

# Development of a web-geographical information system application for plotting tuberculosis cases

Nur Adibah Mohidem,<sup>1</sup> Malina Osman,<sup>2</sup> Farrah Melissa Muharam,<sup>3</sup> Saliza Mohd Elias,<sup>1</sup> Rafiza Shaharudin,<sup>4</sup> Zailina Hashim<sup>1</sup>

<sup>1</sup>*Department of Environmental and Occupational Health, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Serdang;* <sup>2</sup>*Department of Medical Microbiology, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Serdang;* <sup>3</sup>*Department of Agriculture Technology, Faculty of Agriculture, Universiti Putra Malaysia, Serdang;* <sup>4</sup>*Institute for Medical Research, National Institutes of Health, Shah Alam, Selangor, Malaysia*

Correspondence: Malina Osman, Department of Medical Microbiology, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia.

Tel.: +60.389472372. E-mail: malinaosman@upm.edu.my

Key words: Mapping; tuberculosis; programming; spatial analysis; web-geographical information system; Malaysia.

Acknowledgements: we would like to thank Leow Kang Ren for providing feedback for useful discussion during the development of Portal TB Gombak.

Contributions: NAM, MO, conceptualization and methodology; MO, validation; NAM, formal analysis and investigation, data curation, original draft preparation; ZH, resources and funding acquisition; NAM, MO, FMM, SME, RS, ZH, revise and editing; MO; FMM, SME, RS, ZH, supervisions. All authors have read and agreed to the published version of the manuscript.

Conflict of interests: the authors declare no potential conflict of interests.

Ethical considerations: the use of secondary data in this study was approved by the Medical Research Ethics Committee (MREC) of the Ministry of Health with the reference number: NMRR-17-3029-39236. The study subjects were represented by a number instead of their names and all information pertaining to the subjects was kept confidential.

Availability of data and materials: demographic data were extracted from the *MyTB* database; full access to which requires government authorization (<http://mytb.moh.gov.my/>). The original dataset is available from the corresponding author upon request.

Received for publication: 22 January 2021.

Revision received: 28 April 2021.

Accepted for publication: 26 May 2021.

©Copyright: the Author(s), 2021  
Licensee PAGEPress, Italy  
*Geospatial Health* 2021; 16:980  
doi:10.4081/gh.2021.980

*This article is distributed under the terms of the Creative Commons Attribution Noncommercial License (CC BY-NC 4.0) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.*

## Abstract

In the last few decades, public health surveillance has increasingly applied statistical methods to analyze the spatial disease distributions. Nevertheless, contact tracing and follow up control measures for tuberculosis (TB) patients remain challenging because public health officers often lack the programming skills needed to utilize the software appropriately. This study aimed to develop a more user-friendly application by applying the CodeIgniter framework for server development, ArcGIS JavaScript for data display and a web application based on JavaScript and Hypertext Preprocessor to build the server's interface, while a webGIS technology was used for mapping. The performance of this approach was tested based on 3325 TB cases and their sociodemographic data, such as age, gender, race, nationality, country of origin, educational level, employment status, health care worker status, income status, residency status, and smoking status between 1<sup>st</sup> January 2013 and 31<sup>st</sup> December 2017 in Gombak, Selangor, Malaysia. These data were collected from the Gombak District Health Office and Rawang Health Clinic. Latitude and longitude of the location for each case was geocoded by uploading spatial data using Google Earth and the main output was an interactive map displaying location of each case. Filters are available for the selection of the various sociodemographic factors of interest. The application developed should assist public health experts to utilize spatial data for the surveillance purposes comprehensively as well as for the drafting of regulations aimed at reducing mortality and morbidity and thus minimizing the public health impact of the disease.

## Introduction

Tuberculosis (TB), a highly prevalent infectious disease caused by the bacterium *Mycobacterium tuberculosis*, counts as one of 10 most serious communicable diseases in the world and one of the top 10 deadly infectious diseases worldwide. According to the World Health Organization (WHO), about 10 million new TB cases are reported each year and 1.5 million people die from this disease (WHO, 2020). Malaysia is categorized as an intermediate TB burden country, nonetheless the Ministry of Health in Malaysia (MOHM) accounts for 5 to 7 deaths per 100,000 population annually and about 20,000 to 25,000 new TB cases are recorded each year (MOHM, 2019). The mortality rate due to TB is not only the highest in the country compared to other infectious diseases, but transmission of the disease is on the rise; hence prevention measures are mandatory to keep this endemic under con-



trol. Previous studies have found that TB cases are associated with a list of sociodemographic factors such as age (Kang *et al.*, 2020), gender (Ben, 2020), race (Barroso, 2020), educational level (Taherian *et al.*, 2020), employment status (Sweeney *et al.*, 2020), number of health care workers per person (Phyu *et al.*, 2020), income (Kwak *et al.*, 2020), residency (Singh and Ramamohan, 2020), nationality (Cheng *et al.*, 2020), country of origin (Hill *et al.*, 2020), and smoking status (Adegbite *et al.*, 2020). From a public health management perspective, understanding the distribution of TB cases associated with these and also other sociodemographic factors over time is crucial for prevention and control programmes.

All health care providers in hospitals, clinics, disease prevention, control institutions and other designated health care establishments in Malaysia are obligated to promptly report active TB cases. Under the Prevention and Control of Infectious Diseases Act 1988 (Act 342), all laboratory confirmed TB patients are notified within a week and asked to report at the nearest district health office. Then, investigations of these notified TB cases are carried out within 24 hours by public health officers and the results registered into the TB Information System (TBIS) database. When finalized, the data are electronically transferred into the national database MyTB at MOHM.

Spatial epidemiology includes the interpretation and analysis of spatial data consisting of major interrelated approaches such as disease mapping (Cho *et al.*, 2020). Integration of different data including both spatial and non-spatial data assist an understanding of the distribution of TB cases at different scales. The spatial data basically involves the geographic dimension, *i.e.* the location and movement of cases while the non-spatial data refers to various variables, such as those mentioned above (Hoseini *et al.*, 2018; Kiani *et al.*, 2017). The spatial data can either be points showing the locations of reported cases or they can be polygon shapes that appear when point data are aggregated over sub-regions of the study area (Cai *et al.*, 2020). A map provides a simple visualization of the spatial data in the preliminary observation stage. Besides being more comprehensive than data tables and written reports, which require longer time to take in and be interpreted, disease mapping is the key step in epidemiological disease investigation that illustrates spatial variation; highlights the importance of relationships between cases and its risk factors; formulates and validates etiological hypotheses about disease transmission; and assists the development of public health prevention and surveillance programmes (Ferraro *et al.*, 2018).

Spatial statistics have participated in the increased accessibility to data based on geographic information systems (GIS). ArcGIS, a commercial package developed by ESRI (Redlands, CA, USA), has played an important role for epidemiological research by producing software that manages GIS services together with collected field data that not only provides various statistics packages but also connection to various web services available thanks to the Internet. Later on, Quantum Geographical Information System (QGIS) developed an open-source package that shares many features with ArcGIS (QGIS, 2015). Both ArcGIS and QGIS are currently the most useful tools for disease mapping. Nevertheless, the presence of expert on GIS and statistics is often necessary to operate even the simplest analysis and output because many researchers lack the appropriate statistical and programming skills to operate them efficiently.

In view of this, there is a broad range of spatial analysis techniques that can be used, *e.g.*, space-time scan statistic (Takahashi *et al.*, 2020), cluster analysis and hotspot detection (Azimi *et al.*, 2021; Montazeri *et al.*, 2020; Pishgar *et al.*, 2020), and spatial autocorrelation and spatial regression analysis (Goshayeshi *et al.*,

2019). However, running these spatial statistics will result in loss of each TB patient's original data, thus plotting each case's location is the only way to retain the original track. This is possible with GIS, a powerful tool to store, analyze, and display geographically-referenced data (Schultz and MacArthur, 2019) and to conduct various spatial analyses. These are functions that make GIS an excellent platform for creating a map that effectively presents the data and is more comprehensive for users from different educational backgrounds. Many commercial GIS applications are equipped with high functionalities developed by the latest software. Geospatial and geoinformatics experts must bear the cost of server development and annually license fees to continuously use these applications. In contrast, a web-GIS application is designed for a certain function for users that lack programming knowledge and skills at a fraction of the cost. It is thus an alternative to visualize disease pattern because it is easy to use and can be accessed anywhere as long as the user has a laptop or computer with Internet access.

