

# Geo-Spatial Disease Clustering for Public Health Decision Making

Atta-ur-Rahman<sup>1</sup>, Munir Ahmed<sup>2</sup>, Gohar Zaman<sup>3</sup>, Tahir Iqbal<sup>4</sup>, Muhammed Aftab Alam Khan<sup>5</sup>, Mehwash Farooqui<sup>5</sup>, Mohammed Imran Basheer Ahmed<sup>5</sup>, Mohammed Salih Ahmed<sup>5</sup>, Majd Nabeel<sup>1</sup> and Abdullah Omar<sup>1</sup>

E-mail: aaurrahman@iau.edu.sa

<sup>1</sup>Department of Compute Science (CS), College of Computer Science and Information Technology (CCSIT) Imam Abdulrahman Bin Faisal University (IAU), P.O. Box 1982, Dammam, 31441, Saudi Arabia

<sup>2</sup>Barani Institute of Information Technology (BIIT), PMAS Arid Agriculture University, Rawalpindi, 46000, Pakistan

<sup>3</sup>Faculty of Computer Science and Information Technology

Universiti Tun Hussein Onn Malaysia (UTHM), Batu Pahat, 86400, Malaysia

<sup>4</sup>Department of Business Administration, College of Business Administration (CBA)

Imam Abdulrahman Bin Faisal University, P.O. Box 1982, Dammam 31441, Saudi Arabia

<sup>5</sup>Department of Computer Engineering (CE), College of Computer Science and Information Technology (CCSIT)

Imam Abdulrahman Bin Faisal University (IAU), P.O. Box 1982, Dammam, 31441, Saudi Arabia

**Keywords:** geo-spatial mapping, public healthcare, decision making, clustering

**Received:** November 15, 2021

*An explosion of interest has been observed in disease mapping with the developments in advanced spatial statistics, data visualization and geographic information system (GIS) technologies. This technique is known as “Geo-Spatial Disease Clustering,” mainly used for visualization and future disease expansion prediction. Its importance has been overwhelmingly observed since the COVID-19 pandemic outbreak. Government, Medical Institutes, and other medical practices gather large amounts of data from surveys and other sources. This data is in the form of notes, databases, spread sheets and text data files. Mostly this information is in the form of feedback from different groups like age group, gender, provider (doctors), region, etc. Incorporating such heterogeneous nature of data is quite challenging task. In this regard, variety of techniques and algorithms have been proposed in the literature, but their effectiveness varies due to data types, volume, format and structure of data and disease of interest. Mostly, the techniques are confined to a specific data type. To overcome this issue, in this research, a data visualization technique combined with data warehousing and GIS for disease mapping is proposed. This includes data cleansing, data fusion, data dimensioning, analysis, visualization, and prediction. Motivation behind this research is to create awareness about the disease for the guidance of patients, healthcare providers and government bodies. By this, we can extract information that describes the association of disease with respect to age, gender, and location. Moreover, the temporal analysis helps earlier prediction and identification of disease, to be care of and necessary avoiding arrangements can be taken.*

*Povzetek: Analizirana je vloga vizualizacije podatkov in rešitev za naloge prostorskega gručenja bolezni, tj. ugotavljanje in napovedovanje širjenja bolezni.*

## 1 Introduction

Data mining and visualization techniques have been extensively used by many organizations especially in healthcare. It provides very interesting pattern and useful health information, based on which many critical decisions could be made. While visualizing these patterns on the map will be very helpful for the healthcare researchers and stakeholders to observe and predict the disease spread. The importance and application of such studies have been tremendously observed since the COVID-19 pandemic outbreak. Several studies are evident since late 2019 [1-2] in this regard. This has been greatly helping the healthcare professionals and government bodies to observe and control the disease spread. Almost every infected country in the world has similar type of systems and mobile phone applications for this purpose. Map base analysis are more interactive and usable for the end user. The analysis of the healthcare data

in different dimensions like time, gender, age group and location, make it more useful for sake of decision making. Moreover, different stakeholders may be interested in looking at a different perspective of same data. For example, diabetes could be analyzed and visualize in time, gender, or state dimension, to see how disease varied on monthly, quarterly, or yearly basis [3-5]. This research is intended to provide a diverse healthcare data analysis and visualization through maps using data warehouse and Tableau (www.tableau.com) to pull out the unique patterns and interesting information which could lead to the efficient healthcare analysis, successful solutions, and predictions for the future of healthcare in USA.

Rest of the paper is organized as follows. Section 2 contains the review of literature with a focus on existing and related work in the near past. Section 3 thoroughly covers the proposed approach. Section 4 is dedicated to

results and discussions, while section 5 concludes the paper.

## 2 Review of literature

The current era is about application of information and communication technologies (ICT) in various fields. Healthcare is one the most studied and investigated field in this regard [6-8]. This section reviews the related work regarding data visualization of diseases related healthcare data. Data visualization is an important factor in any field and analyzing this data is also very important in decision making. After the brief description of the work carried out by different researchers in the area, their contributions and conclusions have been presented.

### 2.1 Spatial and Temporal Dynamics

*Aedes aegypti*, the yellow fever mosquito along with their other types have been most widely studied in terms of its growth rate. Dengue has been imitated a very fast spreading mosquito-borne viral disease in the world and is now the most prevalent human arbovirus infection with roughly one half of world's population living in various countries especially in tropic zones. Reducing mosquito vector populations and human-vector interactions are currently existing dengue prevention approaches [3-5].

In [3], authors investigated the geo-spatial relationship for dengue virus disease and found interesting patterns related to growth in near future and its possible causes. The study showed that petrol pumps, workshops, rice paddy, marsh/swamp and deciduous forests played a significant role in dengue vector growth. In [4], authors presented a very useful mapping of *Aedes* mosquitoes breeding habitats in urban and peri-urban areas using a fuzzy analytical hierarchical process. The parameters of the process were climatic and physical. The dataset was comprised of satellite imagery and geospatial data collected on the ground. In [5], authors presented a near real-time geospatial data analysis and visualization for dengue fever and its possible growth and infected areas in near future. So that the authorities can take avoidance and prevention measures well before time. The results were classified based on gender, age-group, population density and region. It was concluded that age group (5-24) years was more vulnerable, almost 67%; students, workers and laborers are more vulnerable in terms of profession, almost 88%; in general, males were more vulnerable than female, nearly 10% and based on proximity, public places, market, and religious places were more prone to be infected. In [9], a study is made for the Brazilian oceanic island (urban areas only), Fernando de Noronha (located at S3 °45 to S3 °57 and W32 °19 to W32 °41 and 545 km at northeast of Recife City, the capital of Brazil (Pernambuco)) which is shown in Figure 1. The monitoring system (SMCP-Aedes) describes island areas of urban population, the temporal and spatial distribution of the dengue vector that is based upon on a 103-trap network for *Aedes* egg samples, by using the analysis tools of spatial statistics and GIS. The research was implemented in a combined effort between the staff, local

