<u>Dengue-View</u>: Dengue Epidemic Surveillance Modeling, Visualization, and Response Management System

No.	***Elapsed time (in months)***	***Milestone***	***Deliverables***	***Status***
1.	Apr 16 - July 15, 2013	***Task 1:*** *Integrated and sharable ID databank design and development (both logical and physical)*	**Notable Metrics:** • Physical and logical *design of the distributed ID database (tables etc.)*, which will be later used for the collection, preprocessing, integration, and query of streaming data arriving from different hospital sources, together with the weather, vegetation and human and vector population data. The database meant to enable disease outbreak detection and forecasting across space and time. • A progress report will be submitted detailing all physical and logical design issues and the final distributed ID Database design.	**Submitted**

<u>First Quarter Progress</u>

The project commenced on April 15, 2013 and first milestones were achieved on July 15, 2013. The details of which are as follows:

Task 1: Integrated & Shareable ID Databank: Design and Development

A comprehensive design of the ID Databank has been finalized and includes the following;

- **Databank schema diagram & ED Work Flow Diagram**

The database schema of the three primary partner hospitals has been finalized after detailed analysis of hospital's standard operating procedure. The workflow design is completed after careful study of patient's treatment procedure at emergency departments (ED). The diagram depicts entities and processes of the ED system.

- **Climate Database**

To record various climatic conditions and population (human and vectors) parameters the database schema of climate data is finalized.

- **System Architecture**

A java based enterprise system architecture design is finalized which covers the multi-tier/multi layer client server distributed system architecture illustrating the integration of other modules such as Smart Phone platform for the acquisition of the data from rural health centers.

The implementation status of the ID Databank is summarized as follows:

- The ID databank application consists of Master Server currently deployed at Center of Visual Analytics and Research (CVAR), KICS, UET, Lahore. This server hosts the Master application

	server and the Master Database server and is responsible for collection of multi-source disease surveillance streaming data (including primary, secondary, and tertiary hospital ED data; human population density data; vector concentration data; and temperature and humidity data). • The desktop clients (at CVAR) attached to master server are responsible for the management of the server side relational database and support acquisition, transformation, integration, processing and the visualization of dengue related data. • A group of servers (ED-Servers) is developed and being deployed at different partnering hospitals. The corresponding desktop clients being connected at these servers are used to input information of ID surveillance data with attributes such as patient ID, patient visit time, patient location (residence/spatial coordinates), syndromes symptoms. • A Smart phone based applications is developed for both Windows and Android Mobile platform to collect ED data from remote rural areas. The application consists of suitable GUI forms for the data input and submission to remote servers.			
2.	July 16 – Oct 15, 2013	***Task 2:*** *Real-time, multi-source, space-and-time-stamped streaming ID data capture, integration, and visualization modules*	**Notable Metrics:** • Dengue-View modules for the purpose of capturing, preprocessing, transformation, and integration of the streaming ID and weather data etc. mentioned just above in Task 1, at a centralized location at UET C-VAR. The space-and-time-stamped ID data will be arriving in real-time from multiple (tertiary, secondary and primary) hospital locations from all across Pakistan. • A set of data visualization modules used for displaying the space-time stamped ID related data; weather, vegetation and human and vector population data; the analyses results data; and disease spread patterns etc. in the form of appropriate graphs and charts. The displays will also support and enable user-interactive visual analyses. • A progress report will be submitted detailing visualization module of Dengue-View covering all the space-time stamped ID data streaming, capturing, preprocessing, transformation, and integration details. The report will also include the	**Submitted**

			algorithms used for these purposes. • Demo of Dengue-View modules developed so far.	

The second milestones was focused on development of following modules:

Module 1: Data Acquisition Module (Capturing and Cleaning of Infections Disease Data)

Main focus of this module is to develop a system that captures, cleans, prepares and preprocess real time ID (infectious disease) data for the data mining module. Raw data is acquired from different sources (Primary/Secondary/Tertiary Hospitals) and stored in ID data-bank. This information is then parsed, filtered and processed to identify and resolve typographical errors, missing details and invalid entries.

