of climatic variables on dengue burden in India," *Epidemiology Infection*, vol. 147, p. 147, Jan. 2019.
- [40] S. Polwiang, "The correlation of climate factors on dengue transmission in urban area: Bangkok and Singapore cases," *PeerJ Preprints*, vol. 4, 2016, Art. no. e2322v1, doi: 10.7287/peerj.preprints.2322v1.
- [41] S. Atique, S. S. Abdul, C.-Y. Hsu, and T.-W. Chuang, "Meteorological influences on dengue transmission in pakistan," *Asian Pacific J. Tropical Med.*, vol. 9, no. 10, pp. 954–961, Oct. 2016.
- [42] A. Y. Chang, M. E. Parrales, J. Jimenez, M. E. Sobieszczyk, S. M. Hammer, D. J. Copenhaver, and R. P. Kulkarni, "Combining Google Earth and GIS mapping technologies in a dengue surveillance system for developing countries," *Int. J. Health Geographics*, vol. 8, no. 1, p. 49, 2009.
- [43] J. E. Hernández-Avila, M.-H. Rodriguez, R. Santos-Luna, V. Sánchez-Castaneda, S. Román-Pérez, V. H. Ríos-Salgado, and J. A. Salas-Sarmiento, "Nation-wide, web-based, geographic information system for the integrated surveillance and control of dengue fever in Mexico," *PLoS ONE*, vol. 8, no. 8, pp. 1–9, Aug. 2013, doi: 10.1371/journal.pone.0070231.
- [44] S. Ahmad, M. Asif, R. Talib, M. Adeel, M. Yasir, and M. H. Chaudary, "Surveillance of intensity level and geographical spreading of dengue outbreak among males and females in Punjab, Pakistan: A case study of 2011," *J. Infection Public Health*, vol. 11, no. 4, pp. 472–485, Jul. 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1876034117302770>
- [45] C. Davi, A. Pastor, T. Oliveira, F. B. D. L. Neto, U. Braga-Neto, A. W. Bigham, M. Bamshad, E. T. A. Marques, and B. Acioli-Santos, "Severe dengue prognosis using human genome data and machine learning," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 10, pp. 2861–2868, Oct. 2019.
- [46] P. Siriysatien, S. Chadsuthi, K. Jampachaisri, and K. Kesorn, "Dengue epidemics prediction: A survey of the state-of-the-art based on data science processes," *IEEE Access*, vol. 6, pp. 53757–53795, 2018.
- [47] A. Appice, Y. R. Gel, I. Iliev, V. Lyubchich, and D. Malerba, "A multi-stage machine learning approach to predict dengue incidence: A case study in mexico," *IEEE Access*, vol. 8, pp. 52713–52725, 2020.
- [48] O. Mudele, F. M. Bayer, L. F. R. Zanandrez, A. E. Eiras, and P. Gamba, "Modeling the temporal population distribution of *Ae. aegypti* mosquito using big Earth observation data," *IEEE Access*, vol. 8, pp. 14182–14194, 2020.
- [49] P. S. Sasongko, H. A. Wibawa, F. Maulana, and N. Bahtiar, "Performance comparison of artificial neural network models for dengue fever disease detection," in *Proc. 1st Int. Conf. Informat. Comput. Sci. (ICICoS)*, Nov. 2017, pp. 183–188.
- [50] C. Liu, Q.-Y. Liu, H. Lin, B. Xin, and J. Nie, "Spatial analysis of dengue fever in Guangdong province, China, 2001–2006," *Asia-Pacific J. Public Health /Asia-Pacific Academic Consortium Public Health*, vol. 26, pp. 58–66, Jan. 2013.
- [51] M. Yajid, N. Che Dom, S. N. Camalxaman, and R. Nasir, "Spatio-temporal analysis for identification of dengue risk area in melaka tengah district," *Geocarto Int.*, vol. 35, no. 14, pp. 1570–1579, Feb. 2019.