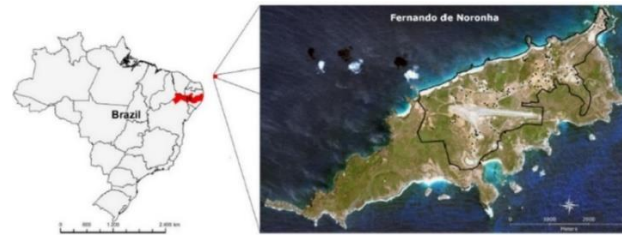


Figure 1: The study site Fernando de Noronha [9].

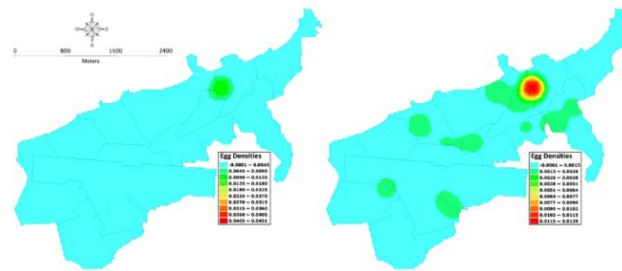


Figure 2: Spatial distribution of egg density [9].

health managers and the scientific team. This report of island characteristics of the infestation by dengue vector provides basic information for the analysis of relationships between dengue cases, the vector of spatial distribution and for the development of integrated the control strategies of vector [9].

KDE maps in Figure 2, present smooth egg density for the same period, as generated for a commonly used legend (from left) marking to compare maps from different months and the changed legend KDE map (from right) for the (similar) month, highlighting portions for high egg population.

### 2.2 Geographical Information System

According to [10], dengue and chikungunya zika are the diseases mostly were considered as there are certain areas where these diseases exist like Latin America and other tropical areas, but due to migration and tourism these diseases became also endemic even where it did not exist. National and international levels institutes are working collectively on the academic study of epidemiological of infectious diseases information in local and national scales. Geographical information systems (GIS) are being develop for epidemiological maps which has been used for dengue, however not in other developing arboviral diseases also not in the Central America. During the study period of (2015), in Honduras reported cases were 19,289 and 85,386 for dengue and chikungunya respectively and with median of ranging with 291 to 1789, 726 cases reported per week for dengue and chikungunya has 1460 with ranging 387 to 4175. Dengue cases were reached at highest ratio that us 25th week during epidemiological while 27th was at whilst for chikungunya. Projected cases ratio for dengue and chikungunya in Honduras for year 2015 by department(s). These maps were generated with the help of a Kosmo GIS. Furthermore, national GIS-based had generated the maps for departments wise and municipalities wise, for distribution of the chikungunya and dengue. The Microsoft Access® Software was the

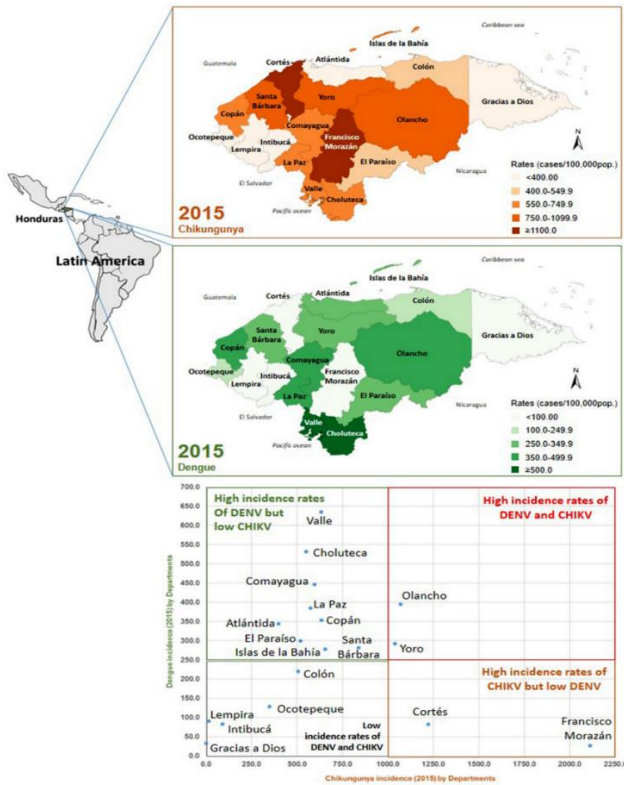


Figure 3: GIS-based map (Geographic Distribution) [10].

platform to design and develop the spatial databases which are used for the improvement of rates in frequency by municipalities, disease, and departments as well for the software of GIS. The “Client GIS” is an open-source software was used, named “Kosmo Desktop 3.0RC1®”. To generate digital maps of yearly cases rates by departments, municipalities and three shapefiles of departments which were joined to a data table (database) through spatial linked operation used. The map is shown in Figure 3.

### 2.3 Global maps

According to [11], nearly “one third” population of the human are unprotected against the danger of dengue (Figure 4). Every year almost fifty to hundred million occurrences are recorded with fever of dengue and 500000 occurrences of severe dengue like dengue fever hemorrhagic requiring hospitalization and from 20,000 to 25,000 cases of deaths, many of them are children. A dengue global risk map has been generated, based upon one database having global amount for disease, Dengue’s data predicted has vector species (two main) *Aedes aegypti* and *Aedes albopictus* and human population density as well. Three different sets of dengue fever have been used to make almost 100 bootstrap models by making sub-sampling of severe dengue like dengue hemorrhagic fever, DHF and all-dengue, output results were used to make a single global risk map for each type of dengue. It includes predictor variables like Land Surface Temperature (weather), thermal data layers having both day and night times, population of human density and a variation of rainfall. The map for dengue global risk presents risk in South America Asia, India,

Central America, and parts of coastal South America, but in generally some areas of Africa. High human population density is the key significance for all dengue risk maps made here. The risk of dengue from Europe at present is measured to be little, but sufficiently ambiguous to warrant monitoring in those zones of greatest predicted suitability of environmental, particularly in counties of northern Greece, portions of Austria, Croatia, Slovenia, Bosnia and Herzegovina, southeastern France, Serbia, Albania, Montenegro, Germany, Italy, and Switzerland, and in smaller areas elsewhere. This map is generated by using the data from all dengue databases and included the dengue’s vector species of modelled distributions of (two vector), *Aedes aegypti* and *Ae. albopictus*, as predictor data layers (which are not selecting in all models). The map represents the probability of suitability, like the relationship from each pixel and any of the other clusters included in the model. The gray parts are different from any of the occurrence or absence clusters that data will has no predictions are made for them.