Implementation Details

Servers deployed at Emergency Department (ED) are capable of collecting live streaming ID data. The corresponding desktop clients, connected to these servers through local network, are used to input ID surveillance data with attributes such as patient's id, visits time, location (resident/spatial coordinates), syndromes, symptoms etc. The ED servers communicate with Master server at CVAR for streaming live data.

Windows and Android Phones are being used for data collection from remote rural areas. Both applications consists of suitable GUI for the input data and uploading to remote server.

Module 2: Data Mining Module (Data Integration and Correlation)

This module consists of Integration and correlation of Spatio-temporal data i.e syndromic surveillance data, vectorial data, weather data etc. It aims to provide insights of the dengue disease through different data mining techniques.

a. Integration

In the integration phase, the addresses of patients are classified using Spatial DB called PakGIS. It is developed at CVAR for storing the geographical information of different areas of each union council of a city. This information includes the longitude/latitude coordinates and geometry of the area. With the help of this information, the geographic coordinates of patients are identified and integrated for further use.

b. Correlation

In the correlation phase, all the input variables of the weather data, vector data and patient data are correlated and their impact in the spread of the dengue is studied.

Implementation Details:

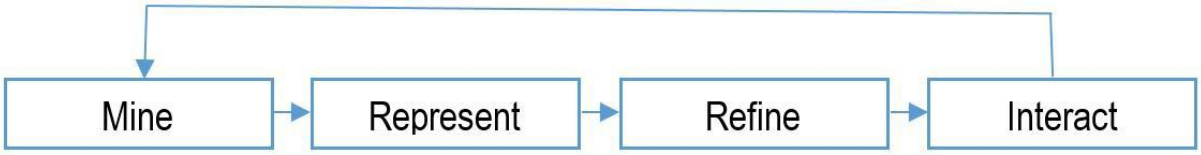
To keep track of geographical locations, spatial database is designed using PostGreSQL (open source object-relational database management system). Post GIS plug ins are added to store and query data representing objects defined in geometric space.

ID data is correlated with weather data using statistical tools (Matlab and R) to find climatic factors affecting diseases spread.

i. Module 3: Visualization Module

The purpose of this module is to develop analytics based software for real-time data visualization. Data is

further processed for visualization using different steps which are shown in following pipeline:



Following two tasks are designed in this modules:

i. Choropleth

Choropleth provides an easy way to visualize how patient density varies across neighboring geographic areas. It uses color schemes which represent different geographical locations with varying data. Specific visualization data can be obtained using filters based on age, sex, time and locations.

ii. Heat maps

Heal maps are generated on the basis of 'No. of dengue cases reported' for a particular geographical location. Overlays related to weather, vegetation, human and vector population, the analyses results data and disease spread patterns can also be added. The display supports and enables user-interactive visual analyses. Alerts are generated in Spatio-temporal realm based on increasing rate of change in patient count density.

Implementation details:

The current implementation of the visualization module is an extension and integration of:

 Open source visualization API called D3 (Data Driven Documents); a JavaScript library widely used for processing geojson files (file format for encoding collections of simple geographical features along with their non-spatial attributes).

 Google API is embedded for generating heat maps based on counts for a particular geographical location for alert generation.

Visualization Control Room (VCR) follows a four step procedure. First is the database record retrieval, to retrieve patient records from PakGIS database. Second, is to apply filters based on age, sex and location as per user requirement. Third is data visualization which visualizes our filtered data and reveals hidden trends. The final step provides interactivity, in which the user is able to interact with the visualizing data such as panning, zooming etc., to obtain different sets of results.

\Syndrome specific information is generated after classifying complaints data to a particular syndromic category. For data visualization, anonymization is maintained by synthesizing center points for any geographical location. User may select the syndromic hotspot from generated heat maps and call/analyze historical time series data in the area to look for unusual trends or potential outbreaks.

