

Tuberculosis Prevention Model in Developing Countries based on Geospatial, Cloud and Web Technologies

Innocent Mwila¹

Department of Electronics and Electrical Engineering
University of Zambia
Lusaka, Zambia

Jackson Phiri²

Department of Computer Science
University of Zambia
Lusaka, Zambia

Abstract—Information is important when making decisions. Decisions which are based on gut feeling and made in the absence of evidence always tend to be less effective in most situations. This is also the case when it comes to Tuberculosis (TB) disease control and prevention intervention planning and implementation. The lack of evidence-based information upon which decisions for action to help with the prevention of spread of TB has proved to be less effective in the prevention of the disease as TB keeps spreading. The aim of this paper was to design and develop a prototype system that would provide TB program managers with information and tools which can be used to make decisions which can effectively influence the fight against the spread of TB through the application of cloud computing, geospatial data analysis and web technologies. The system would improve disease monitoring and tracking through the use of the identified technologies, by displaying the geographical distribution of TB cases in the communities on a mapping application as well as providing reports which TB program managers can use to make decisions when planning and implementing disease control and prevention activities.

Keywords—Evidence-based; monitoring; cloud computing; geospatial data analysis; mapping; web technologies; information; decision-making; tuberculosis prevention

I. INTRODUCTION

TB is one of the leading infectious diseases in the world with an annual average of 10 million cases being reported between 2016 and 2018, and causing an average of 1.4 million deaths in the same period which translates into 25% of deaths globally [1][2][3][4] and therefore, making TB the number one cause of death by a single infectious agent [1]. TB caused 17,000 deaths in Zambia in 2015 [5] and an average of 36,000 cases have been recorded annually between 2016 and 2018 in the country [6]. Zambia is among the 30 high TB burden countries in the world [7]. TB is an infectious disease which spreads from an infected person to the non-infected, and one way to prevent the spreading of the disease is by identifying individuals who are infected and then put measures in place to curtail the spread, to protect the ones not infected and Information and Communication Technologies (ICTs) can be utilized for this purpose, to improve surveillance [8].

ICTs play a significant and cornerstone role in the development of countries and the world at large [9] and the ability to make decisions, do business and communicate in this

information age dictates the rate at which social, political and economic development takes place as well as the advancement towards epidemic control [10]. ICTs have been used in different ways to improve disease surveillance [11][12] and thereby preventing further spreading or at least reducing the rate at which some diseases are spread.

ICTs have been identified as an important enabler in the fight against diseases and TB is not an exception [13][14]. Disease surveillance is the continuous, systematic monitoring, tracking, collection and analysis of data and the provision of information which leads to decisions being made and action taken to prevent and control infectious diseases [15]. The use of modern technologies has improved disease surveillance.

ICTs, particularly cloud computing, geospatial data analysis and web technologies have been less utilized in TB monitoring and tracking in Zambia. This has resulted in little being known about the geospatial distribution of TB cases and untimely to non-availability of other information about the disease. Information is not readily available when required for decision making and as such it is unclear as to which areas and communities TB program managers should urgently direct the resources in the fight against TB. This paper looks at how cloud computing, geospatial data analysis and web technologies can be applied to empower TB program managers with information needed to make decisions as they plan and implement disease control and prevention activities.

The aim of this paper was to design and develop a prototype system that would provide information which TB program managers can use for decision-making to help in the prevention of the spread of TB, through improved monitoring and tracking using cloud computing, geospatial data analysis and web technologies.

Firstly, the paper established factors which escalate the spread of TB in Zambia and then designed a model that would be used to monitor and track the spreading of TB in the communities based on cloud computing, geospatial data analysis and web technologies, and finally, based on the designed model, developed a prototype system to help with decision-making in the fight against the spreading of TB.

The remaining part of the paper is divided into six sections as follows; The literature review is given in Section II while

the related works are covered in Section III. In Section IV, the paper gives the details on the research methodology. The results are given in Section V, the discussion in Section VI and finally the conclusions and recommendations are given in Section VII.

II. LITERATURE REVIEW AND RELATED WORKS

In disease surveillance, the speed at which and the accuracy with which diseases or outbreaks are reported draws a difference between a successful or a failing disease control and prevention campaign. Leveraging the ability of ICTs such as real time reporting with the ability of Geographic Information Systems (GIS) to identify and display the geographical position of objects are being used to improve disease surveillance and this has resulted in improvements in disease prevention activities. The review looked at the TB disease, how it is caused, the factors which promote the spread of the disease, and how cloud computing, geospatial data analysis and web technologies have been used to support decision-making in countries such as India and South Africa to prevent the of spread TB through improved monitoring and tracking. The review also looked at some related works which have been done in the recent past in the health sector.

A. Tuberculosis and its Transmission Factors

TB is an airborne disease which mainly affects the lungs and can spread to other parts of the body [1]. It is caused by the Mycobacterium Tuberculosis bacterium and spreads from one individual to the other by the exhaling of the bacteria from the infected person's lungs and the subsequent inhaling of these bacteria by a non-infected individual [1][16]. The disease is preventable and being aware of the factors which lead to its spread is key in stopping further spread. Fig. 1 shows some of the TB transmission risk factors.

Susceptibility is the likelihood of one to catch a disease due to a weakened immune system [16] and the weaker the immune system the higher the likelihood of one to get infected with TB when exposed to the TB causing bacteria. The amount of the TB bacteria which an infected person can expel into the air when they cough or otherwise is known as infectiousness [17]. Environment factors such as the concentration of infectious droplets in the air, the exposure to the TB infected individuals in small enclosed spaces and inadequate ventilation [16][17] are among the prominent drivers in the spread of the disease. Ventilation is the flow of air in and out of a space to get rid of stale particles from that space [18]. Inadequate ventilation results in insufficient dilution or removal of infectious TB bacteria from the environment [16][17] and this supports the transmission of TB. Proximity and length of exposure to the TB causing bacteria are other factors pointed out by [16][17] to be key in the spread and transmission of TB, as longer duration of exposure to a person with infectious TB accelerates the risk of transmission of the disease to the people around that person. Proximity and prolonged exposure to the TB causing bacteria in homes and other settings is mainly promoted by crowding [16][17]. A house is said to be crowded if the ratio of the number of its occupants to that of the total number of rooms of that house is greater than or equal to 1.5 [19][20].

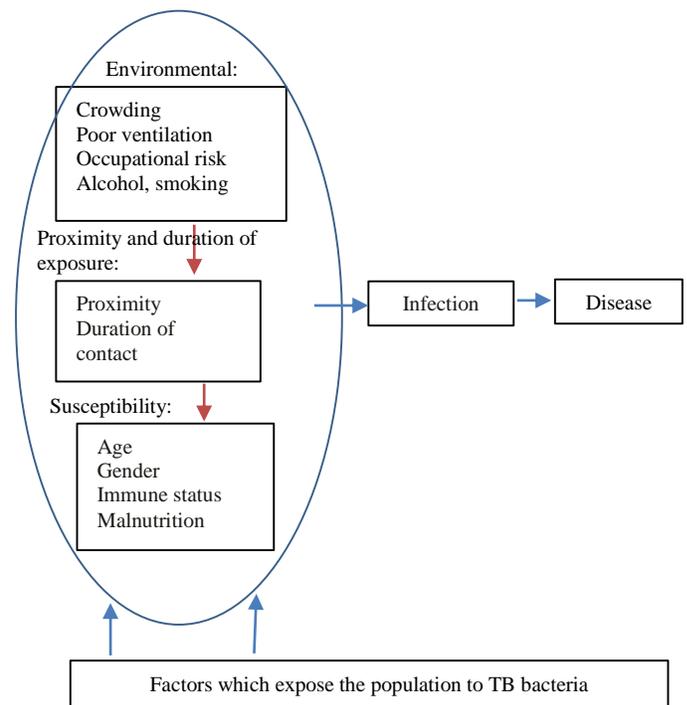


Fig. 1. TB Transmission Risk Factors [16][17].

Influencing the disease transmission factors which we have control over is important in the fight against the spread of TB and much of the emphasis in this fight must be placed on altering or being aware to these factors.