Approaches identical to above mentioned, have been made for other regions of the world and for various diseases. For example, in [12-13], authors proposed map-based data analysis and visualization technique for Alzheimer’s disease. In [14], authors proposed a neural correlation between the psychedelic state and psilocybin determined by functional Magnetic Resonance Imaging or functional MRI (fMRI) studies. In [15], authors studied the geospatial relationship of chikungunya epidemic in South India. Similarly, in [16], the aim was to disease estimation and visualization for small area using R language.

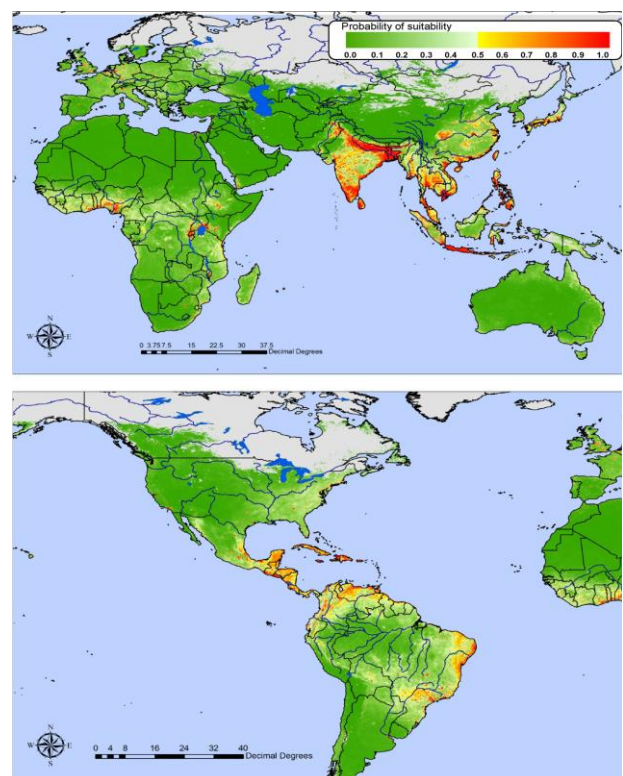


Figure 4: Global Risk map (for dengue) [9].

### 2.4 COVID-19 disease clustering

Thousands of studies have been conducted around the globe since the covid-19 pandemic has emerged [17]. These studies can be categorized in various types. For example, disease prediction based on the symptoms like cough sound, chest x-ray, and other historic patient data using variety of techniques [18-21]. Similarly, various apps and systems were developed to observe and forecast the disease spread on temporal and spatial bases [22].

## 3 Proposed approach

This section contains the whole description about the proposed approach used in this research. Data is taken from a Healthcare IT Company. Data preparation and data pre-processing (warehouse) are the strategies used to generate results and data representations. Data warehouse has been used to pull out the information from the system. Because it is very complex so for understanding, ER diagram is given in Figure 5. We pick the data of year 2016 for warehouse in this experiment. This system will help to create awareness about the disease for the guidance for patient and health care stakeholders. By this, we can extract data that describes source of disease in terms of age, gender, and the state as well for an arbitrary disease.

### 3.1 Data warehouse

An enormous impact has been created by data warehousing technology in the business world, with its help data turned into information for big competitive benefits. Data warehousing in a medical field, have traditionally been administrative in nature, focusing on patient management and billing, organizational aspects of hospitals that were improved by using the data warehouse techniques are not much different from the contemporary enterprises.

Technology though changed rapidly, and now more difficult areas of medical data management could be handled. The information technology maintained fetching

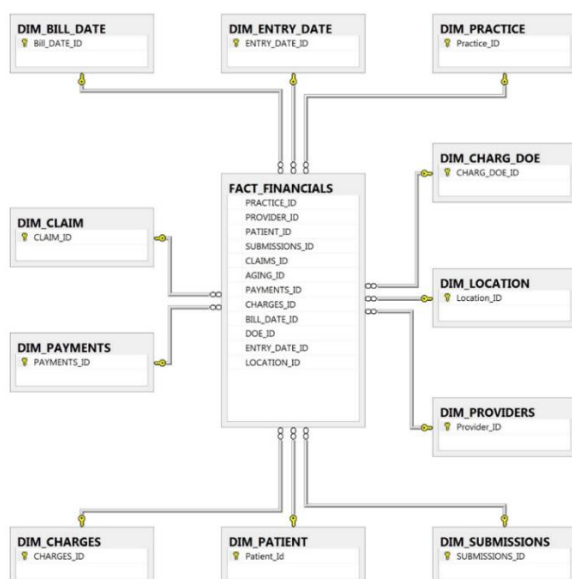


Figure 5: Proposed Database Schema.

Table 1: Warehouse Characteristics.

Name	Description
Subject Oriented	The situation when data/information is referring to a specific subject. For example, if an origination wants to analyze data for marketing department only. The Data warehouse devotion for a particular subject is the key factor for the subject oriented warehousing.
Integration	When data is fetched from different sources and stored coherently, by identifying differing sets of data into standard format. Origination can easily resolve their problems, discrepancies among units of measurement and will produce better output.
Nonvolatile	If data remains unchanged for the new developments. Once data stored into database, data should not be modified. It can be ensured by the data comparison.
Time variant	When data is stored into the system for the specific time and can be modified in different time intervals. Using a huge size of data and its spread over a long-time interval then analysts can divide it into different patterns and business associations.

process of historic medical data analysis, particularly in universities and hospitals. In healthcare system, the information about patient’s age, gender, location, and gender is taken at the time of service. Figure 5 describes database schema. The characteristics of the warehouse being built for sake of data visualization and analysis are enlisted in Table 1. It also includes the description of each type.