3.	Oct 16 – Jan 15, 2014	***Task 3:*** *Real-time integrated syndromic surveillance sub-system for DE outbreak detection and reporting*	• Real-time classification of streaming ED data into syndromic categories; the dengue related patient data will be separated at this point for further analyses. • Syndromic surveillance based early disease	Submitted

			outbreak detection and alarm reporting modules. • GIS marriage/incorporation of the modules, with the help of Google Maps. • A progress report will be submitted detailing syndromic surveillance module algorithmic and functionality details. • Demo of Dengue-View syndromic surveillance module based disease outbreak detection along with GIS based disease mapping using Google maps. • A conference paper will be ready for submission.	

Third milestone started on 1st November 2013 and was focused on developing real time syndromic subsystem for dengue epidemic (DE) outbreak detection and reporting.

Module 1: Modeling Emergency Department data

Due to unavailability of Emergency department (ED) data PI has simulated patient's emergency department visit patterns for infectious diseases. For each town input data was modeled by histogram generation followed by estimating probability density function for each disease. Similarly the project team has used user-defined probability density function for generating patient's age and gender. In later stages the user will be provided with provision to model patient's age and gender distribution through an interactive tool. Historical ED department stats data was collected to model patient's arrivals in emergency departments to model seasonal trends. This will be further extended to incorporate days of the week effects, long term temporal trends using seasonal trend decomposition (STL) followed by spatial distribution using kernel density estimation.

Module 2: Syndromic Classification

This module relies on chief complaints data collection from partnering hospitals, i.e. Ganga Ram and Mayo hospitals; which is being processed into one of nine syndromic categories (respiratory, gastro-intestinal, hemorrhagic, rash, fever, neurological, botulinic, shock/coma and others) using neural networks. Syndromic classification will be perused for early detection of any of the above mentioned general categories of infectious diseases in Pakistan–dengue the specific that we will target on. Syndromic classifier is easily amenable to incorporate other syndromes and diseases so can be used in any geographical location in Pakistan with minor changes. For current system we particularly focus on separating the illness cases that are representative of any of the four serotypes of dengue with added reference to geographical and temporal origin of the disease event associated with patient in question.

This module is well integrated with input data module, to generate syndromic categories based on ED data and later on outbreak detection followed by visualization, to provide Visual Analytics environment. Visual analysis of syndromic clusters is also helpful in identifying risk groups in correlation with pollutant hazards with reference to geographical locations; thus elaborating the dominant factors causing the epidemic spread.

Module 3: Out-Break Detection

Main objective of this system is to develop methodology for out-break detection based on spatio-temporal syndromic data. Different standard event detection algorithms by center for disease control (CDC) are being compared; all these algorithms are based on cumulative summation (CUSUM) that include C1-MILD (C1), C2-MEDIUM (C2), and C3-ultra (C3) each requiring less than three years of historical data. Based on sensitivity, specificity and time to detection parameters for each algorithm previously stated we will finalize aberration detection method.

Research team has developed disease injection module so that the users can create known outbreaks in order to test the detection tools and algorithms. For this we have used approach of parameterizing clusters using simple mathematical models. Users can create clusters for their supplied baseline data by selecting the cluster's center relative to hospital location, temporal duration and magnitude of severity. Future work will provide user with ability to set the probability that the disease will affect a certain age and gender more, by linked interactive probability distribution tool to control these parameters.

Module 4: Visual Analytics Environment

Classification of emergency department data was followed by multivariate time series analysis to provide users with ability to explore data for common patterns that help in outbreak detection. Multiple plots including line charts scatter plots, area charts are developed to provide visual analytics environment that support users with contextual information about significance of an event. Moreover data is being clustered based on syndromic categories; for time series data clustering each time step is being treated as a separate attribute, good choice of colors helps understanding the cluster results leading to comparison of cluster results to allow analyst to judge goodness of grouping based on interpretability of spatial patterns.