Examples of web-GIS applications for case detection in high-risk areas that are easy to use and do not require the manual processing of the spatial data include the Map Maker tool (EMMa) of the European Centre for Disease Prevention and Control (ECDC) designed for communicable disease surveillance (ECDC, 2020); the Dot Mapper to plot the molecular clusters of TB cases spatiotemporally introduced by Smith and Hayward (2016); and the SpatialEpiApp, a web application using Bayesian models, scan statistics and R-integrated nested Laplace approximations to obtain risk estimates of lung cancer mortality (Moraga, 2017).

Disease mapping application of this kind with just a slightly different design would improve our understanding of the epidemiological characteristics of TB and be useful for prevention and targeted control strategies. To the best of our knowledge, a web application of interactive TB maps using sociodemographic variables has so far not been implemented in Southeast Asia. We aimed at creating a tool as a handy source for MOHM's TB investigations with appropriate ethical approval. Our approach is mainly addressed to public health officers responsible for case investigations but lacking GIS and programming knowledge and skills.

## Materials and methods

### Study area

This cross-sectional study used retrospective data of TB cases in the district of Gombak, Selangor, Malaysia (Figure 1). This is one of the administrative districts in Selangor, covering a total area of 650.08 km<sup>2</sup> between longitudes 101°34' and 14°6' East and latitudes 3°16' and 27.3° North. The district is divided into four sub-districts (mukims): Rawang, Batu, Setapak, and Hulu Kelang. Gombak's residential population was about 815,200 in 2018, which ranks fourth (after Petaling, Hulu Langat, and Klang) among the eight districts in Selangor; with approximately 90% of the population living in urban areas (Census, 2018).

Gombak has a tropical climate, which is hot and humid throughout the year, with an annual average temperature of 27.1°C and an annual rainfall of approximately 2535 mm (Malaysian Meteorological Department, 2018). The terrain is hilly in the eastern and parts of the northern and western regions, which are mostly covered by forests within the altitudes of 100 and 500 m above the mean sea level, while the central and south-western areas are lowlands with an average elevation between 30 and 70 m above the mean sea

level (Nur and Choy, 2016). The district is a part of the Klang Valley, a major municipal area in the Southwest, and Kuala Lumpur's rapid urbanization has spread to the southern part of Gombak in recent decades. Aside from that, the health sector comprises two government hospitals, eighteen government clinics, three private hospitals and three private clinics (Official Portal of Gombak Land and District Office, 2020). In 2018, Selangor accounted for the highest TB burden in the country, with 5,071 cases, and Gombak reported the highest TB incidence among its districts at approximately 700 cases each year (MOHM, 2018).

## Data sources

### Patient and sociodemographic data

In this study, the TB data between 1<sup>st</sup> January 2013 and 31<sup>st</sup> December 2014 were collected from the Rawang Health Clinic in Rawang, while TB data between 1<sup>st</sup> January 2015 and 31<sup>st</sup> December 2017 were collected from the TB/Leprosy Unit of Gombak District Health Office in Batu Caves. The confidential data were retrieved upon the approval from the Director of Gombak District Health Office *via* MyTB and double-checked from TBIS to prevent errors. The TB patients were diagnosed based on X-ray findings, pathogen detection and pathological diag-

nosis according to the criteria set by the MOHM (2008).

The sociodemographic variables of the current study of TB patients included age (years <15, 15-64, 64>), gender (male, female), race (Malay, Chinese, Indian, others), nationality (Malaysian, non-Malaysian), country of origin (Malaysia, Indonesia, Myanmar, Bangladesh, others), educational level (secondary school and below, secondary school and above), employment status (employed, unemployed), health care worker status (health care worker, non-health care worker), residency status (urban, rural), income status (permanent income, no permanent income), and smoking status (smoker, non-smoker). These variables were selected through the consideration of the real situation in Malaysia from the references of previous literatures (Tan *et al.*, 2020; Mohidem *et al.*, 2021). Additionally, we also included data such as individual IDs and date of diagnosis (month, year) when the TB diagnosis was confirmed.

Out of the 3590 TB cases retrieved, 181 cases (5.04%) were excluded because they concerned residents outside the study zone or the patient in question was not diagnosed within the study period. A further 2.46% of the cases were removed since they could not be geocoded due to incorrect, missing or unclear address information. This led to a final inclusion of 3325 cases (97.54%) to be aggregated at the mukim level.

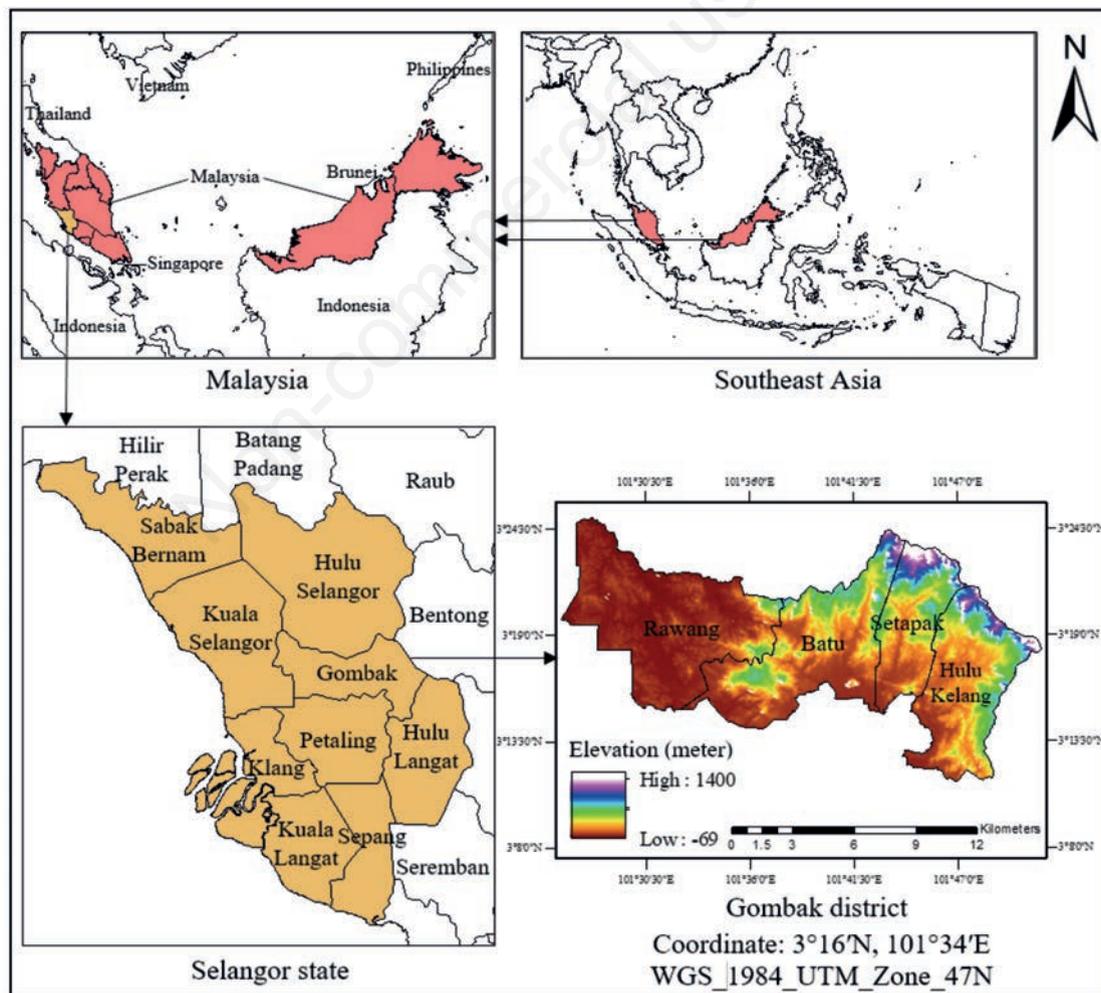


Figure 1. Geographical location of the study area at Gombak, Selangor, Malaysia.



### Geographical data

The Universal Transverse Mercator coordinates of the patients' residences were geocoded based on their home addresses recorded in the database using Google Earth™ version 7.15 (Alphabet Inc., Mountain View, CA, USA). However, some of the patients' addresses in rural areas were not detected by Google Earth; thus, coordinates of the bigger geographical areas such as village and street names were recorded. The address includes the location of the patients' home and institutional settings such as prison, nursing homes, hostels, and homeless shelters. The spatial unit of the dataset was point data, making sure that detailed information of all cases was retained. Altitude data showing topographic variation were derived from the 30m digital elevation model obtained from the Shuttle Radar Topography Mission provided by the United States Geological Survey Earth Explorer.