### 3.2 Data warehouse integration process

All the available data must be stored in a way that it consolidated into an information base functional data for company. This process is known as ETL (Extract, Transform, Load) [23-25] that as certain steps which are given below.

#### 3.2.1 Extraction

This step refers to data eliminating from its tale and making it available for further use. All the required data are fetched without disturbing the performance of source systems like response time and locking in a negative way. Cleaning phase is the first step in the process of ETL, in which data quality is confirmed from unification of data. The unification rules are making unique identifiers like gender, categories, phone numbers and zip codes changes into standard format and validation applied on address fields converting then into the proper format.

### 3.2.2 Transformation

In this step, we applied several rules to modify data from source into same dimensions so it can have same measurement of units. This ETL step also includes the joining of data from different sources that creates aggregates and surrogate keys also the validation process and new keys.

### 3.2.3 Loading

In this phase first disable all the constraints, then the indexes. Then starts the data loading process and then enables both constraints and indexes after the data loading

is complete. This step normally targets the loading from a database.

### 3.2.4 Design of dimensional model

Design of dimensional model needs to meet the requirements of industry standards which must have all the business needs and covers information that can be easily available. Components of the model are given in Table 2.

### 3.2.5 Data preprocessing

Data pre-processing is the important step, mainly used in identifying the missing values, false data, and repeated information from the dataset. We use the data that describes disease. The description is given in Table 3. The location (state) value is extracted from place of service where doctor office is located. For gender single character is used. Data is divided into four quarters. It can also be viewed month wise.

Table 2: Components of Dimensional Model.

Name	Description	Value
Dimension	Dimensions is the major component of design comprised of the individual keys and non-overlapping keys. The main purposes of dimensions are filtering, grouping, and labeling the dataset. Dimension tables can contain textual descriptions. Next column shows the dimensions.	Dim_Bill_Date, Dim_Entry_Date, Dim_Practice, Dim_Claim, Dim_Payments, Dim_Charge_Code, Dim_Location, Dim_Providers, Dim_Charges, Dim_Patient and Dim_Submission
Fact Table	Fact table data has measures or dependent attributes. Here the fact table is providing statistics for financial data broken by patient, claims, charges, and locations dimensions etc. Fact table usually contains historical data from operational system, it mainly has foreign key values which have many dimensions and numeric measure values on which the aggregation can be performed. The attributes in proposed Fact Table named FACT_FINANCIALS, are given in next column.	Foreign Key Column: practice_id, provider_id, patient_id, submission_id, claim_id, aging_id, payments_id, charges_id, bill_date_id, doe_id, entry_date_id, location_id. Measure: Patient Account.

### 3.3 Tools

In this research, two tools are used, SQL is used for the data warehouse and Tableau is used for data representation and visualization. Tableau is the leading visualization tool in market and is being used by several well reputed companies. It has so many visualization types and quite interactive interface. It also has a lot of functionalities to perform advanced filtrations and advance aggregations etc.

Figure 6 shows the Tableau home screen pertaining to the proposed system. On the left column, there are the navigation controls, in the middle column recently opened workbooks and sample workbooks can be accessed while in the third and last column pattern discovery options are enlisted.

In Figure 7, Tableau reader for the proposed warehouse can be seen where patient diagnosis code

Table 3: Data Description.

Name	Description	Value (Type)
Location	The location (state) value is extracted from place of service where doctor office is located.	String (Varchar)
Gender	Gender of the Patient exists in database as male/female. In dataset that is used for analysis purpose M/F values are there.	Char [M/F]
Month	Month wise data is dumped in data warehouse.	Number
Quarters	Data is divided in four equal quarters of a year.	Number

analysis is presented. Here the main screen shows the outcome on the map while the right-side column contains the control of attribute values that can be changed to view the data on the map. For example, the controls given are years in terms of date of service, gender, age range, quarters and month containing date of service.

Explanation of the controls is given in the Figure 8 (a-e).

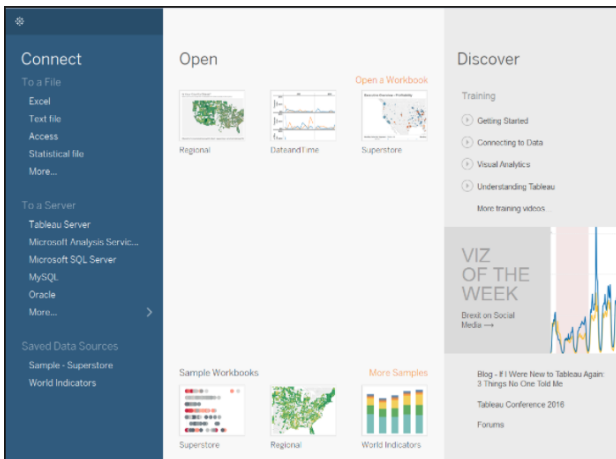


Figure 6: Tableau Home Screen.

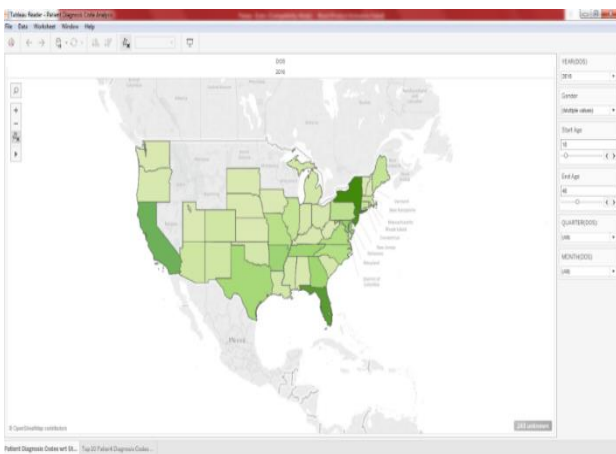


Figure 7: Main Screen.

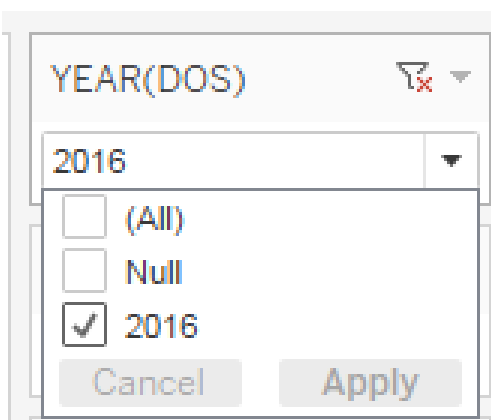


Figure 8 (a): Year selection.