- [52] M. Sanna and Y.-H. Hsieh, "Temporal patterns of dengue epidemics: The case of recent outbreaks in kaohsiung," *Asian Pacific J. Tropical Med.*, vol. 10, no. 3, pp. 292–298, Mar. 2017.
- [53] K. D. Sharma, R. S. Mahabir, K. M. Curtin, J. M. Sutherland, J. B. Agard, and D. D. Chadee, "Exploratory space-time analysis of dengue incidence in trinidad: A retrospective study using travel hubs as dispersal points, 1998–2004," *Parasites Vectors*, vol. 7, no. 1, p. 341, 2014.
- [54] D. Hamer and M. Lichtveld, "Spatial distribution of epidemiological cases of dengue fever in suriname, 2001–2012," *West Indian Med. J.*, vol. 64, pp. 344–350, May 2015.
- [55] S. Ahmad, M. Asif, M. Majid, M. Iqbal, and M. Adeel, "An effective study of geographical pattern and intensity of dengue outbreak among males in Punjab, Pakistan," *LIFE, Int. J. Health Life-Sci.*, vol. 3, no. 3, pp. 117–137, Jan. 2018. [Online]. Available: <https://grdspublishing.org/index.php/life/article/view/1089>
- [56] L. A. Baig, "Outbreak of dengue fever in pakistan and its prevention," *Ann. Jinnah Sindh Med. Univ.*, vol. 5, no. 2, p. 99, 2019.
- [57] N. Bostan, S. Javed, S. A. M. A. S. Eqani, F. Tahir, and H. Bokhari, "Dengue fever virus in Pakistan: Effects of seasonal pattern and temperature change on distribution of vector and virus," *Rev. Med. Virology*, vol. 27, no. 1, Jan. 2017, Art. no. e1899.
- [58] S. Ahmad, M. A. Aziz, A. Aftab, Z. Ullah, M. I. Ahmad, and A. Mustan, "Epidemiology of dengue in Pakistan, present prevalence and guidelines for future control," *Int. J. Mosq. Res.*, vol. 4, no. 6, pp. 25–32, 2017.
- [59] D. Murad, N. Badshah, and S. M. Ali, "Mathematical approach for the dengue fever transmission dynamics," in *Proc. Int. Conf. Appl. Eng. Math. (ICAEM)*, Sep. 2018, pp. 1–5.
- [60] D. Murad, N. Badshah, and S. M. Ali, "Mathematical modeling and simulation for the dengue fever epidemic," in *Proc. Int. Conf. Appl. Eng. Math. (ICAEM)*, Sep. 2018, pp. 1–3.
- [61] W. Shabbir and J. Pilz, "Bayesian spatio-temporal analysis for dengue fever in major cities of Pakistan," in *Proc. RSEP Conferences*, 2019, p. 1.
- [62] W. W. C. Information. (2020). *Climate and Average Weather in Pakistan*. [Online]. Available: <https://weather-and-climate.com/average-monthly-Rainfall-Temperature-Sunshine-in-Pakistan>
- [63] Z. Fatima, "Dengue infection in pakistan: Not an isolated problem," *Lancet Infectious Diseases*, vol. 19, no. 12, pp. 1287–1288, Dec. 2019.