### Web-GIS application

#### Input data preparation

Portal TB Gombak demonstrated its features and implementation with a case study based on TB cases and its sociodemographic variables from years 2013 to 2017 in Gombak. Initially, the application user is required to prepare several categories of TB data, *i.e.* patients ID, address coordinates, mukim, month of diagnosis, year of diagnosis, age, gender, race, nationality, country of origin, educational level, employment status, health care worker status, income status, residency status, and smoking status. This information was represented in the data file as the numbers given in Table 1 forming a .csv file consisting of one column per data type and one row per TB case *i.e.* each TB patient.

#### System architecture and design

The web application is hosted on a dedicated server with a presented architecture comprising of CodeIgniter Framework as the application layer, ESRI Map and Web ArcGIS Service as service layer, and SQL Server as the data layer. CodeIgniter is a web application development tool created by EllisLab in 2006

(<https://codeigniter.com/>) that can be used to build websites based on the Hypertext Preprocessor (PHP) framework. It has a wide range of functionality to help developers speed up the formation of web applications, and the developers can publish libraries and plug-ins to CodeIgniter (Solanki *et al.*, 2017). Other advantages include its flexibility and efficiency in learning, updating, and integrating the Library and Helper, its use of the model-view-controller (MVC) pattern, which ensures that the structure of the generated code is more structured and has clear standards, thereby providing more security for web applications (Hustinawati *et al.*, 2014). The CodeIgniter Framework consists of four main components: the HyperText Markup Language 5 (HTML5), the Cascading Style Sheet (CSS), the JavaScript, and the PHP. Together, they build the server's interface that, in combination with webGIS technology, constitutes the server for the JavaScript application programming interface (API) on ArcGIS software (Figure 2). Application was hosted as a web page with its own URL (<https://tbgisgombak.com/arcgis/gis?i=1607073469154>) and accessed through the internet using a web browser. The SQL Server establishes a spatial database to store the attribute data, spatial data, and file data. In this study, the application interfaces allow users to insert and modify the TB data directly from the browser and control the application's layout using CodeIgniter Framework, which automatically changes the input. Further details of the setup are given below.

#### Application layer

The application layer refers to the application part of the whole system for the web application. The HTML5 and CSS widget was used for interactive web data visualization (HTML and CSS Standards, 2020). HTML5 was used because it has the elements to support multimedia without using Flash or other third-party plugins. Apart from that, PHP was utilized as the MapServer to import the data file to the website page using HTML5 code, acting as a platform to display spatial information. To achieve an online display of spatial inundation information, MapServer has four necessary files: a start webpage, spatial information and the HTML5 template. This platform is displayed on browser pages and as soon

**Table 1. Data preparation for plotting of tuberculosis cases in Gombak, 2013 and 2017.**

Data used	Type of information
Patient ID	Unique identifier for each case
X-coordinate	Longitude of the address of patient
Y-coordinate	Latitude of the address of patient
Mukim	Batu=1, Hulu Kelang=2, Rawang=3, and Setapak=4
Month of diagnosis	January to December designated from 1 to 12, respectively
Year of diagnosis	2013, 2014, 2015, 2016, 2017
Age	<15 years old=1, 15-64 years old=2, >64 years old=3
Gender	Male=1, Female=2
Race	Malay=1, Chinese=2, Indian=3, Others=4
Nationality	Malaysian=1, Non-Malaysian=2
Country of origin	Malaysia=1, Indonesia=2, Myanmar=3, Bangladesh=4, Others=5
Educational level	Secondary school and below=1, Higher than secondary school=2
Employment status	Employed =1, Unemployed=2
Health care worker status	Health care worker=1, Non-health care worker=2
Income status	Permanent income =1, No permanent income=2
Residency status	Urban=1, Rural=2
Smoking status	Smoker=1, Non-smoker=2

as the spatial data are captured, the results are presented in the users' web browsers according to the template files. Besides, the templates files also show the webpage image and the control panel for the design layer, such as opening/closing, zooming, panning, and other operations.

What can be said about JavaScript is that it is a scripting language written by the users by embedding into HTML5 documents accessible offline. It supports the development of highly responsive interfaces and maintains user interface presentation as well as offering dynamic functionality without waiting for server responses. Furthermore, a JavaScript programme can control what is visible on the screen by showing, hiding or changing the characteristics of the page elements (text, tables, page regions, and buttons) defined in HTML5. The purpose of JavaScript is to geocode reported cases and its sociodemographic variables of the study area. An additional feature of the web application is the geocoding of the TB patients' addresses according to the base map using JavaScript. Geolocation detection by JavaScript is compatible with different devices such as laptops, tablets and computer devices, creating a responsive web application and cross platform, as well as supporting several browsers such as Chrome, Mozilla Firefox, Safari, and Opera. Data Tables display interactive tables containing information on the TB cases and their sociodemographic variables. It supports filtering, pagination, and sorting, which helps locate and provide information on a particular area or show the areas with higher or lower values.

#### Service layer

The service layer is the web server connection between the user, GIS server, and database server. It is responsible for the web access control service, whereby it handles user base information, systems, and control. ESRI Map is a free, editable world map server that allows the sharing of visualization of the target areas without copyright restrictions (ESRI, 2020). The user interface base map in our system is such a map with several basic controls for panning, zooming, and switching views, which is very convenient when the map contains small areas. This part of the interface is built when the page is loaded, and the starting map location is a default location. Furthermore, the ESRI Map is selected so that the user will be able to operate the web app like the other mass-market internet mapping sites such as Google Earth and Google Map. In this study, the ESRI Map is used as a platform to facilitate the detection of high-risk areas of TB cases and provide visualization on the web.

Contrary to the complex, multifunctional, and established version of the GIS system, Web ArcGIS Service integrated with the Internet is broadly used in many fields. It focuses on the conversion of data format as well as usual function inquiry and display. This study used ArcGIS JavaScript API 3.7 developed by ESRI as the Web GIS technology (JavaScript API for ArcGIS Guide, 2020). Furthermore, the API key was added to the JavaScript code of the ESRI Map along with latitude and longitude of the TB cases. Thus, API displays the spatial data on the maps in the created web app, which substantially simplifies the information and allows sharing with other users without any software.

#### Data layer

The data layer is a function module dealing with database operation, all types of data and control commands, whereby any information must be synchronized with the database. My Structured Query Language (MySQL) is an open-source relational database management system that can retrieve rows of data, and there is a request of string named 'Query.' The SQL Server 2008

database preserves spatial information for the reported TB data in the current study.

#### Web geographical information system technology

Unlike the complex stand-alone versions of a GIS system, the web GIS technology combined with the Internet is extensively used in all fields. This is because it focuses on the conversion of data format as well as the common function survey, whereby as an effective tool for displaying spatial data. Furthermore, rate of data transmission and user convenience should also be recognized in the advancement of technology. The client uses a web browser, and the server provides the client with information and services. Web GIS technology has not only been able to perform the functions of conventional GIS yet must also be able to perform the functions of data acquisition, storage, processing, compilation, management analysis, and visualisation of spatial information. In this study, the functions of web-GIS application are map visualization.

## Results

### Features of the web-geographical information system application

The web application can be entered by single or multiple users (a group of users) at one time. It consists of four tabs: i) an 'INTRODUCTION' tab where the user can read the summary of the information about the web; ii) a 'GIS APPLICATION' tab

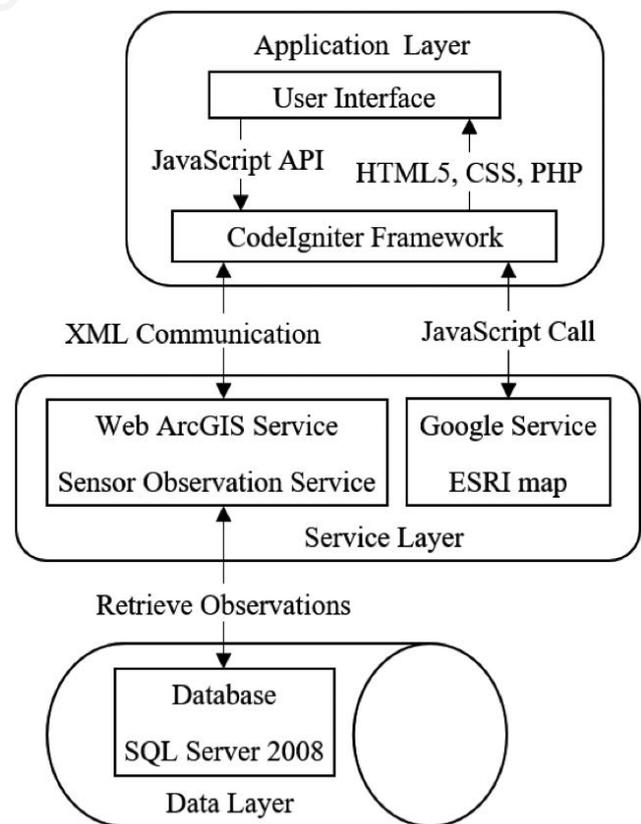


Figure 2. Architecture of the Portal TB Gombak.