Figure 8 (b): Year selection.

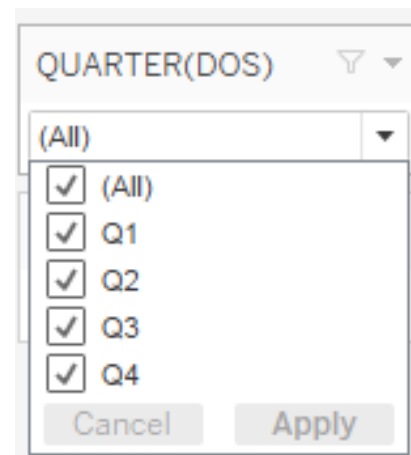


Figure 8 (c): Year selection.

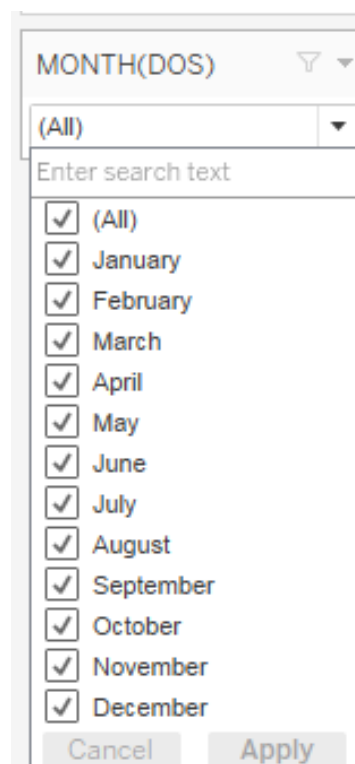


Figure 8 (d): Months of selected year.

**Top 10 Patient Diagnosis Codes wrt State**

Prac State	Dx Code1	Distinct count of Patient I..
AL	Z00.129	44
	I10	17
	Z23	27
	Z00.00	5
	E11.9	1
	J06.9	2
	B35.1	1
AR	L60.0	1
	I10	97
	Z00.129	118
	Z00.00	87
	R07.9	31
	L60.0	23
	J06.9	23
	E11.9	11
	Z23	11
	B35.1	4

Figure 8 (e): Data In table.

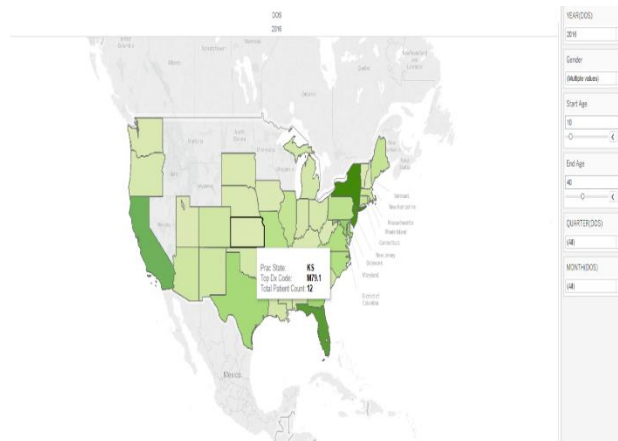


Figure 9: Location based analysis.

## 4 Results and discussion

In this section, results derived from the HealthCare IT Company datasets for year 2016 are presented. The data is taken from the designed data warehouse. The dataset contains the data for 2016 (the year of service) of patients for all the USA states. Following sections are dedicated to various analysis and visualization performed on various dimensions of the data warehouse.

### 4.1 Location based

Location has been extracted from the state (address) string of patient. Where patient live and not where they take service or the states of the service provider (Doctor/Clinic/Hospital). This is important because disease mapping/clustering is with respect to the location of patient. The results in Figure 9 depict that the patients (Male and Female) from Kansas (KS) State have been reported for the diagnostic (Dx) code (M79.1) which corresponds to a disease called Myalgia which is about muscular pain. It did include the time. Moreover, it is

Table 4: Comparison.

Parameter	[5]	[9]	[10]	Proposed
Disease	Dengue fever	Dengue larval growth	Dengue and Chikungunya	All types of diseases with registered US diagnostic code (Dx Code)
Region (Fixed or Multiple)	Fixed: City Muang of Phitsanulok Province, Thailand	Fixed: Fernando de Noronha, Brazilian oceanic island	Honduras Multiple states but not all states of US	Multiple All states of US
Gender based analysis	Yes	No	No	Yes
Time based analysis	No	Yes	Yes	Yes
Age based analysis	Yes	No	No	Yes
Location based analysis	Yes	No	Yes	Yes

based on whole provided data of all quarters with age limit 10 to 40 years.

### 4.2 Temporal

Time is one of the most important factors in USA Medical Health Sector. It is used to analyze various aspects for revenue generation for insurances and billing companies. Specially the first three months are the key to such analyses [26]. The results shown in Figure 10 (a & b) are from the two different states Iowa (IA) and New Hampshire (NH), respectively. The visual statistics illustrate that disease rate is much lower in these states based on the dataset. In IA, the patients are reported for Dx code I73.9, that corresponds to peripheral vascular disease (PVD). Similarly, in NH, the patients are reported for Dx code G47.33, that corresponds to Obstructive sleep apnea disease were reported during the given time range.

### 4.3 Age based

Below results are patients above then age 65. In USA mostly, patient can get Governmental insurances (Medicare and Medicaid) for treatment of various diseases. Because Government funds the patients above 65 years. The results given in Figure 11 (a & b) are for states California (CA) and New York (NY) for the patient with age more than 65 respectively. The results in Figure 11a, show that during year 2016, there were 10686 patients were reported for Top Dx code B35.1 that corresponds to Tinea Unguium disease which is most common fungus infection of the nails. That concludes that

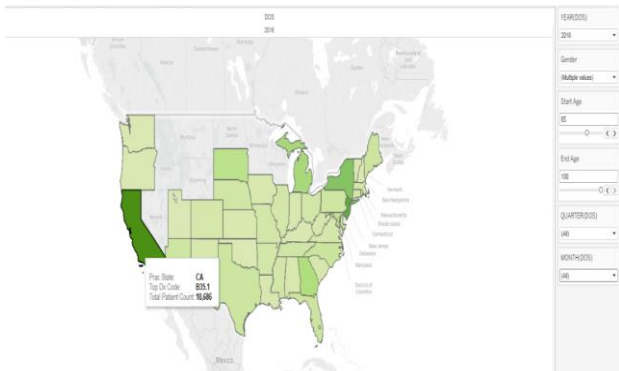


Figure 11: Age based analysis (CA State).