B. ICTs in Disease Surveillance and Spread Prevention

ICT is the capturing, processing, storing and transmission of information electronically using a variety of communication methods and devices such as satellite networks, mobile phones, wired and wireless networks as well as other digital forms [10]. The prevention of the spread of TB can be enhanced through better disease surveillance and ICTs can be harnessed for this purpose. ICTs have been used in different ways to improve disease surveillance [21][12] and thereby preventing further spreading or at least reducing the rate at which some diseases are spread. Using ICTs, in particular, the leveraging of cloud computing, improved data management (through database technologies), geospatial data management and web technologies in TB disease tracking and monitoring can help to prevent the spread of new cases or reduce the rate at which the disease is spread as has been the case in India, United States and Ireland [22][15].

C. Geo Spatial Data Analysis

Spatial data indicates where objects are within a given coordinate system while geo is the absolute or relative position of an object on the Earth's surface. Geospatial data, therefore, represents the positioning of an object on earth's surface and is associated with a coordinate system and so can be transformed onto a map. Vector data is used to depict a given point on a map and is in form of latitude and longitude pairs, which show the location of objects on a map and can be analyzed [23] and used to action decisions. The scrutinizing of spatial data to interpret the relationship of objects and other

associated attributes on the earth's surface is known as Geospatial Data Analysis [23] and Geographic Information Systems (GIS) are software systems used to capture, store, retrieve, analyze and display spatial data [24]. The advantages of GIS which includes cost effectiveness, handling of large volumes of spatial data and the ability for real time reporting have made GIS to be widely applied in disease surveillance and decision-making [25].

D. Cloud Computing

Cloud computing is the use of remotely hosted servers for data management instead of local computers and servers [26]. The servers, other computer resources, hardware and software are hosted offsite and made available via the internet by service providers. Cloud computing offers several advantages such as the pulling of resources to utilize economies of scale and scalability, cost efficiency, device independency, high computing power, location independency, redundancy and improved availability [26] while allowing users to use low specification computers to take advantage of maximum computing capabilities.

Cloud computing has presented itself as tool that can be harnessed for disease surveillance and help with the prevention of the spread of diseases by providing early warning systems for infectious diseases so that timely steps can be taken to avoid epidemic breakout [27]. Whereas traditional systems of disease monitoring are time consuming, slow and use less efficient data collection and monitoring systems, cloud computing provides speed, accuracy and efficiency in the epidemic outbreak detection using intelligent systems, databases, improved analytics, modeling, visualization and ontology mapping offering real time options for decision-making [27].

E. Disease Surveillance using Web Technologies

Web technologies collectively describe the types of technology used in communication over the Internet. The web of connections enables people to communicate and access files and documents located on other computers (and other electronic devices) in the network using web browsers. A web browser is a program that runs on client computers and requests for information, files, documents or services from other computers [28]. Examples of web browsers include Internet Explorer, Firefox, Chrome and Opera. The ability of web technologies to report disease occurrences in real time, at low cost and the simultaneous access to the data and other reports by various team players in disease surveillance regardless of one's location as long as internet connectivity is available are motivating factors. Although there are challenges which web-based applications are associated with, such as inaccurate information interpretation and privacy issues, it is recommended that web applications are applied in disease surveillance as the strengths, opportunities and advantages by far outweighs the challenges of tradition surveillance systems which are often expensive and inefficient due to inherent delays between disease occurrence and notification [29][30]. Quick and strategic decision-making are important in disease monitoring and enhanced surveillance can be made possible using web technologies [30] in epidemiological investigations as well as management and protection of exposed individuals.

F. Mobile Applications in Disease Surveillance

The convenience of being able to capture data from any place and send it to a database in a remote location using a device that one can move with to any place of choice makes mobile applications very handy for data collection [31][32] especially in disease monitoring and tracking. A mobile application is a software system that is programmed for and runs on a small movable handheld device and is accessible from any place.

Mobile applications are programmed for operating systems such as Android, Symbian and Windows, and they run on smart phones [33]. Main advantages of mobile applications include quick and real time communication, and less computation which results in less power consumption [31][34], which makes them a reliable option when power is a constraint as is the case in many developing countries. Mobile applications have been used in disease surveillance and have demonstrated reliability, accuracy and creativity upon which players in the health sector can rely for infectious disease monitoring and tracking [32][35].

III. RELATED WORKS

A. South Africa

Spatial data analysis was used in a research conducted by [36] in South Africa to determine the geographical distribution of TB cases in Western Cape's suburbs with the highest incidences of the disease. GIS was used to determine the distribution of the TB cases and the results of the research study showed that the TB cases were unevenly distributed in the communities with 1,835 out of the 5,345 (34.3%) residential homes housing at least 1 case of TB in the past decade while 3 or more cases occurred in 483 houses.

The study showed that the TB cases in a high incidence community spread unevenly though cases repeatedly occur in some households. The study further showed that there were higher incidences in households with the smallest plots in the communities, suggesting overcrowding. The lessons learned from this study were used to concentrate TB prevention services in specific areas as informed by GIS technologies.

B. India

The aim of the study under review was to identify geographical clusters with the prominent number of TB cases to design interventions of dealing with the escalating incidence of TB. In this study, reference [37] used spatial scan statistics to identify purely spatial and space-time clusters of TB.

The research was conducted in Almora District, Uttaranchal State, India. A spatial scan statistics software program was used to test the presence of statistically significant spatial as well as space-time clusters of TB and to identify their approximate locations. The geospatial analysis was done via GIS technologies with input data used being digital maps provided by the local government agency. The results of the study identified TB hotspots in the district. GIS eases the analysis and clearly displays the spatial patterns of disease distribution and can be used to provide insight into the extent and distribution of TB and more importantly, let TB

program managers use this information in the allocation of resources to fight the disease, intensify remedial measures and plan out future strategies for impactful TB control.

C. China

Timely information provision was an important factor in this study whose objective was to use mobile applications and Web GIS for decision support to respond to infectious disease emergencies. Reference [33] used geo data to display infectious disease emergencies via a Google Mapping API and the information collected by data capturers in the field using mobile phone applications was being shared via a 3G wireless network with decision makers in their respective workplaces who would then make decisions about the course of action in the fight against the spread of the infectious diseases. In this study, baseline data, geo coordinates of locations where infectious disease occurred and other disease parameters were captured throughout China, sent to the server and then the Web GIS application on the computer of the decision makers would process and summarize the data into reports to be used to effect action.

The architecture of the system under review, shown in Fig. 2 had two ends, the data collector interface sitting on an application of a phone or mobile computer and the decision maker interface, on a web browser of a computer.

The reviewed work illustrates how mobile and web-based applications can be used in disease surveillance to enhance decision-making in the face of infectious diseases and [33] recommends the use of the technologies used in this study for disease investigations especially in developing countries.

D. eIDSR

In the run up to and during the early stages of the 2014 to 2016 Ebola outbreak in Sierra Leone, the national health surveillance system, Integrated Disease Surveillance and Response (IDSR) was based on texting, calling or hand delivery of disease surveillance reports from health facilities to the district office for entry into the national system [38]. However, as the Ebola outbreak was peaking, this business model became inadequate as Ebola cases and other diseases' information needed to be availed in real time and accurately in order to allow prevention officers to act and prioritize their activities based on the information, and this led to the designing and development of electronic Integrated Disease Surveillance and Response (eIDSR) system.

eIDSR uses mobile devices for data entry from communities and health facilities, and the data is sent to the web-based national database which runs on DHIS2.

This system helped Sierra Leone to respond to the Ebola outbreak efficiently with reporting rates improving, from less than 40% before eIDSR deployment to greater than 97% with the error rate dropping by 45%.

E. eHealth Cloud

eHealth Cloud is a cloud-based health care system that was developed in Bangladesh to heterogeneously connect the

patients, doctors, the government and other players in the health care system to replace the paper-based system [39]. The lack of electronic medical information repository, the inability for doctors to see patient medical history, the impracticality of analyzing medical data and the absence of a heterogeneous communication platform among the health system players were major issues with the paper-based system.