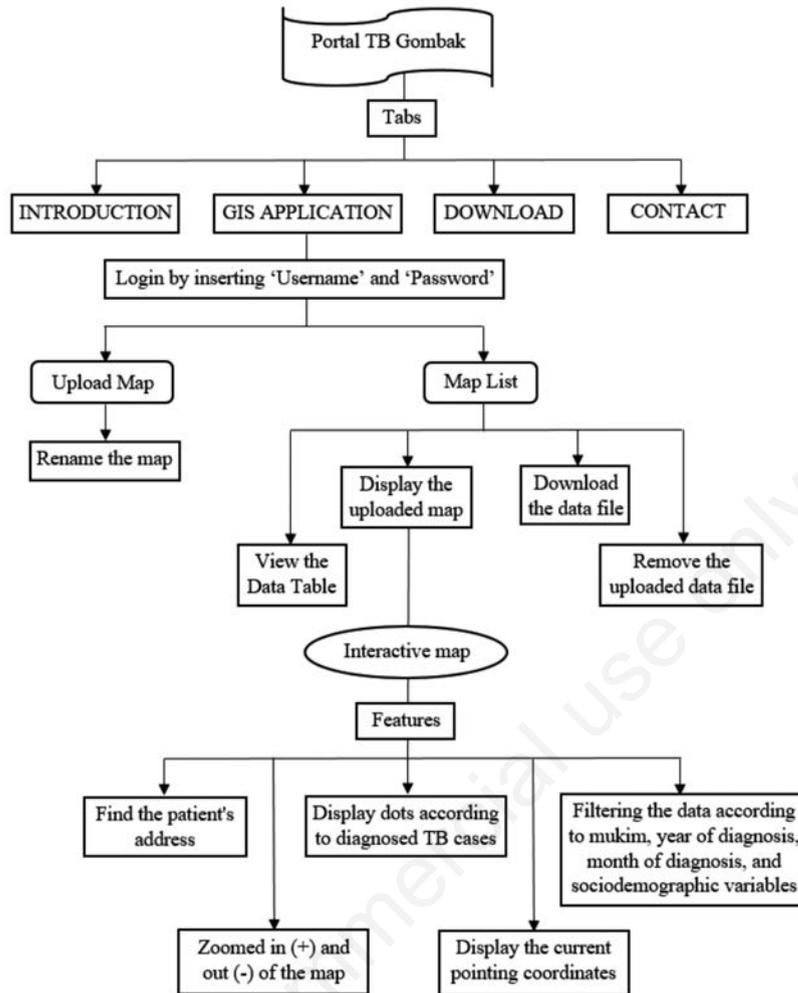


Figure 3. The flow of Portal TB Gombak.

where the data file can be uploaded and downloaded; and TB cases and its sociodemographic variables can be plotted and displayed; iii) a 'DOWNLOAD' tab containing a flow diagram on designing the web, publication for references regarding the sociodemographic factor of TB cases in Malaysia, and a user manual on how to use the web; and iv) a 'CONTACT' tab whereby the user can send queries to the authors via e-mail and letter using their affiliation address. The flow of design for Portal TB Gombak is presented in Figure 3. The main output of this web app is an interactive map, a user-friendly tool for plotting TB cases in the Gombak. It allows users to filter the distribution of TB cases based on location of cases, mukim, year of diagnosis, month of diagnosis, and sociodemographic variables. Once the webpage is accessed, the dashboard appears as shown in Figure 4. It is not an open-access web app; therefore, the user is required to insert a specific 'Username' and 'Password' in the columns provided to access the web app.

After login, the user can click the 'GIS APPLICATION' tab to upload the data file through the 'Upload Map' column (Figure 5). The data file can be inserted by clicking the 'Choose File' button and selecting the respective files. Then, the user needs to specify the name of data file in the 'Map Name' column and upload it by clicking the 'Submit' button. The 'Reset' button is present if the user wants to delete the uploaded data file or re-upload a new data

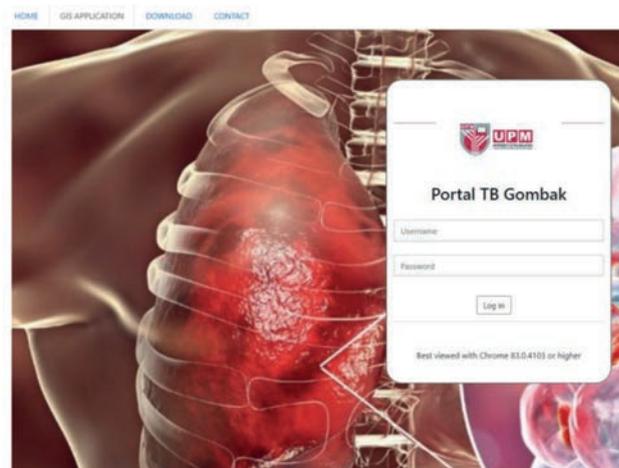


Figure 4. Dashboard of Portal TB Gombak's user interface. The picture shows the first page asking for 'Username' and 'Password' that are required before the system can be accessed after which those tabs 'HOME', 'GIS APPLICATION', 'DOWNLOAD' and 'CONTACT' will appear.

file. Meanwhile, the map ID is the unique identifier for the data of each TB patient from 1<sup>st</sup> January 2013 to 31<sup>st</sup> December 2017. The data file can be exported by clicking the ‘Download’ button to download the data in the plotted map.

Information of the uploaded data file can be displayed as a map through the ‘Map List’ column under the ‘GIS APPLICATION’ tab. By click the ‘Data’ button, the user can view the table form of the data consisting of latitude, longitude, mukim, year of diagnosis, and month of diagnosis. Users can also view and filter selected information from the table by entering a keyword in the ‘Search’ column and they can sort the rows in an ascending or descending order for data in any of the columns. However, this study cannot show all the data view variables because the large data size would cause a lag in the web application, thus needing a longer time to be loaded. Uploaded maps can be removed by clicking the ‘Remove’ button. To log out from the web application, there is a ‘Logout’ button on the top right corner of the dashboard. This study plotted the TB cases in a point form through the coordinate of longitude and latitude to prevent the loss of raw information for each patient. This means that all TB cases from 2013 to 2017 can be displayed by clicking the ‘Map’ button to generate a map on a new dashboard on the screen (Figure 6).

After clicking the ‘Map’ button in the ‘Map List’ column of the ‘GIS APPLICATION’ tab, the server will automatically generate the data as an interactive map. The distribution of TB cases from January 2013 to December 2017 refers to grey-coded dots, whereby

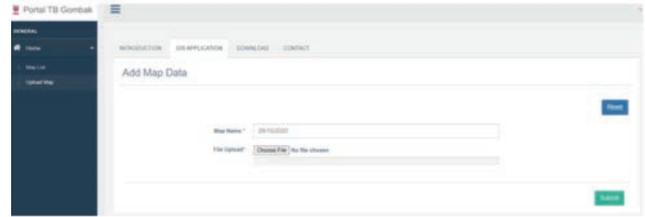


Figure 5. The ‘Upload Map’ column.

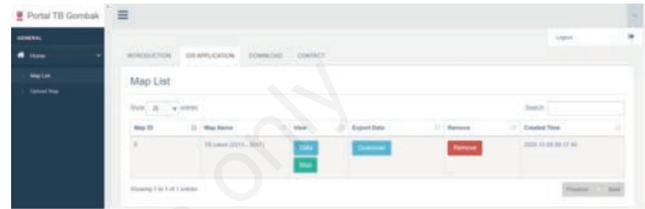


Figure 6. The ‘Map List’ column.