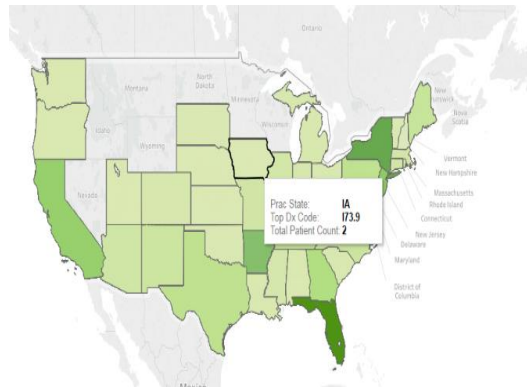


Figure 10 (a): Time based analysis (IA State).

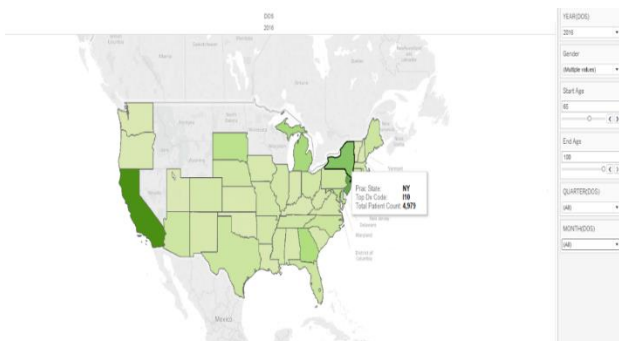


Figure 11: Age based analysis (NY State).

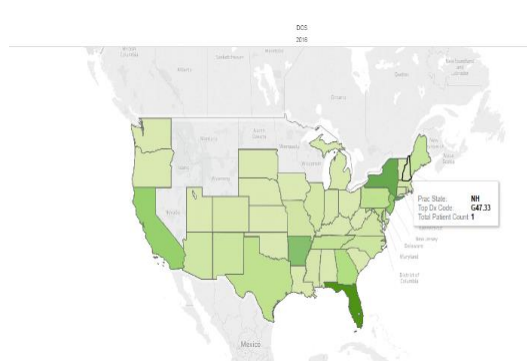


Figure 10 (b): Time based analysis (NH State).

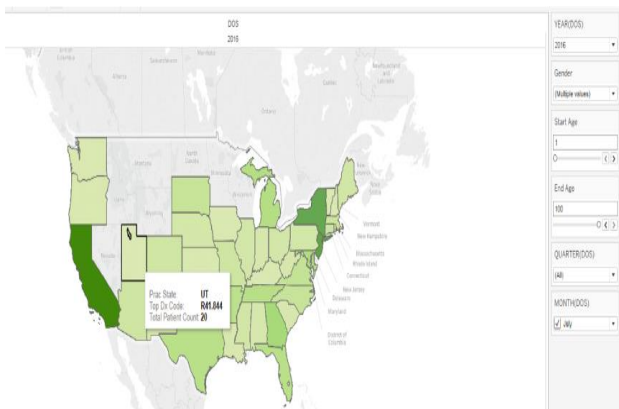


Figure 12: Month based analysis.

patient above 65 years are prone to this disease in California. Similarly, according to Figure 11b, 4979 patients were reported for Top Dx code I10 that corresponds to hypertension disease. That concludes that patient above 65 years are prone to this disease in New York [27-30].

#### 4.4 Monthly analysis

There are two types to report the insurance i.e., electronic and paper. USA government prefers to its doctors to use electronic way for fast processing. Figure 12 shows that in Utah (UT) state, the top Dx code for the month of July was R41.844 that corresponds to Frontal lobe and executive function deficit disease [31-40].

#### 4.5 Comparison

In this section, the proposed scheme has been compared with similar techniques in the literature qualitatively. The selection of the technique is based on the data, map, mapping, and visualization type. Table 4 shows the said comparison. From comparison it is apparent that proposed scheme has two major advantages over the other schemes. First one is complete range of diseases not just one type [41-46]. Second advantage is that the data can be analyzed and visualized for many dimensions like gender, location, time, and age-group. Moreover, it covers all the states of US while the schemes in [5] and [9] are for a specific city or zone. The scheme in [10], though provides analysis for several states of US, just works for Dengue and Chikungunya diseases analyses only for time and location based. Analysis for age-group and gender is mentioned as their future work.

The analysis made above, can be beneficial for healthcare stakeholders and government bodies for better decision making. In the current covid-19 pandemic outbreak, the proposed scheme can be beneficial in variety of ways. The spread can be visualized and observed in spatial and temporal perspectives. Further benefits may be obtained like:

- The patients may be guided based on their disease to refer to specialist doctors based on map history.
- Based on the analysis, advertisement for awareness on location of patients for a particular disease can be done.



- Based on temporal analysis disease growth rate can be monitored and remedies like vaccinations etc. can be initiated.

Based on gender, age group, state etc., alerts may be sent to public to take precautionary measures before they enter in that age group etc.

## 5 Conclusion

In this paper, Geospatial disease clustering has been proposed and designed which focuses on data visualization of healthcare domain. Map visualization is done by using data warehouse and Tableau tool. Data is collected and prepared from a Health Care IT Company. Datasets of different patients, states, genders, locations, doctors, clinics, and hospitals are transformed into data warehouse for service of year 2016 records consisting of different tables that contain all information related disease diagnostic code (Dx Codes). Data is transformed into desired shape for visualization and analysis purpose. The system has been developed which reads data of services, according to desired dimensions (Gender, Age, Month and Quarters) from dataset. Different results have been created which display association between different factors that can be used for decision making by the medical authorities. There are many dimensions that show disease trend. For decision making, in future, further analyses on these trends can be conducted. Data from other sources may also be gathered and data mining techniques may be investigated for further analysis. For example, to incorporate the covid-19 analysis, the appropriate datasets, databases, and other forms of data can be incorporated into the designed data warehouse easily and the analyses can be obtained on the fly.