A three tier model as shown in Fig. 3 on page 5 which was designed to address the problems noted in the paper based system has a client and server, and in between them is logic layer which implements the application logic rules and other system functions such as data querying and transfer as well as handling data security issues for eHealth Cloud.

eHealth Cloud offers the unified connected solution for all health care players and allows for seamless interaction and information sharing in the health care system with the capability of storing large volumes of medical data such as laboratory results, diagnostic information and prescriptions.

The front end of eHealth Cloud was developed using the Rich Internet Application platform and is accessible from both desktop and mobile devices while the cloud (database) server was implemented using SimpleDB. Representational State Transfer (REST) was used to link the front end to the cloud server which does all the processing so as to take the processing burden off the clients of eHealth Cloud.

F. FluMob

With the increasing cases of Influenza in Singapore, there was need for a real time monitoring system for tracking and reporting the disease and Flumob system was developed as a response to this need [40]. Flumob is used by health care workers to report influenza cases in real time using a mobile interface and web browsers of computers to the central server. Flumob generates reports which are analyzed and used for disease control and prevention purposes.

Flumob's nonfunctional requirements implementation are Windows server 2008 R2, Apache, PHP, MySQL, Android studio, and xCode for iOS development in terms of software, and a central server machine with four cores Core2 Intel Xeon Processor, 8 GB of RAM, and 500 GB of storage space.

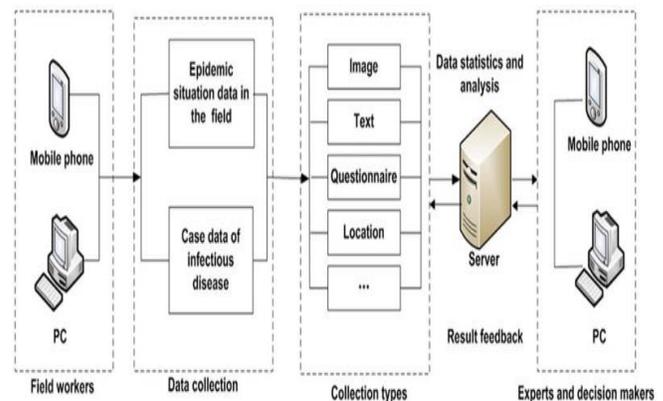


Fig. 2. Architectural and Application Data Flow Graphical Design [33].

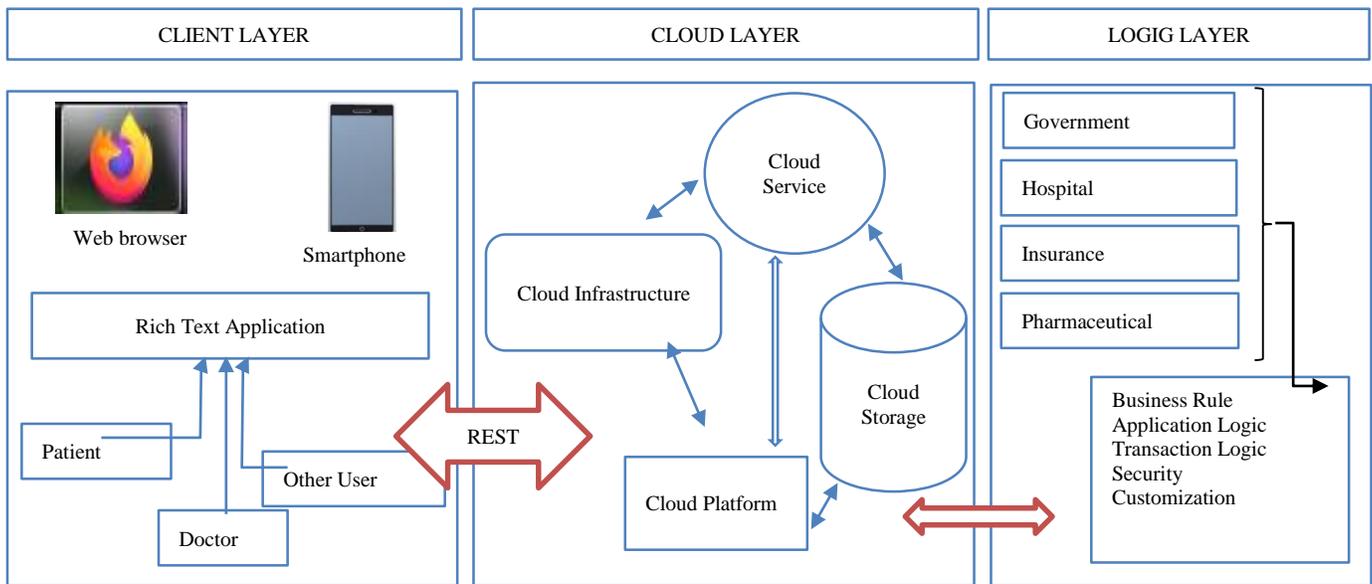


Fig. 3. Three Tier cloud System Design [39].

IV. METHODOLOGY

The aim of this paper was to design and develop a prototype system that would provide information which TB program managers can use for decision-making to prevent the spread of TB.

Firstly, the paper established factors which escalate the spread of TB in Zambia using a baseline study and then designed a model that would be used to monitor and track the spreading of TB in the communities based on cloud computing, geospatial data analysis and web technologies, and finally, based on the designed model, developed a prototype system to help with decision-making in the fight against the spread of TB using the Cloud Computing Architecture, Google Mapping API for displaying reported TB cases on a map, PHP for developing the web application frontend and MySQL as the data management platform.

A. Data Collection and Population Study Sample

The baseline study was carried out in Lusaka district of Lusaka province of Zambia. TB patients on TB treatment formed the sample study population. Participants of the baseline study were identified from clinical records at health facilities from which they were obtaining their TB drugs and randomly picked. The health facilities from which participants of the baseline study were recruited included Chawama Level One Hospital, Chipata Level One Hospital, Kalingalinga Clinic, Kanyama Level One Hospital and Mtendere Clinic Table I. shows the five health centers of Lusaka district with the respective number of participants.

The questionnaire for interviewing participants for the baseline study data collection was built around the transmission risk factors identified in the literature review (under Tuberculosis and Its Transmission Factors) and data collection was done using KoboCollect application running on Android as this allowed for the capturing of geo coordinates of participants' homes.

TABLE. I. POPULATION STUDY SAMPLE

Facility Name	Participants
Chawama First Level Hospital	31
Chipata First Level Hospital	59
Kalingalinga Clinic	25
Kanyama First Level Hospital	81
Mtendere Clinic	49

B. System Modelling

a) *Current Business Model:* Currently, the TB prevention routine is reactive and is mainly activated once a confirmed bacteriological TB case has been identified. The prevention routine in this model includes a contact tracing activity in which members of the household in which the bacteriologically confirmed TB case has been identified being screened for TB symptoms. In cases where funds are available, TB prevention activities are conducted in these communities in a one fits all manner, for example distribution of literature which teaches coughing etiquettes – the best way of coughing so as not to spread TB. During specific times of the year such as the TB month, which is commemorated in March annually, mass education about the disease is done, for example, through sending of text messages to people on a certain mobile communication network. These messages are generic and are not informed by any evidence. The flow chart in Fig. 4 shows the summary of the current TB prevention model in the communities.

b) *Proposed Business Model:* The proposed business model took the steps in the current model and incorporated the use of ICTs such as cloud computing, geospatial data analysis and web technologies to capture, store and process data. TB is preventable and being aware of the factors which lead to its spread is key in stopping the further spread. Influencing factors such as proximity and exposure as well as other

environmental factors which we have control over is important in the fight against the spread of TB and much of the emphasis in the fight must be focused on altering or being aware to these factors and spatial data analysis of already existing TB cases using GIS can help identify the geographical patterns of transmission and spread of the disease and therefore inform strategic decisions in the fight against TB. Once the data has been processed into maps and reports, TB managers can then use this information to plan and implement TB prevention activities in the communities.

The proposed model seeks to establish the likely causes of the given TB case, this process also involves the assessment of the patient's homes. This information is important when designing prevention activities. The likely causes of the disease are captured together with the details of the patient's house as well as the geo coordinates of that house and sent to the database in the cloud. The web application running on the TB program managers' computer retrieves maps showing the geo distribution of TB cases in the communities and other reports upon which disease prevention and control activities can be planned and implemented. The flow chart in Fig. 5 on page 7 shows the proposed business model.