Visualization toolkit will maintain spatiotemporal history to keep track of movement of syndromic hotspots across space and time. User is provided with ability to switch between multiple views i.e. geographical trend analysis based on temporal data. Syndromic hotspots of infectious disease (ID) can be duly selected and immediately analyzed in corresponding linked views. Users may select Syndromic hotspot from generated heat-maps and call/analyze for historical time series data in the area to look for unusual trends or potential outbreaks. Such multi-linked views allow users to quickly form and test hypotheses, thereby reducing the time needed to reject false positives and confirm true outbreaks.

The current Visual Analytics environment provides user with following functionality

- Time series charts for daily weather data i.e. temperature, rainfall, humidity to find correlation between weather patterns and emerging infectious diseases.
- Time series charts for visualization of patient records in partnering hospitals to keep track of hospital work load in terms of number of patients being entertained.
- Spatial charts for visualization of patients' location and regions under threat of dengue outbreak.
- Data clusters based on geographical location of syndromic categories for analysis of risk groups based on demographics data i.e. age, gender, geographical locations, mobility patterns and occupation.

Implementation details:

The implementation of the visualization module is an extension and integration of:

• Open source visualization API called Highcharts; a JavaScript library widely used for processing geojson files (file format for encoding collections of simple geographical features along with their non-spatial attributes).

	• Google API and Leaflet is embedded for generating heat maps based on counts for a particular geographical location for alert generation.			
4.	Jan 16 – Apr 15, 2014	***Task 4:*** *Statistical and geo-spatiotemporal DE spread modeling, analysis, and visualization*	**Notable Metrics:** • Development of mathematical models for 4-serotype geo-spatial time-series modeling of the spread of dengue viruses using STL approaches; deploying these modules as part of the Dengue-View. • A progress report will be submitted detailing 4-serotype geo-spatial time-series modeling module of the spread of dengue viruses using STL approaches. The report will also include algorithmic and functionality details. • Demo of Dengue-View for 4-serotype geo-spatial time-series modeling along with GIS based time-series mapping using Google maps. • A conference paper will be ready for submission.	**Submitted**

Fourth milestone started on 1st February 2014 and was focused on the development of statistical/stochastic techniques for 4-serotype geo-spatiotemporal DE spread modeling, analysis, and visualization. Timeline of major modules included in this milestone is given as under

No	Tasks	Feb				Mar				Apr			
		W1	W2	W3	W4	W1	W2	W3	W4	W1	W2	W3	W4
1	**Historical Trend Analysis**	100%											
2	**Modeling Mosquito Reproduction Behavior**	100 %											
3	**Modeling Mosquito Transmission Behavior**					80%							
4	**Spatial Distribution of Dengue Transmission**									20%			

(The green area shows the percentage of task completed. White area represents inactivity of the respective tasks)

5.	Apr 16 – July 15, 2014	***Task 5:*** *Simulation-based geo-spatiotemporal prediction/forecasting of*	**Notable Metrics:** • Development of mathematical simulations for discovery and detection of DE outbreak patterns,	**Submitted**

		dengue hotspots and outbreaks	together with forecasting of DE hotspots across space and time. • Application of sliding window techniques and Maximum Likelihood Estimation. • A progress report will be submitted detailing geo-spatiotemporal prediction/forecasting module. The report will also include algorithmic and functionality details. • Demo of Dengue-View simulation-based geo-spatiotemporal prediction/forecasting module along with GIS based dengue hotspots and outbreaks mapping using Google maps. • The first workshop will be arranged at UET Lahore for training and technology transfer purposes, based on the progress so far.	
6.	July 16 – Oct 15, 2014	***Task 6:*** *Prototyping the visual analytics modules of Dengue-View system for GIS-based DE spread modeling, outbreak prediction, visualization, and alerts generation*	**Notable Metrics:** • Prototyping, testing, and deploying the Dengue-View modules for GIS based statistical DE spread modeling, as well as spatiotemporal disease outbreak and disease hotspot prediction/forecasting; • Prototyping, testing, and deploying the epidemic scenario simulation and hypotheses testing capability of the Dengue View. • A progress report will be submitted detailing visual analytics module. The report will also include algorithmic and functionality details. • Demo of Prototyping, testing, and deploying of modules for GIS-based mapping and visualization of the pertinent	**Submitted**