- [64] G. of Pakistan. *National Institute of Health*. Accessed: Sep. 2019. [Online]. Available: <https://www.nih.org.pk/>
- [65] PITB. *Directorate General of Health Services, Government of the Punjab*. Accessed: Sep. 2019. [Online]. Available: <https://dghs.punjab.gov.pk/>
- [66] G. O. K. P. Health Department. *Directorate General Health Services*. Accessed: Sep. 2019. [Online]. Available: <http://www.dghskp.gov.pk/>
- [67] G. O. S. Health Department. *Directorate General Health Services Sindh*. Accessed: Sep. 2019. [Online]. Available: <https://www.sindhhealth.gov.pk/>
- [68] G. O. T. P. Punjab Information Technology Board. *Dengue Activity Tracking System dengue Activity Tracking System*. Accessed: Sep. 2019. [Online]. Available: <https://www.pitb.gov.pk/dats>
- [69] N. Healthcare. *Mirth Connect*. Accessed: Sep. 2019. [Online]. Available: <https://www.mirthcorp.com/community/issues/browse/MIRTH/?selectedTab=%com.atlassian.jira.projects-plugin:summary-panel>
- [70] U. H. Network. *H17 Application Programming Interface. University Health Network*. Accessed: Sep. 2019. [Online]. Available: <https://hapifhir.github.io/hapi-h17v2/>
- [71] O. C. Ibe, "2—Basic concepts in stochastic processes," in *Markov Processes for Stochastic Modeling*, 2nd ed. O. C. Ibe, Ed. Oxford, U.K.: Elsevier, 2013, pp. 29–48. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/B9780124077959000025>
- [72] F. Khalique, S. A. Khan, W. H. Butt, and I. Matloob, "An integrated approach for spatio-temporal cholera disease hotspot relation mining for public health management in punjab, pakistan," *Int. J. Environ. Res. Public Health*, vol. 17, no. 11, p. 3763, May 2020, doi: 10.3390/ijerph17113763.
- [73] A. Ali, Z. Fatima, B. Wahid, S. Rafique, and M. Idrees, "Cosmopolitan lineage of dengue virus serotype 2 is circulating in pakistan: A study from 2017 dengue viral outbreak," *J. Med. Virology*, vol. 91, no. 11, pp. 1909–1917, Nov. 2019.



**FATIMA KHALIQUE** received the M.S. degree in computer science from Uppsala University, Sweden, in 2007. She worked in industry and academia. She was a Lecturer with the National University of Sciences and Technology (NUST) and the National University of Modern Languages (NUML), Islamabad, Pakistan. She was also a Software Developer with Zhongxing Telecom Engineering (ZTE), Islamabad. She is currently a Ph.D. Scholar with NUST. She is also an Oracle Certified Professional. Her research interests include data mining, health informatics, artificial intelligence, data analytics, and machine learning algorithms.



**RUBINA SHAHEEN** (Member, IEEE) received the B.S. degree from the National University of Sciences and Technology (NUST), Islamabad, Pakistan and the M.S. degree from the Centre for Advanced Studies in Engineering (CASE). She was a Software Developer with Digitania, a leading multinational software house. She has a diverse experience in industry and academia. She is currently a Lecturer with the Balochistan University of Information Technology, Engineering and Management Sciences (BUIITEMS). Her research interests include software requirement engineering, system design and testing, data mining, health analytics, and various areas in artificial intelligence.



**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 of computer and software engineering (C and SE) with the College of EME, National University of Sciences and Technology. He is an inventor of five awarded U.S. patents. He has over 260 international publications. His book on *Digital Design* (John Wiley & Sons). He is being followed in national and international universities. He has more than 22 years of industrial experience in companies, USA, and Pakistan. He has founded the Center for Advanced Studies in Engineering (CASE) and the Center for Advanced Research in Engineering (CARE). CASE is a Premier Engineering Institution that runs one of the largest post graduate engineering programs in the country. He has already graduated 50 Ph.D., and more than 1800 M.S. degree in different disciplines in engineering, whereas CARE, under the 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, like Australia, Singapore, Hong Kong, Malaysia, and so on. Dr Khan was a member of the Board of Governance of many entities with the Ministry of IT and Commerce. He was also a member of the National Computing Council and the National Curriculum Review Committee. He was the Chairman of Pakistan Association of Software Houses (P@SHA). He received the Tamgh-e-Imtiaz (Civil), the Highest National Civil Award, Pakistan, the National Education Award in 2001, and the NCR National Excellence Award in Engineering Education.

...