C. Application Design

a) System Architectural Design: From the proposed business model, the architectural design of the system is as shown in Fig. 6 and works as illustrated; the system consists of the data capturing application which sits on the mobile device (Smartphone, Table, Laptop etc.) and is accessible via a web browser. The user captures data (including geo coordinates of patients' homes) on the mobile device and sends the data to the database in the cloud for storage. The

web application on the decision maker's computer, embedded with a GIS application is used to display geo locations of TB cases and to generate summary reports which can be used for strategic decision-making by TB program managers to help with TB prevention.

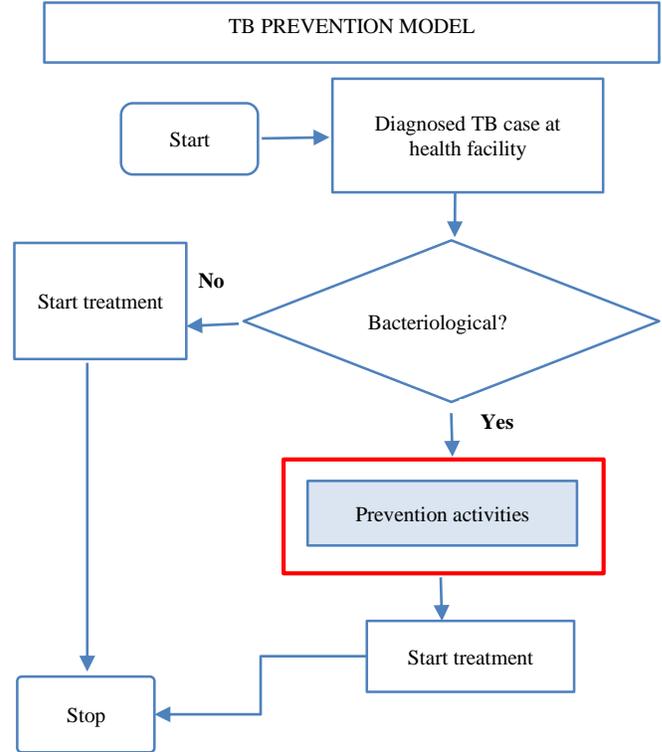


Fig. 4. Current Business Model Flow Chart.

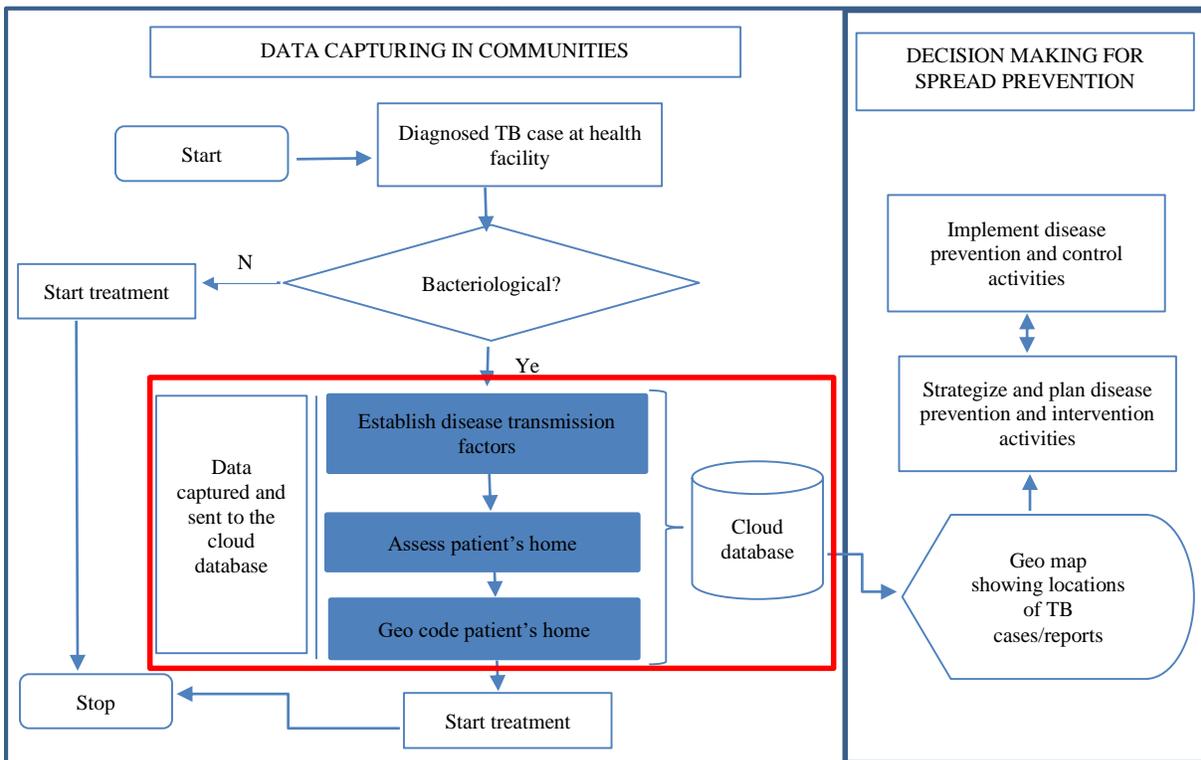


Fig. 5. Proposed Business Model Flow Chart.

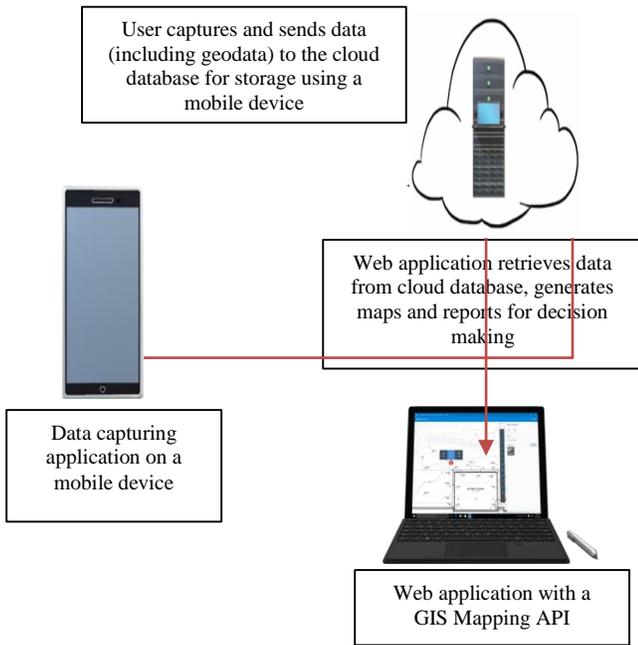


Fig. 6. Architectural System Design.

b) *Use Case Diagrams:* The system has two actors, the data capturing personnel, to capture data and the TB program

manager, who is the system administrator. The TB program manager views maps, runs summary reports and makes decisions about disease control and prevention activities based on the availed information. The TB program manager being the system administrator can also create user accounts for other system users. Fig. 7 and Fig. 8 respectively shows the use case diagrams for the data capturer and TB program manager.

c) *Sequence Diagram:* Using the architectural system design (Fig. 6) and the two use case diagrams (Fig. 7 and Fig. 8), the logical sequence of activities of the system prototype can be summarized as shown in the sequence diagram in Fig. 9 on the next page.

D. Data Design

MySQL is the Relational Database Management System (RDBMS) that was used to develop the data management platform for the prototype. MySQL is used to store, process and retrieve the data in the system. The data design of the application consists of seven tables, each table being the entity with its attributes. Fig. 10 shows the conceptual design of the system prototype database in the form of an Entity Relationship Diagram (ERD). It shows the database entities.

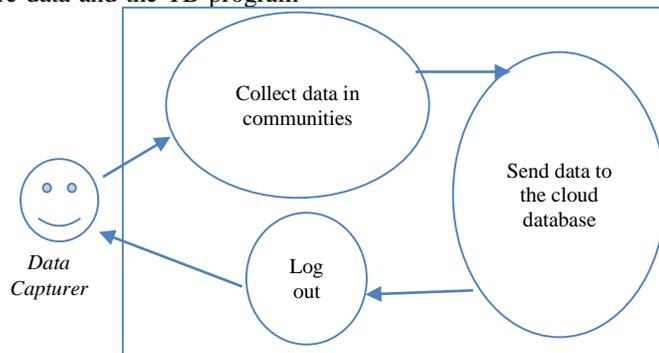


Fig. 7. Data Capturer use case Diagram.