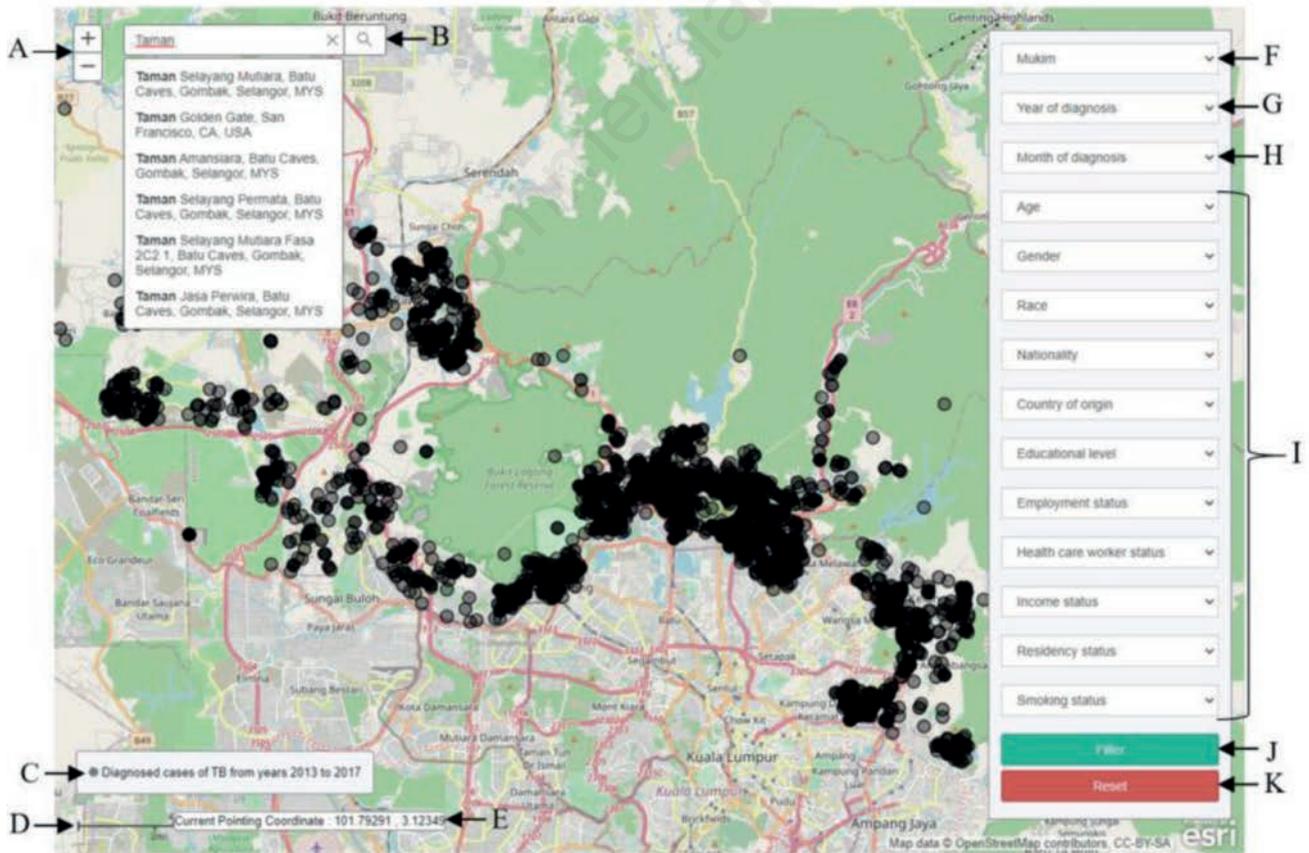


Figure 7. Example of the map with plotted TB cases. A) ESRI Map panning and zooming controls; B) search address; C) year of infection; D) scale bar; E) coordinate; F) mukim; G) year of diagnosis; H) month of diagnosis; I) sociodemographic characteristics; J) filter; K) reset button.

its legend is displayed at the bottom-left side of the dashboard. The user can also find patient addresses by searching a word at the dashboard's top-left side, which consists of the three buttons: 'ESRI World Geocoder'; 'Congressional Districts'; and 'Senator'. Besides, plotted cases can be changed and filtered according to mukim, year of diagnosis, month of diagnosis, and variables of the sociodemographic factor by clicking the 'Filter' button on the bottom-right side of the dashboard. Furthermore, the filter can be removed to display all cases by clicking the 'Reset' button. The map can be zoomed in (+) and out (-) by clicking the panning and zooming controls on the top-left side of the dashboard and scrolling the mouse to magnify or minimize the map. When the mouse is moved to any area, the dot's size representing TB cases remains the same until the user zoomed in to the maximum size of the targeted location when information of the targeted areas will be displayed. The 'scale bar' represents distance in km and the plotted coordinates in decimal degrees can also be seen on the bottom-left side of the dashboard (Figure 7). Clicking on a dot displays additional information on that reported case which includes mukim, year of diagnosis, month of diagnosis and each sociodemographic variable. This information is useful when investigating the locations of each sociodemographic variable (Figure 8).

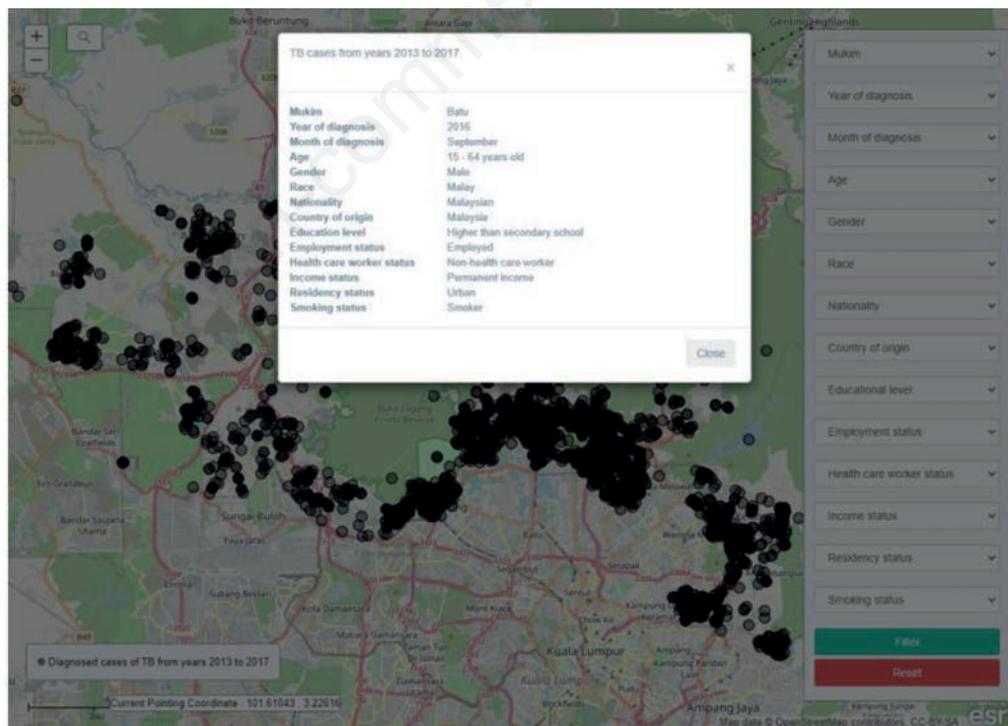
### Implementation of Portal TB Gombak

A case study was carried out to demonstrate how the system operates, *e.g.*, for anonymized data of sociodemographic variables for reported TB patients. By analyzing the uploaded data in the developed web application, a higher number of TB in southern, south-eastern, north-western and western Gombak were identified, whereas lower TB cases were detected in the North and North-East, while moderate TB cases were recorded in the South-West.

Using the 'Filter' button in the web app, the user can filter mukim and sociodemographic variables and the visualisation on the map will enhance the surveillance of risk areas.

Of the mukims, Batu revealed the highest number of TB cases (1466), while Setapak showed the lowest, *i.e.* 236 cases. The range in the annual period was between 574 and 700 cases, while in the monthly period, it was between 39 and 61 cases. The highest number of TB cases was found to be among patients aged from 15 to 64 years, followed by those aged more than 64 years and less than 15 years. Patients in the 15 to 64 range were highly concentrated between the South and South-East, and moderately distributed in the north-western, western and south-western parts of the region. Additionally, patients more than 64 years old were moderately dispersed and those less than 15 years were slightly scattered in the same zone with those in the 15 to 64 range. In terms of gender, the males presented a higher number of TB cases compared the females across the district. The infection was most common in the Malays followed by the Chinese, Indian and others. The 'others' refers to people not belonging to the three major race groups in Malaysia. Interestingly, this group included more TB cases than the two major race groups, *i.e.* the Chinese and Indian.

The patients who categorized as having received secondary education and below included those who had not undergone formal education through the Ministry of Education Malaysia and those who had completed preschool, primary school, and secondary school. On the other hand, patients categorized as having received secondary education and higher included those who had completed pre-university studies and similar. The plot map (Figure 9) shows that there was a higher number of TB cases among patients with secondary education and below compared those with higher edu-



**Figure 8.** Finding additional information on reported tuberculosis (TB) cases. The requested information appears by clicking on a dot (the grey dots were displayed when a TB case was plotted at a larger distance with other neighbouring cases while the black dots were appeared when the cases were plotted overlapped *i.e.* the same location of reported cases or when the cases were plotted at a smaller distance with other neighbouring cases, *i.e.* occurred when the user zooming out the dashboard).