			DE information along with the disease spread scenarios/processes, in the form of interlinked user interfaces/displays. • A conference paper will be ready for submission.	

7. The 6th module of the project was focused on development of a Prototype, which includes GIS based Dengue Epidemic spread modeling, outbreak prediction, visualization and alerts generation modules of the system. The software architecture in this task consists of following modules:

- Modeling Mosquito Reproductive Behavior
- Temporal Outbreak for Surveillance Data.
- Visualization Module.
- Spatiotemporal Outbreak

The underlying technology used for the implementation of these modules is Spring Framework – a famous Java platform for developing multitier web applications.

No	Tasks	Aug				Sep				Oct			
		W1	W2	W3	W4	W1	W2	W3	W4	W1	W2	W3	W4
1	**Modeling Mosquito Reproduction Behavior**	**100%**											
2	**Temporal Outbreak for Surveillance data**	**100 %**											
3	**Visualization Module**					**100%**							
4	**Spatiotemporal Outbreak**									**100%**			

8.	Oct 16 – Jan 15, 2015	***Task 7:*** *Development of (an alternative) social networking based DE spread model and its integration within Dengue-View*	• Development and integration in Dengue-View of a far more sophisticated agent-based, human-vector social networking model for 4-serotyple DE spread and forecast modeling (A Ph.D. student will be separately working on this problem). • Applying MIDAS styled agent-based modeling techniques to the spread and forecast of dengue. • A progress report will be submitted detailing social networking based DE spread modeling module. The report will also include algorithmic and functionality details. • Demo of Dengue-View this social networking based DE	**Not Due**

			spread modeling module. • A Journal paper will be ready for submission.	
9.	Jan 16 – Apr 15, 2014	***Task 8:*** *Integration of DE threat response planning and management capability module into the Dengue-View environment*	**Notable Metrics:** • Threat response planning and management capability in Dengue View. • Dengue View can now simulate disease spread and disease prevention/control scenarios, for studying the effects of different pre-emptive counter-measures. • A progress report will be submitted detailing response planning and management module. The report will also include algorithmic and functionality details. • Demo of Dengue-View this threat response planning and management modules. • A conference paper will be ready for submission.	**Not Due**
10.	Apr 16 – July 15, 2015	***Task 9:*** *Validation of Dengue-View system for disease event prediction, planning, counter-measure actions, and response of the counter-measures*	**Notable Metrics:** • This is primarily a validation and verification task. • Real DE scenarios are planned and this system is verified through feedback from domain experts in the epidemic control field. The experts would advise/guide in planning various counter-measures and then evaluate the effects of the counter-measures, refining them where required. • A progress report will be submitted on validation and verification of Dengue-View and results of test scenarios will be included in the report. • A journal paper will be ready for submission.	**Not Due**
11.	30	***Task 10:*** *Integration of additional hospitals, all*	**Notable Metrics:** • Ensuring that the Dengue-	

		across Pakistan, into the Dengue-View system	View architecture is scalable and flexible enough to accommodate/integrate additional hospitals into the system, as and when required. • A progress report will be submitted on integration of other hospitals.	
12.	30	***Task 11:*** *Training and technology transition to partners all across Pakistan*	**Notable Metrics:** • Training materials about dengue epidemic monitoring; to be disseminated through on-line services to the stakeholders. • Transition of Dengue-View system to the end users. • The second workshop will be arranged for training and technology transfer purposes, based on the progress made so far.	**Not Due**