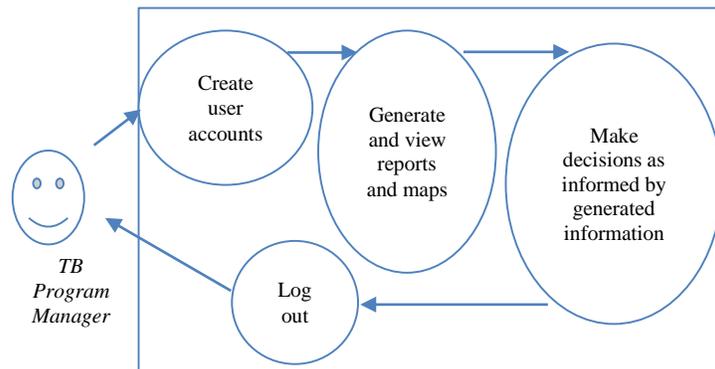


Fig. 8. TB Program Manager use Case Diagram.

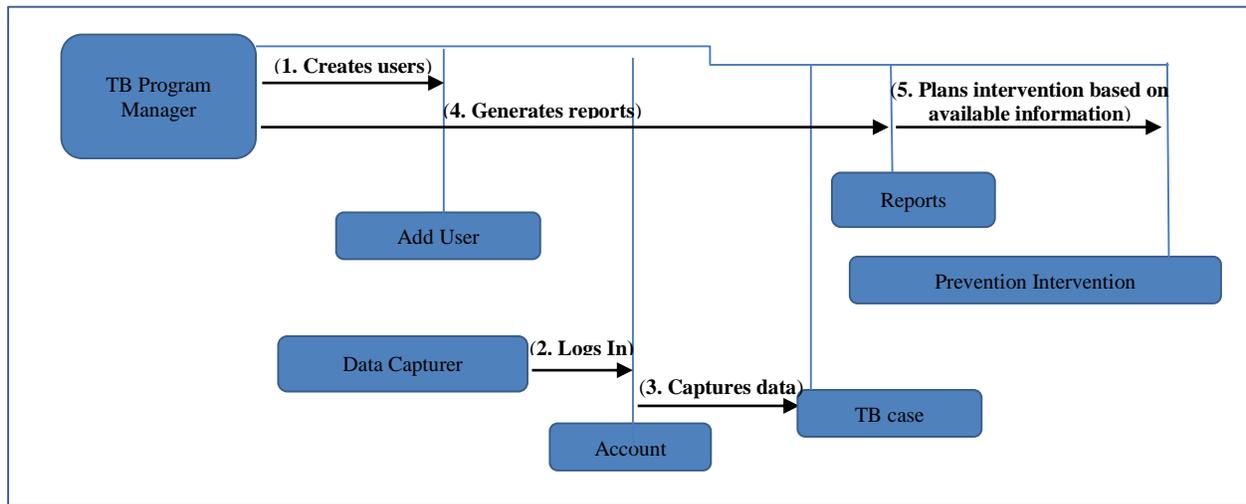


Fig. 9. System Sequence Diagram.

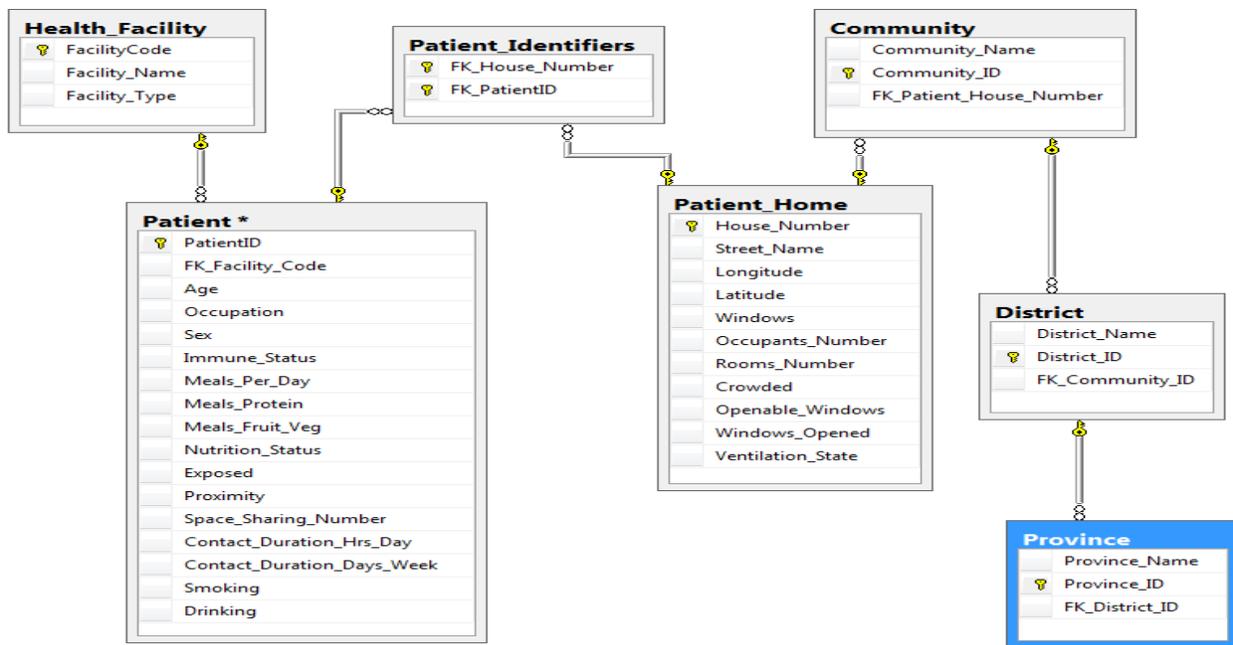


Fig. 10. ERD of the System Prototype.

The relationships depicted in the ERD can be summarized as;

- A Health Facility offers diagnostic and treatment services to Patients.
- Patients live in a Home.
- A Home is located in particular Community.
- A Community is within a District.
- A District is in a given Province.

V. RESULTS

The first section looks at the results as the output of the research activity. It starts by highlighting the results of the baseline study using descriptive statistics and regression

analysis. The analysis was done using SPSS and results have been disaggregated by some of selected TB transmission factors identified in the literature review and how likely they are to escalate the spread of TB in Zambia, Lusaka district in particular. The second section shows some modules of the developed prototype system and also highlights the requirements needed for the prototype system to function.

A. Baseline Study Results

The baseline study had 245 participants drawn from five TB diagnostic and treatment health centers around Lusaka district. Table II shows the Frequency Table of distribution of participants across the five health facilities. The results of the survey show that 82% of participants were aged between 20 and 50 years old as shown in Table III while 65% of participants were male as shown in Table IV which displays the disaggregation of participants by sex.

TABLE. II. PARTICIPANTS' FREQUENCY TABLE

Number	Participants distribution			
	Facility	Frequency	Percent	Cumulative Percent
1	Chawama Level One Hospital	31	12.7	12.7
2	Chipata Level One Hospital	59	24.1	36.7
3	Kalingalinga Clinic	25	10.2	46.9
4	Kanyama Level One Hospital	81	33.1	80
5	Mtendere Clinic	49	20	100
	Total	245	100	

TABLE. III. PARTICIPANTS BY AGE GROUP

Group	Participants age groups			
	Age group	Frequency	Percent	Cumulative Percent
1	<20	9.4	9.4	9.4
2	20-29	85	34.7	44.1
3	30-39	77	31.4	75.5
4	40-49	40	16.3	91.8
5	50-59	13	5.3	97.1
6	60-69	4	1.6	98.8
70+	3	1.2	100	100
	Total	245	100	

TABLE. IV. PARTICIPANTS BY SEX

Sex ID	Participants by sex			
	Sex	Frequency	Percent	Cumulative Percent
1	F	85	34.7	34.7
2	M	160	65.3	100
	Total	245	100	

a) *Exposure Due to Proximity*: The only way TB can be transmitted from the infected person to the one not infected is by exposure to the TB causing bacteria and this exposure is mainly due to being with someone who has active TB. Out of the 245 participants surveyed, 134 remembered being and sharing space with someone who was either on TB treatment or presented with the TB symptoms such as prolonged coughing, fever, chest pains or was unintentionally losing weight during the last one year prior to being diagnosed with TB. The bar graph in Fig. 11 shows that 55% of the respondents could have contacted TB due to exposure to the TB causing bacteria as a result of being in the same space and spending time with someone who had TB.