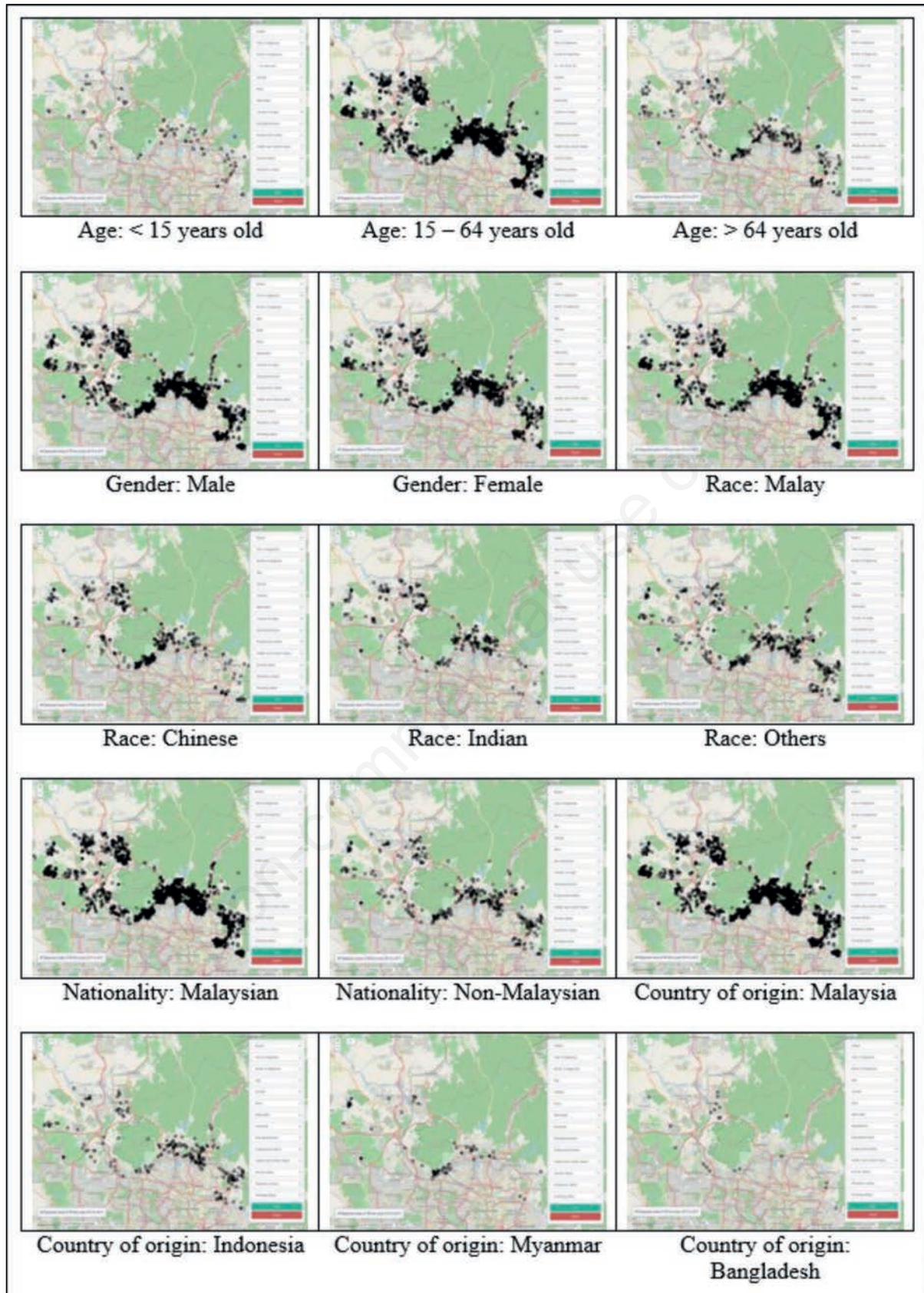


Figure 9. Distribution of sociodemographic variables of reported tuberculosis cases in Gombak in the period 2013-2017.

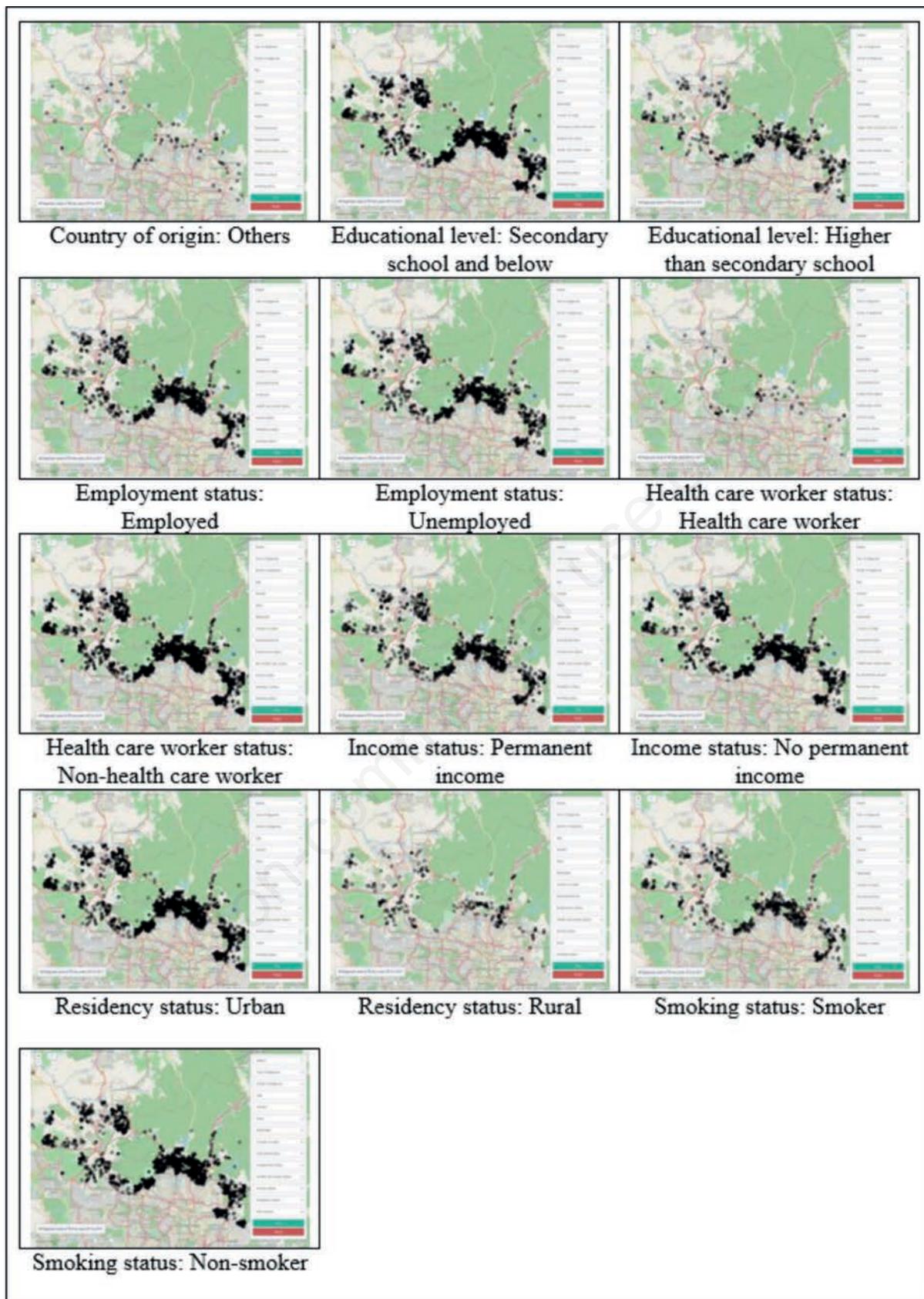


Figure 9. Distribution of sociodemographic variables of reported tuberculosis cases in Gombak in the period 2013-2017.

cation. However, there were no clear differences regarding the distribution of cases between these two categories, whereby the same pattern of geographical distribution was also seen between employed and unemployed patients and between those with permanent employment and those without. In contrast, few number of TB cases were scattered among health care worker compared to non-health care workers over the district. Furthermore, a higher number of TB cases was also displayed in urban areas compared to rural ones. In terms of behaviour, although most of the patients were non-smokers, yet the displayed distribution of patients being smokers was quite high (Figure 9).