## References

- [1] Khalid Farooq, Rai; Ashiq, Murtaza; Siddique, Nadeem; Rehman, Shafiq Ur; Adil, Hafiz Muhammad; and Ajmal Khan, Muhammad, "A Bibliometric Review of Highly Cited and Hot Papers on Coronavirus and COVID 19" (2021). *Library Philosophy and Practice* (e-journal). 5238. <https://digitalcommons.unl.edu/libphilprac/5238>.
- [2] Shueb, S., Gul, S., Nisa, N.T., Shabir, T., Ur Rehman, S. and Hussain, A. (2021), "Measuring the funding landscape of COVID-19 research", *Library Hi Tech*, Vol. ahead-of-print No. ahead-of-print. <https://doi.org/10.1108/LHT-04-2021-0136>.
- [3] Sarfraz, M.S., Tripathi, N.K., Tipdecho, T., Thongbu, T., Kerdthong, P. & Souris, M. (2012) Analyzing the spatio-temporal relationship between dengue vector larval density and land-use using factor analysis and spatial ring mapping. *BMC Public Health* 2012, 12:853
- [4] Sarfraz, M.S., Tripathi, N.K., Faruque, F.S., Bajwa, U.I., Kitamoto, A. & Souris, M. (2014) Mapping urban and peri-urban breeding habitats of *Aedes* mosquitoes using a fuzzy analytical hierarchical process based on climatic and physical parameters. *Geospatial Health* 8(3), 2014, pp. S685-S697.
- [5] Sarfraz, M.S., Tripathi, N.K. & Kitamoto, A. (2014) Near real-time Characterization of urban environments: a holistic approach for monitoring dengue fever risk areas, *International Journal of Digital Earth*, 7:11, 916-934.
- [6] A. Rahman, M.H. Salam, S. Jamil (2013) Virtual Clinic: A Telemedicine Proposal for Remote Areas of Pakistan, Conference: 3rd World Congress on Information and Communication Technologies, Vietnam.
- [7] A. Rahman, A. Bakry, K. Sultan, M.A.A. Khan, M. Farooqui, D. Musleh (2018) Clinical Decision Support System in Virtual Clinic, *Journal of Computational and Theoretical Nanoscience* 15(6):1795-1804.
- [8] A. Rahman, J. Alhiyafi (2018) Health Level Seven Generic Web Interface, *Journal of Computational and Theoretical Nanoscience* 15(4), DOI: 10.1166/jctn.2018.7302.
- [9] Regis, L. N., Acioli, R. V., Silveira Jr, J. C., de Melo-Santos, M. A. V., da Cunha, M. C. S., Souza, F., ... & Monteiro, A. M. V. (2014). Characterization of the spatial and temporal dynamics of the dengue vector population established in urban areas of Fernando de Noronha, a Brazilian island. *Acta tropica*, 137, 80-87.
- [10] Zambrano, L. I., Sierra, M., Lara, B., Rodríguez-Núñez, I., Medina, M. T., Lozada-Riascos, C. O., & Rodríguez-Morales, A. J. (2017). Estimating and mapping the incidence of dengue and chikungunya in Honduras during 2015 using Geographic Information Systems (GIS). *Journal of infection and public health*, 10(4), 446-456.
- [11] Rogers, D. J., Suk, J. E., & Semenza, J. C. (2014). Using global maps to predict the risk of dengue in Europe. *Acta tropica*, 129, 1-14.
- [12] Sanz-Arigitia, E. J., Schoonheim, M. M., Damoiseaux, J. S., Rombouts, S. A., Maris, E., Barkhof, F., ... & Stam, C. J. (2010). Loss of 'small-world' networks in Alzheimer's disease: graph analysis of FMRI resting-state functional connectivity. *PloS one*, 5(11), e13788.
- [13] Keihaninejad, S., Ryan, N. S., Malone, I. B., Modat, M., Cash, D., Ridgway, G. R., ... & Ourselin, S. (2012). The importance of group-wise registration in tract based spatial statistics study of neurodegeneration: a simulation study in Alzheimer's disease. *PloS one*, 7(11), e45996.
- [14] Carhart-Harris, R. L., Erritzoe, D., Williams, T., Stone, J. M., Reed, L. J., Colasanti, A., ... & Hobden, P. (2012). Neural correlates of the psychedelic state as determined by fMRI studies with