Linear regression analysis results in Table V show the P Value of $p < .001$, implying a significant relationship between exposure to the TB causing bacteria and TB cases. The coefficient of 23 and 111 being the constant of the linear regression equation means that TB cases will occur if individuals are exposed to the TB causing bacteria.

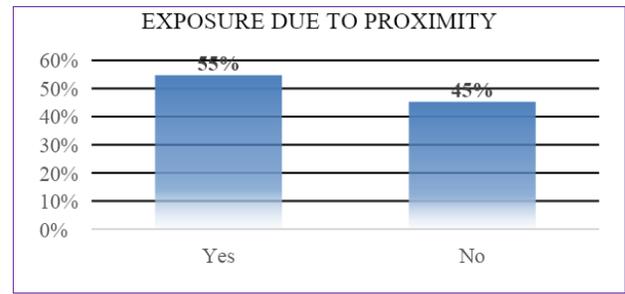


Fig. 11. Exposure Due to Proximity Graph.

b) *Ventilation*: While a number of participants were exposed to the TB bacteria due to proximity, the problem of TB spreading could have been compounded by poor ventilation with 165 (which is 67%) of the 245 participants' homes which were surveyed having poor ventilation.

Linear regression analysis results show a perfect negative coefficient correlation, -1.000 between ventilation and TB cases. The linear regression equation has -85 as coefficient and 165 being the constant, implying that TB cases are likely to occur as long as ventilation is poor. Table VI shows the regression analysis results of the baseline survey on ventilation.

c) *Crowding*: The results of the baseline study showed that crowding was eminent with 202 of the surveyed participants' houses falling into the crowded category, translating into 82%. Only 43 out of the 245 houses were not crowded. With exposure to the TB causing bacteria and poor ventilation being common, crowding is likely to escalate the spread of TB.

The Linear regression analysis results show the significance of $.000$ with coefficient of 159 and 43 being the constant of the linear regression equation and 1.000 as the correlation coefficient, implying a perfect positive correlation between crowding and the spreading of TB. The bar graph in Fig. 12 and the linear regression analysis results in Table VII summarizes the baseline survey results on crowding.

TABLE. V. EXPOSURE VS TB CASES REGRESSION ANALYSIS COEFFICIENTS OUTPUT

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	111	0		1931292773.353	.000
Exposure	23	0	1	-233920361.389	.000

a. Dependent Variable: TB Cases

TABLE. VI. VENTILATION VS TB CASES REGRESSION ANALYSIS COEFFICIENTS OUTPUT

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	165	.000		1170045913	.000
Ventilation	-85	.000	-1.000	-427078195	.000

a. Dependent Variable: TB Cases

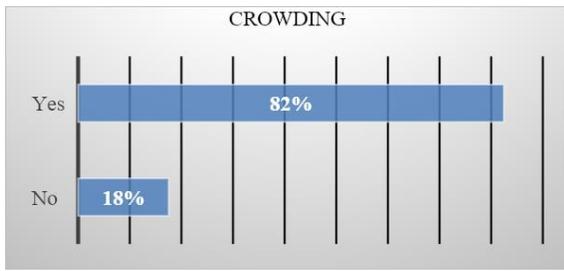


Fig. 12. Results of the Crowding Survey.

TABLE. VII. CROWDING VS TB CASES REGRESSION ANALYSIS COEFFICIENTS OUTPUT

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	43	.000		231471142.26	.000
Crowding	159	.000	1.000	603979776.00	.000

a. Dependent Variable: TB Cases

d) *Nutrition*: Good nutrition is a deterrent to the development of active TB. When exposed to the TB causing bacteria, a person with a bad nutrition status is highly likely to develop active TB compared to one who has a good nutrition status. The results showed that 92 (38%) participant's nutrition status was good while 153 (62%) participants had a bad nutrition status.

Linear regression results show the significance of .000, Coefficient of -61 and 153 as the Constant, implying that TB cases are likely to occur when exposed individuals have a bad nutrition status. The bar graph in Fig. 13 and the linear regression results in Table VIII summarizes the nutrition status of the participants.

e) *Smoking*: The participants who were smoking prior to being diagnosed with TB made up 44% which was 108 out of the 245 and those who never smoked were 137 which translates into 56% of the participants. The pie chart in Fig. 14 shows the comparison between the smoking and non-smoking participants.

f) *Occupation*: The study also sought to disaggregate the participants by occupation. The results reported 34 different occupation types for the 245 participants. There were 55 small scale traders who were selling in small shops and markets, 18 drivers who included bus, taxi or truck drivers, 9 bus conductors, 6 restaurant workers who worked as waiters or waitresses, 21 were school pupils, 18 who did general work and 52 were unemployed. Table IX shows selected occupation types of the participants.

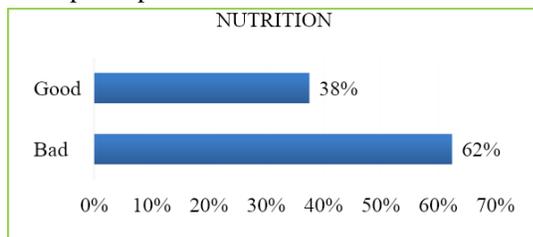


Fig. 13. Nutrition Status 12 Months Prior to TB Diagnosis.

TABLE. VIII. SMOKING VS TB CASES REGRESSION ANALYSIS COEFFICIENTS OUTPUT

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	153	.000		.	.
Nutrition	-61	.000	-1.000	.	.

a. Dependent Variable: TB Cases

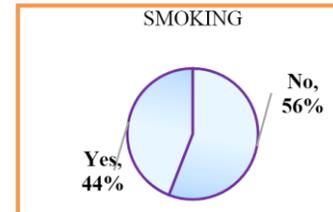


Fig. 14. Participants who Smoked Prior to TB Diagnosis.

TABLE. IX. SOME OCCUPATION TYPES OF PARTICIPANTS

Occupation	Number
Trader	55
Pupil	21
Bus conductor	9
Bus driver	8
Restaurant	6
Lorry/Truck driver	5
Taxi driver	4

The highlighted occupation types have one thing in common, they expose individuals to a lot of different people daily as they carry out their duties and mainly, in poorly ventilated spaces. This puts them at an increased risk of contracting TB compared to the people who have other occupation types.

B. Prototype System

Based on the proposed business model, the architectural system and the data designs, a prototype system has been developed. It is accessible for data entry on mobile devices (Smartphones, Tablets, Laptops, etc.) via a web browser. The system's front end was developed using PHP, embedded with the Google Mapping API for displaying geo distribution of TB cases in communities on the TB program managers' side. MySQL was used as the RDBMS.

a) *Functional Requirements*: Input: TB case details comprising parameters from which the likely cause of disease as well as geo coordinates of the location where the case was reported from.

System Processes: Data storage, retrieval, aggregation and summarization to produce maps and reports.

Output: Reports/Geo maps

b) *Hardware Requirements*: Basic Smartphones, Tablets or mobile computers can be used for data capturing as long as a web browser has been installed.

For good performance of the web application on the managers' side, a computer with the following hardware specification is recommended:

2.3 GHz processor speed or higher; 2GB of RAM or higher; 1GB of free disk space at least; and Internet connection.

c) *Some Modules of the Prototype System:* The system has two user types, the data capturers and program managers who need to make decisions to prevent the spread of TB. The data capturing module's primary function is to capture TB cases' data in communities and as such it only has a Data Entry button which allows the user to capture and send data to the server, and the Logout button for logging out of the system. Fig 15 on page 12 shows the data capturer's landing screen.