TB cases among non-Malaysians were distributed at the certain areas of the district which includes Batu Arang, Bandar Country Homes, Rawang, Bandar Tasik Puteri, Kundang, Sungai Buloh prison, Kuang, Sungai Buloh, Kepong, Selayang, Batu Caves, International Islamic University Malaysia, Ukay Perdana, Bukit Antarabangsa, Kampung Dato' Keramat, and Bukit Kembara (Figure 10). Patients from Indonesia contributed the highest cases which disseminated from south to South-East, North-West, and South-West, followed Myanmar which there is a notable group cases in the south and South-West and Bangladesh which appear to be dispersed across the district (Figure 9). At the International Islamic University Malaysia, there are five TB cases among international students *i.e.* those who are non-Malaysian. The other 13

non-Malaysian TB patients were among the inmates and prison workers at the Sungai Buloh prison. Accordingly, plotting of TB cases demonstrated the use of this web app in targeting interventions for specific risk population which is inmates and prison workers that contributes 220 cases.

## Discussion

Disease mapping is the most common approach in the surveillance of a disease and the Portal TB Gombak web application could increase the understanding of the spatial epidemiology of TB by assessing and visualizing the geographically-referenced data in a user-friendly way. This web application allows a convenient data presentation tool to detect patterns of TB distribution on the interactive maps that show disease progression in time and space. Experts with local knowledge would identify potential areas of transmission of TB that could assist targeted prevention and control measures of the disease through enlarged screening and contract tracing.

Another advantage of this web application is its web browser interface, which provides intuitive features of the base map (ESRI map) similar to those found on the internet such as Google Earth and Google Map that are familiar to users. It will allow the acquisition of spatial data by public health officers to investigate epidemiological disease without skills and knowledge for spatial

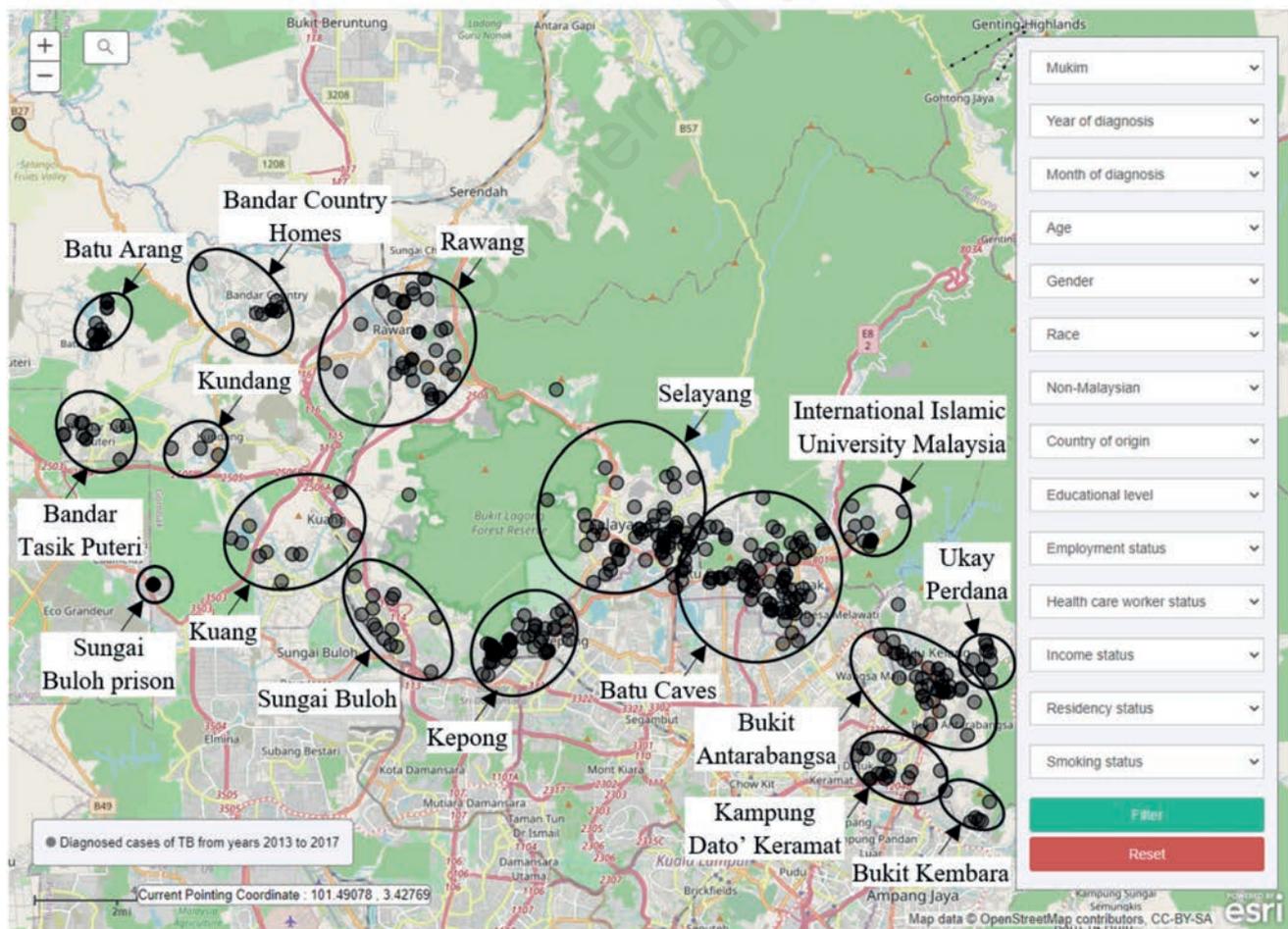


Figure 10. Location of reported tuberculosis (TB) cases among non-Malaysians in Gombak in the period 2013-2017.



statistics and programming. On top of that, this web app is an independent platform that does not require installing particular GIS software, thus significantly reducing server development cost and time. Users only require a web browser to access the entire functionality of this web application.

A future extension to Portal TB Gombak is to add a mobile version of the currently available web app for computers and laptops. The mobile application will enable live updates in the database on-site during surveillance. Furthermore, the server's effective architecture can manage user requests and rapidly respond by organizing the design flow in plotting TB cases within the same software environment without having to transfer data between different types of packages. Moreover, the advantage of using CodeIgniter Framework is the large user-base and online community that can assist in many circumstances. For instance, this kind of support could be advantageous during an outbreak and epidemic, where data can be updated regularly with convenience.

The current study showed how to follow distribution of TB cases and hotspot locations identified from earlier published literature (Mohidem *et al.*, 2021). Using Portal TB Gombak, researchers can conduct preliminary screenings of TB cases in the potential risk areas based on the localities of previously shown hotspot location and high-density regions making it possible to investigate the areas where cases have previously been reported and promptly conduct surveillance at the targeted areas. In addition, Portal TB Gombak's flexible nature was designed for convenient updates in the future, thus offering the opportunity for further web development. In this study, the web application only provides categorical variables of a dataset, *i.e.* sociodemographic variables plotted on the map. However, its functionality allows upgrading by adding specific datasets. In the future, researchers could replace or add the datasets of categorical variables with the datasets of continuous variables such as real age and total household income of TB patients, as well as adding environmental variables, such as air quality index, particular matter (PM<sub>2.5</sub>, PM<sub>5</sub>, PM<sub>10</sub>) and concentration of pollutants (carbon monoxide, nitrogen dioxide, sulphur dioxide, *etc.*), rainfall, relative humidity, temperature and more.

Researchers could customize and enlarge the dot colours representing the time axis by adding the more dataset in the Data Table. Furthermore, a smaller temporal unit such as daily or weekly could also be implemented. The operation of the web application by changing the display on the interactive map by enabling marker dragging, adding area data with polygons, overlaying images or using alternative base map tiles of various kind. This possibility enhances the web application to be useful for public health experts who analyse geostatistical or point process data. Additionally, the web app has the flexibility to be upgraded and extended to a model for predicting future TB outbreaks thereby assisting surveillance.

Although this study focused on TB data, this web application can be extended to any geocoded health data in other areas, *e.g.*, vector borne diseases such as dengue, food-borne diseases and sexually transmitted diseases. Nevertheless, there are several limitations to this web-GIS mapping application. Firstly, although the web application is a user-friendly tool, it is not supported and needs a technical expert to install and debug if necessary. It is written in HTML5, and extensive experience with ArcGIS JavaScript API 3.7 is essential for the initial set up. Secondly, it is important to realize the limitation of disease mapping as an approach for epidemiological disease investigations. The mapping does not account for the overall TB cases because the data input must include patients' addresses to plot the cases according to latitude and longitude values. Therefore, researchers must remove incom-

plete addresses and homeless people before uploading the data into the server, hence preventing the intervention for specific risk populations. Finally, while the web app is useful in the early steps to visualize the distribution of TB cases, it cannot conduct spatial statistics such as kernel density, hotspot, ordinary least squares, and geographically weighted regression that are available in GIS. Thus, this web app is not a replacement for formal software for the statistical analysis of spatial data.