- psilocybin. Proceedings of the National Academy of Sciences, 109(6), 2138-2143.
- [15] Talawar, A. S., & Pujar, H. S. (2010). An outbreak of chikungunya epidemic in South India-Karnataka. *International Journal of Research and Reviews in Applied Sciences*, 5(3), 229-34.
- [16] Moraga, P. (2018) Small Area Disease Risk Estimation and Visualization Using R. *The R Journal* Vol. 10(1), pp. 495-506. July.
- [17] R. A. Naqvi, M. F. Mushtaq, N. A. Mian, M. A. Khan, A. Rahman et al., "Coronavirus: a "mild" virus turned deadly infection," *Computers, Materials & Continua*, vol. 67, no.2, pp. 2631–2646, 2021.
- [18] M. I. B. Ahmed, A. Rahman, M. Farooqui, F. Alamoudi, R. Baageel, A. Alqarni, "Early Identification of COVID-19 Using Dynamic Fuzzy Rule Based System," *Mathematical Modelling of Engineering Problems*, vol. 8, no. 5, pp. 805-812, 2021.
- [19] K. S. Alqudaihi, N. Aslam, I. U. Khan, A. M. Almuhaideb, S. J. Alsunaidi et al., "Cough sound detection and diagnosis using artificial intelligence techniques: challenges and opportunities," *IEEE Access*, vol. 9, pp. 102327-102344, 2021.
- [20] R. Zagrouba, M. A. Khan, A. Rahman, M. A. Saleem, M. F. Mushtaq et al., "Modelling and simulation of covid-19 outbreak prediction using supervised machine learning," *Computers, Materials & Continua*, vol. 66, no.3, pp. 2397–2407, 2021.
- [21] A. Rahman, K. Sultan, I. Naseer, R. Majeed, D. Musleh et al., "Supervised Machine Learning-based Prediction of COVID-19," *Computers, Materials & Continua*, vol. 69, no.1, pp. 21-34, 2021.
- [22] N. Min-Allah, B. A. Alahmed, E. M. Albreek, L. S. Alghamdi, D. A. Alawad et al., "A survey of COVID-19 contact-tracing apps," *Computers in Biology and Medicine*, vol. 137, p. 104787, 2021.
- [23] A. Rahman, F.A. Alhaidari, "Querying RDF Data", *Journal of Theoretical and Applied Information Technology* 26(22):7599-7614, 2018.
- [24] A. Rahman, F.A. Alhaidari, "The Digital Library and the Archiving System for Educational Institutes", *Pakistan Journal of Information Management and Libraries (PJIM&L)*, vol. 20 (1), pp. 94-117, 2019.
- [25] M. Ahmad, M.A. Qadir, A. Rahman et al., "Enhanced query processing over semantic cache for cloud based relational databases." *J Ambient Intell Human Comput* (2020). <https://doi.org/10.1007/s12652-020-01943-x>.  
[www.appliedmedicalsyste.ms.com](http://www.appliedmedicalsyste.ms.com)
- [26] A. Rahman et. al (2019) A Comprehensive Study of Mobile Computing in Telemedicine: Second International Conference, ICAICR 2018, CCIS, pp. 413-425, Shimla, India.
- [27] N. Aldhafferi, A. Alqahtani, A. Rahman, M. Azam (2018) Constraint Based Rule Mining in Patient Claim Data. *Journal of Computational and Theoretical Nanoscience* 15(3):1064-1071.
- [28] A. Rahman, Kiran Sultan, Dhiaa Musleh, Nahier Aldhafferi, Abdullah Alqahtani, and Maqsood Mahmud, "Robust and Fragile Medical Image Watermarking: A Joint Venture of Coding and Chaos Theories," *Journal of Healthcare Engineering*, vol. 2018, Article ID 8137436, 11 pages, 2018.
- [29] A. Rahman, M. Mahmud, K. Sultan, N. Aldhafferi, D. Musleh (2018) Medical Image Watermarking for Fragility and Robustness: A Chaos, Error Correcting Codes and Redundant Residue Number System Based Approach. *Journal of Medical Imaging and Health Informatics* 8(1):1192-1200.
- [30] A. Rahman, Kiran Sultan, Nahier Aldhafferi, Abdullah Alqahtani, and Maqsood Mahmud (2018) "Reversible and Fragile Watermarking for Medical Images," *Computational and Mathematical Methods in Medicine*, vol. 2018, Article ID 3461382, 7 pages. <https://doi.org/10.1155/2018/3461382>.
- [31] A. Rahman, A. Bakry, K. Sultan, M.A.A. Khan, M. Farooqui, D. Musleh, "Clinical Decision Support System in Virtual Clinic", *Journal of Computational and Theoretical Nanoscience*, 15(6):1795-1804, 2018.
- [32] M.T. Naseem, I.M. Qureshi, A. Rahman, M.Z. Muzaffar, "Robust and fragile watermarking for medical images using redundant residue number system and chaos," *Neural Network World*, vol. 30, no. 3, pp. 177-192, 2020.
- [33] A. Rahman, S. Dash, & A.K. Luhach, "Dynamic MODCOD and power allocation in DVB-S2: a hybrid intelligent approach." *Telecommun Syst*, vol. 76, pp. 49–61, 2021. <https://doi.org/10.1007/s11235-020-00700-x>.
- [34] A. Rahman, "GRBF-NN based ambient aware realtime adaptive communication in DVB-S2." *J Ambient Intell Human Comput* (2020). <https://doi.org/10.1007/s12652-020-02174-w>.
- [35] I.A. Najm, J.M. Dahr, A.K. Hamoud, A.S. Alasady et al., "OLAP Mining with Educational Data Mart to Predict Students' Performance," *Informatica* 46 (2022): 11–19.
- [36] A. Rahman, S. Abbas, M. Gollapalli, R. Ahmed, S. Aftab et al., "Rainfall Prediction System Using Machine Learning Fusion for Smart Cities," *Sensors*, vol. 22, no. 9, pp. 1-15, 2022. <https://doi.org/10.3390/s22093504>.
- [37] N. M. Ibrahim, D. G. I. Gabr, A. Rahman, S. Dash, A. Nayyar, "A deep learning approach to intelligent fruit identification and family classification," *Multimedia Tools and Applications*, 2022.

<https://doi.org/10.1007/s11042-022-12942-9>.

- [38] T. M. Ghazal, H. AlHamadi, M.U. Nasir, A. Rahman, M. Gollapalli, M. Zubair, M.A. Khan, C.Y. Yeun, "Supervised Machine Learning Empowered Multifactorial Genetic Inheritance Disorder Prediction", *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 1051388, 10 pages, 2022. <https://doi.org/10.1155/2022/1051388>.
- [39] M Gollapalli, A. Rahman, D. Musleh, N. Ibrahim et al., "A Neuro-Fuzzy Approach to Road Traffic Congestion Prediction," *Computers, Materials and Continua*, vol. 72, no. 3, pp. 295-310, 2022.
- [40] A. Rahman, A. Alqahtani, N. Aldhafferi, M.U. Nasir, M.F. Khan, M.A. Khan, and A. Mosavi. 2022. "Histopathologic Oral Cancer Prediction Using Oral Squamous Cell Carcinoma Biopsy Empowered with Transfer Learning" *Sensors* 22, no. 10: 3833. <https://doi.org/10.3390/s22103833>.
- [41] A. Rahman, S. Abbas, M. Gollapalli, R. Ahmed, S. Aftab, M. Ahmad, M.A. Khan, and A. Mosavi. 2022. "Rainfall Prediction System Using Machine Learning Fusion for Smart Cities" *Sensors* 22, no. 9: 3504. <https://doi.org/10.3390/s22093504>.
- [42] G. Zaman, H. Mahdin, K. Hussain, A. Rahman, J. Abawajy and S. A. Mostafa, "An Ontological Framework for Information Extraction from Diverse Scientific Sources," *IEEE Access*, vol. 9, pp. 42111-42124, 2021. doi: 10.1109/ACCESS.2021.3063181.
- [43] N.A. Sajid, M. Ahmad, M.T. Afzal, A. Rahman, "Exploiting Papers' Reference's Section for Multi-Label Computer Science Research Papers' Classification," *Journal of Information & Knowledge Management*, vol. 20 (2), pp. 1-21, 2021.
- [44] A. Rahman, S. Dash, A.K. Luhach, N. Chilamkurti, S. Baek, Y. Nam, "A Neuro-Fuzzy Approach for User Behavior Classification and Prediction", *Journal of Cloud Computing*, 8(17), 2019.
- [45] A. Rahman, F.A. Alhaidari, D. Musleh, M. Mahmud, M.A. Khan, "Synchronization of Virtual Databases: A Case of Smartphone Contacts", *J. Comput. Theor. Nanosci.*, vol. 16 (3), pp. 1740-1757, 2019.