When logged in, the system takes note of the GPS location of the device and captures the geo coordinates of that position

and shows the position of the device on the Google Mapping Application. This is for the purpose of recording the geo coordinates of the location of the TB case.

When the user clicks on the Add Entry button, the data capturing page loads onto the device. This is the page onto which the parameters of the TB case are recorded. These parameters are patient attributes and other factors which could be used to establish what could have caused the TB case in question and could be used when planning and implementing disease control and prevention activities.

It's important to note that the Location Access permission on the device being used for data capturing has to be On as shown in Fig. 16 when capturing data. This enables the application to access and pick the GPS location of the patient's place of residence.

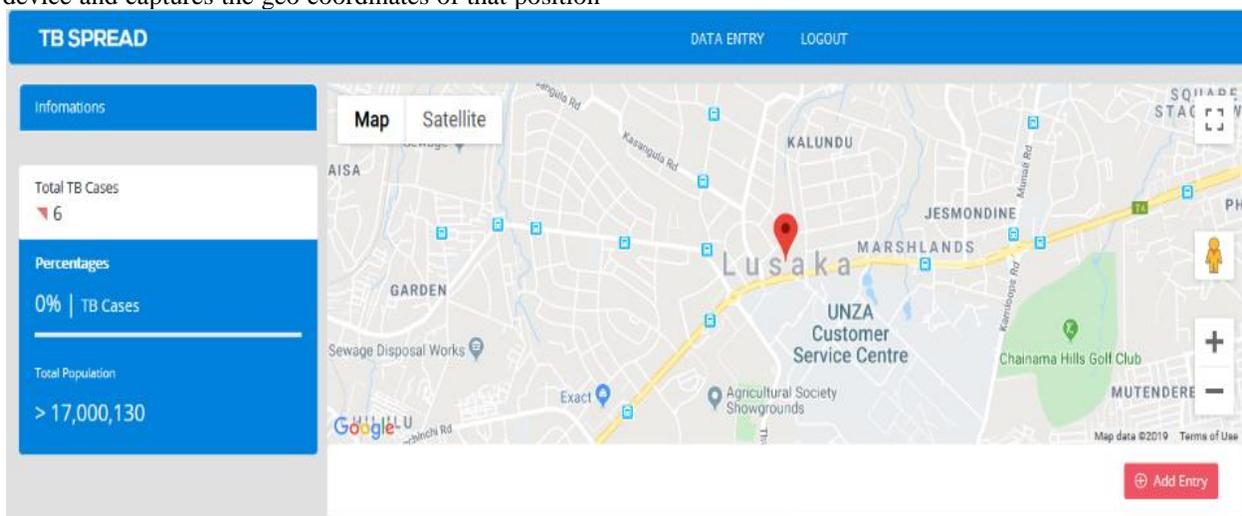


Fig. 15. Data Capturer Landing Page.

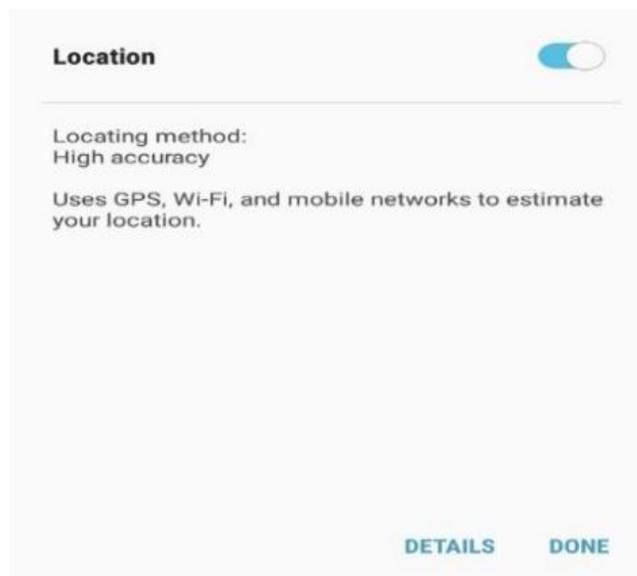


Fig. 16. Location Access Permission of the Data Capturing Device.

6. What is your occupation (before you started TB treatment)

Occupational Risk

7. Were you smoking before you started TB treatment?

Yes No

8. Were you drinking alcohol before you started TB treatment

Yes No

GPS Coordinates

Latitude: -15.396864
Longitude: 28.3336704

Post Data

Fig. 17. Automatically Captured GPS Coordinates of the Data Entry Location.

The data capturer enters all the parameters of the TB case being reported except for the spatial data of the patient's location which is automatically picked based on the geo location of the device, which is the spatial location the TB patient. Fig. 17 is the screen shot of the data capturing page with automatically captured GPS coordinates of the location from which data is being captured.

To know the geographical distribution of TB cases for decision-making about the disease control and prevention activities which need to be carried out in a given area, TB program managers can view maps. The maps show the geographical location of TB cases, and these are likely targets when planning and implementing TB control and prevention activities. Fig. 18 shows the screen shot of some TB cases geographical distribution on a map

Additional information is availed to TB program managers through reports. The reports are accessible through the Reports button on the home page. The information on the reports can be used to craft interventions which can address the prominent causes of the disease in a given area. For example, if most patients were smoking prior to being diagnosed with TB, non-smoking messages can be prioritized

when planning and implementing control and prevention activities. If the reports show that most patients were drinking prior to being diagnosed with TB, prevention activities must focus on drinking places as these could be breeding spaces for TB transmission. The screenshot in Fig. 19 on page 13 shows the draft report of the system prototype with summary patient parameters.



Fig. 18. Satellite Image Screen Shot of the Geographical TB Cases Distribution.

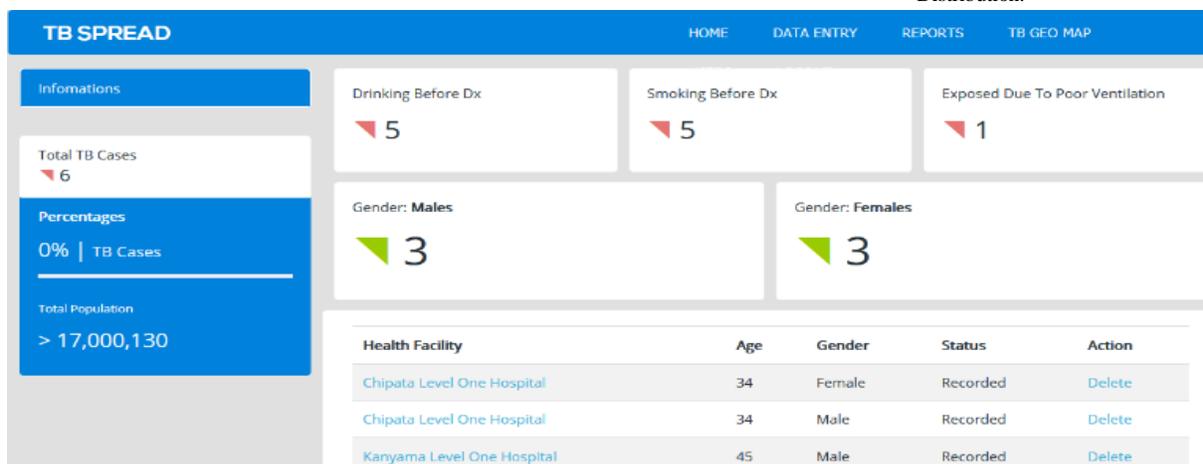


Fig. 19. Draft Report with Summary Patient Parameters.

VI. DISCUSSION

The results of the data analysis show that crowding is the major driver of the TB transmission in Lusaka. While crowding is eminent, 67% (165 out of 245) of the participants' homes had poorly ventilated spaces which they shared with the people who were on TB treatment or presented with symptoms of the TB disease such as coughing and unexplained weight loss. From these findings, it can be established that crowding and poor ventilation resulted into prolonged exposure to the TB causing bacteria by individuals who contracted the disease. Exposure in the face of a weakened immune system and bad nutrition is likely to escalate the spread of TB. To reduce the transmission rates of the disease, prolonged exposure to the TB causing bacteria should be minimized and this can be achieved by improving ventilation systems of homes and other spaces where people are found. This can be achieved by having more openable windows installed in building such as houses, offices, bars, churches, schools and other public places and the windows should always be open whenever there are people. The baseline survey results show that crowding is a common and is one of the drivers of TB transmission and with the established poor ventilation patterns, TB is likely to keep spreading. To minimize the spreading of the disease, crowding must reduce. Another way of reducing the rate at which the disease develops, individuals must ensure they eat food with good nutrition value such as fruits, vegetables and food containing proteins as a good nutrition status is a deterrent to the development of active TB when an individual is exposed to the TB causing bacteria. Good nutrition status strengthens the immune system thereby reducing the susceptibility of individuals to the TB bacteria.