## Conclusions

This study developed the Portal TB Gombak, an interactive web application to visualize the spatial distribution of TB cases. It has a user-friendly interface and allows surveillance without using GIS and programming skills. Specifically, it provides a convenient and rapid mapping of TB cases and their sociodemographic characteristics. Most importantly, a better understanding of TB distribution using this web application could enhance public health interest by facilitating targeted control measures through screening and contact tracing before treating and diagnosing the patients.

## References

- Abascal E, Herranz M, Acosta F, Agapito J, Cabibbe AM, Monteserin J, Chiner-Oms Á, 2020. Screening of inmates transferred to Spain reveals a Peruvian prison as a reservoir of persistent Mycobacterium tuberculosis MDR strains and mixed infections. *Sci Rep* 10:1-8.
- Abbasi T, Luithui C, Abbasi SA, 2020. A model to forecast methane emissions from topical and subtropical reservoirs on the basis of artificial neural networks. *Water* 12:145.
- Adegbite BR, Edoa JR, Achimi Agbo P, Dejon-Agobé JC, Essone PN, Lotola-Mougeni F, Zinsou JF, 2020. Epidemiological, mycobacteriological, and clinical characteristics of smoking pulmonary tuberculosis patients, in Lambaréné, Gabon: a cross-sectional study. *Am J Trop Med Hyg* 103:2501-5.
- Azimi A, Bagheri N, Mostafavi SM, et al, 2021. Spatial-time analysis of cardiovascular emergency medical requests: enlightening policy and practice. *BMC Public Health* 21:7.
- Barroso EG, 2020. Factors associated with household contacts' tuberculosis testing and evaluation. *Public Health Nurs* 37:705-14.
- Ben JM, Ben AH, Koubaa M, Hammami F, Damak J, Ben JM, 2020. Is there gender inequality in the epidemiological profile of tuberculosis?. *Tunis Med* 98:232.
- Cai J, Xie Y, Deng M, Tang X, Li Y, Shekhar, S2020. Significant spatial co-distribution pattern discovery. *Comput Environ Urban Syst* 84:101543.
- Census, 2018. Population distribution by local authority areas and mukims. Department of Statistics Malaysia, 2001. Available from: <https://www.selangor.gov.my>
- Cheng J, Sun YN, Zhang CY, Yu YL, Tang LH, Peng H, Zhao JM, et al, 2020. Incidence and risk factors of tuberculosis among the elderly population in China: a prospective cohort study. *Infect Dis Poverty* 9:13.
- Cho J, You SC, Lee S, Park, D, Park B, Hripesak G, Park RW, 2020. Application of epidemiological geographic information system: an open-source spatial analysis tool based on the OMOP Common Data Model. *Int J Environ Res Public Health* 17:7824.

- ESRI, 2013. ArcGIS Desktop. 102nd ed. Environmental Systems Research Institute, Redlands, CA, USA.
- ECDC, 2020. Available from: <https://www.ecdc.europa.eu/en/publications-data/ecdc-map-maker-tool-emma> Accessed: December 13, 2020.
- Ferraro OE, Guido D, Zambianchi R, Lanfranchi S, Oddone E, Villani S, 2018. Mortality for neurological diseases and pesticides: etiological hypotheses by a spatial analysis in the province of Pavia. *Med Lavoro* 109:420-34.
- Goshayeshi L, Pourahmadi A, Ghayour-Mobarhan M, Hashtarkhani S, Karimian S, Shahhosein Dastjerdi R, Eghbali B, SeyfiE, Kiani B, 2019. Colorectal cancer risk factors in north-eastern Iran: A retrospective cross-sectional study based on geographical information systems, spatial autocorrelation and regression analysis. *Geospat Health* 14:793.
- Hamzah IS, Sarifin MR, Aziz MSA, Abdullah MFA, 2020. Malaysia as attraction of international foreign workers. *J Crit Rev* 7:2020.
- Hill AN, Cohen T, Salomon JA, Menzies NA, 2020. High-resolution estimates of tuberculosis incidence among non-US-born persons residing in the United States, 2000-2016. *Epidemics* 33:100419.
- Himawan AK, 2014. Performance analysis framework codeigniter and CakePHP in website creation. *Int J Comput Appl* 94:6-11.
- Hoseini B, Bagheri N, Kiani B, Azizi A, Tabesh H, Tara, M, 2018. Access to dialysis services: a systematic mapping review based on geographical information systems. *Geospat Health* 13:3-10.
- Kang W, Du J, Yang S, Yu J, Chen H, Liu J, Zong P, 2020 The prevalence and risks of major comorbidities among inpatients with pulmonary tuberculosis in China from a gender and age perspective: A large-scale multicenter observational study. *Eur J Clin Microbiol Infect Dis* 40:787-800.
- Kiani B, Bagheri N, Tara A, Hoseini B, Tabesh H, Tara M, 2017. Revealed access to haemodialysis facilities in northeastern Iran: Factors that matter in rural and urban areas. *Geospat Health* 12:584.
- Kwak N, Winters N, Campbell JR, Chan ED, Gegia M, Lange C, Yim JJ, et al, 2020. Changes in treatment for multidrug-resistant tuberculosis according to national income. *Eur Respir J* 56:2001394.
- Malaysian Meteorological Department, 2018. General climate of Malaysia. Ministry of Science, Technology and Innovation, Kuala Lumpur, Malaysia. Available from: <http://www.met.gov.my/>
- Ministry of Health Malaysia, 2019. Annual Report 2000-2005. Disease Control Division, Ministry of Health, Kuala Lumpur, Malaysia.
- Ministry of Health of Malaysia, 2019. Annual report 2018: TB control programme in Malaysia. Ministry of Health, Kuala Lumpur, Malaysia.
- Mohidem NA, Hashim Z, Osman M, Shaharudin R, Muharam FM, Makeswaran, 2018. Demographic, socio-economic and behavior as risk factors of tuberculosis in Malaysia: a systematic review of the literature. *Rev Environ Health* 33:407-21.
- Montazeri M, Hoseini B, Firouraghi N, 2020. Spatio-temporal mapping of breast and prostate cancers in South Iran from 2014 to 2017. *BMC Cancer* 20:1170.
- Moraga P, 2017. SpatialEpiApp: A Shiny web application for the analysis of spatial and spatio-temporal disease data. *Spat Spatiotemporal Epidemiol* 23:47-57.
- Nur HA, Choy L, 2016. Analysis of land use and land cover changes in Gombak, Selangor using remote sensing data. *Sains Malays* 45:1869-77.
- Official Portal of Gombak Land and District Office, 2020. Available from: <https://www2.selangor.gov.my/>
- Phyu MH, Sriplung H, Kyi MS, San CC, Chongsuvivatwong V, 2020. Comparison of latent tuberculosis infections among general versus tuberculosis health care workers in Myanmar. *Tropical Med Infect Dis* 5:116.
- Pishgar E, Fanni Z, Tavakkolinia J, 2020. Mortality rates due to respiratory tract diseases in Tehran, Iran during 2008-2018: a spatiotemporal, cross-sectional study. *BMC Public Health* 20:1414.
- QGIS Geographic Information System, 2015. Open Source Geospatial Foundation Project, 2.10.1 edn.
- Schultz SD, MacArthur R, 2019. Geographic information systems and rural data. Rural data, people, and policy: information systems for the 21st century. Routledge, London, UK, pp. 189-204.
- Singh H, Ramamohan V, 2020. A model-based investigation into urban-rural disparities in tuberculosis treatment outcomes under the Revised National Tuberculosis Control Programme in India. *PLoS One* 15:e0228712.
- Smith CM, Hayward AC, 2016. DotMapper: an open source tool for creating interactive disease point maps. *BMC Infect Dis* 16:1-6.
- Solanki N, Shah D, Shah A, 2017. A Survey on different framework of PHP. *Int J Latest Technol Eng Manag Appl Sci* 6:155-8.
- Sweeney S, Vassall A, Guinness L, Siapka M, Chimbindi N, Mudzengi D, Gomez GB, 2020. Examining approaches to estimate the prevalence of catastrophic costs due to tuberculosis from small-scale studies in South Africa. *Pharmacoeconomics* 38:619-31.
- Taherian A, Akhlaghi M, Sadat Hosseiniun Z, Shahrestanaki E, Tiyuri A, Sahebkar M, 2020. Investigating the effect of education on knowledge and practice in preventing tuberculosis in eastern Iran. *Int J Health Promot Educ* 58:83-91.
- Takahashi K, Shimadzu H, 2020. Detecting multiple spatial disease clusters: information criterion and scan statistic approach. *Int J Health Geogr* 19:1-11.
- World Health Organization, 2020. Global tuberculosis report 2020. Available from: <https://www.who.int/teams/global-tuberculosis-programme/data> Accessed: December 24, 2020.