On the hand being male raises the likelihood of contracting TB to 65% compared to the 35% chances for the females. Analysis of the data showed that the average age of the participants was 33 years. Male participants made up 65% of the study population and 61% of the participants drank alcohol prior to TB diagnosis. These three variables, age, sex (males) and alcohol drinking coupled with a high rate of unemployment could be inferred to be driving factors in the spread of TB as men mostly spend time drinking alcohol in poorly ventilated and crowded drinking places such as taverns and bars. Equipping individuals who frequent such places with awareness knowledge of TB transmission and reducing on the times spent in such spaces can reduce the rate at which the disease is spread.

From the prototype development perspective, it can be established that when properly harnessed, ICTs can be used to capture, process information and provide it in way and formats which are easy for decision maker to make decisions.

Though there could be other factors which promote the spread of TB, TB program managers can put in interventions which can help to reduce the rate at which the disease is spread if they have this information and other reports readily availed by the system. Coupling this information with the geospatial display of TB cases on a mapping application can help TB program managers to put resources where it matters the most in order to fight TB.

VII. CONCLUSIONS AND RECOMMENDATIONS

A. Conclusion And Recommendation

This paper presents a way of making well informed decisions in the face of constrained resources to help with the fight against TB. It utilizes the flexibility of cloud computing, the accuracy and reliability of geo spatial analysis and display of information on a mapping application made feasible by GIS and the ability to create reports and help to make decisions via the web application. It is recommended that information is always used as a cornerstone for decision-making in TB prevention and this paper makes this recommendation a reality.

B. Future Works

While this paper focused on the use of ICTs to help TB program managers make decisions based on the information availed to them by the system, it is desirable in future, to have the system it self suggest the best intervention activities to be carried out in a given community to the TB program managers. Having the system recommend the intervention packages to be implemented will likely result in more efficient and impactful interventions because the system will be making decisions and not humans, thereby reducing the inherent human biasness in the control and prevention activities planned and implemented.

The Reports module need to be enhanced and expanded to cover more parameters upon which TB program managers can base decisions when planning and implementing disease control and prevention activities. It also needs to be visually improved to quicken decision-making.

ACKNOWLEDGEMENT

We are very grateful to everyone who contributed towards the development of this paper, especially to the supervisors for the guidance, knowledge and encouragements, and to the National TB Program Manager for the encouragement and the support rendered when we were seeking permission from the Ministry of Health to conduct the baseline survey.

REFERENCES

- [1] World Health Organization, Global tuberculosis report 2019, World Health Organization, Geneva, Switzerland, 2019.
- [2] P.J. Dodd, C. M. Yuen, C. Sismanidis, J. A. Seddon, H.E. Jenkins, "The global burden of tuberculosis mortality in children: a mathematical modelling study," *Lancet Glob Health* Vol 5 September 2017, pp. e898–906.
- [3] G. Churchyard. et al, 2017, "What we know about tuberculosis transmission: an overview," *The Journal of Infectious Diseases*, JID 2017;216 (Suppl 6), pp. S629-s635.
- [4] World Health Organization, "Global tuberculosis report 2018", World Health Organization, Geneva, Switzerland, 2018.
- [5] Government of the Republic of Zambia - Ministry of Health, National Strategic Plan (NSP) for Tuberculosis Prevention, Care and Control (2017-2021) - Towards Elimination, Lusaka, Zambia, 2016.
- [6] Mutembo et al, 2019, "Urban-rural disparities in treatment outcomes among recurrent TB cases in Southern Province, Zambia," *BMC Infectious Diseases* (2019) 19:1087.
- [7] UNAIDS, UNAIDS DATA 2018, Joint United Nations Programme on HIV/AIDS, Geneva, Switzerland, 2018.
- [8] I. Mwila and J. Phiri, "Geospatial, cloud and web based model for evidence-based decision-making in tuberculosis prevention," in

- International Conference in ICT (ICICT 2019) – Lusaka, Zambia (20th – 21st November 2019), ISBN 978-9982-70-985-9, Page 110 – 115.
- [9] A. Ali. 2011. “Exploring the aspects of digital divide in a developing country,” *Issues in Informing Science and Information Technology* Volume 8, 2011.
- [10] A. Joshi, et al, 2013, “The role of information and communication technology in community outreach, academic and research collaboration, and education and support services (IT-CARES),” *Perspectives in Health Information Management*, Fall 2013.
- [11] J. Doosje, A. Bosman, M. L. A. Heijnen, V. van Kuyvenhoven, A. A. Warris and J. van Steenberg, 2004, “Improving the use of information and communication technology in infectious disease control,” *European Journal Of Public Health* Vol. 14 2004 NO. 4.
- [12] A. Rukshan, A. Miroshan, “An ICT-based real-time surveillance system for controlling dengue epidemic in sri lanka,” unpublished.
- [13] World Health Organization. *Digital Health For The End Tb Strategy: An Agenda For Action*, Geneva, Switzerland, 2015.
- [14] W. Curioso, et al, *Information and communication technologies for prevention and control of HIV infection and other STIs*, Washington, United States, 2017.
- [15] B. C. K. Choi, 2012, “The past, the present and the future of public health surveillance,” *Scientifica*, Volume 2012, Article ID 875253.
- [16] P. Narasimhan, J. Wood, C. R. MacIntyre and D. Matai, 2013. “Risk factors for Tuberculosis,” *Pulmonary Medicine* Volume 2013, Article ID 828939.
- [17] B. Mathema, 2017, “Drivers of Tuberculosis Transmission,” *The Journal of Infectious Diseases*, JID 2017:216 (Suppl 6), pp. S645-s653.
- [18] P. Wargocki, 2013, “The Effects of Ventilation in Homes on Health,” *International Journal of Ventilation* ISSN 1473-3315 Volume 12 No 2 September 2013.
- [19] M. Ruiz-Castell, et al, 2015. “Household crowding and food insecurity among inuit families with school-aged children in the Canadian Arctic,” *American Journal of Public Health* March 2015, Vol 105, No. 3.
- [20] S. Kevin, R. L. Blake. K. A. Simic, *Measuring Overcrowding in Housing Prepared for: U.S. Department of Housing and Urban Development Office of Policy Development and Research*, Econometrica, Inc. Bethesda, Maryland, United States, 2007.
- [21] J. Doosje, et al, 2004. “Improving the use of information and communication technology in infectious disease control,” *European Journal Of Public Health* Vol. 14 2004 No. 4.
- [22] United States Centre for Disease Control and Prevention (CDC), *Know Your Epidemic: Disease Surveillance For Targeting Efforts To End TB*, 2015.
- [23] N. Sabi'u, S. N. Muhammed, N. Zakari and M. S. Khalil, 2015, “Vector data model in GIS and how it underpins a range of widely used spatial analysis techniques,” *Dutse Journal of Pure and Applied Sciences* 1(1) June 2015 pp. 122 - 132.
- [24] Rezaeian, et al, 2007. “Geographical epidemiology, spatial analysis and geographical information systems: a multidisciplinary glossary,” *J Epidemiol Community Health* 2007; 61:98–102.
- [25] R. Zerbe, et al, *An Analysis of Benefits from use of Geographic Information Systems By King County Washington*, Richard Zerbe and Associates, Seattle WA, 2012.
- [26] K. Soniya, A. S. Kumar. “A basic study on cloud computing,” *IOSR Journal of Computer Engineering (IOSR-JCE)* e-ISSN: 2278-0661, p-ISSN: 2278-8727, pp 43-46.
- [27] S. Chowdhary and A. Srivastava, 2013, “Cloud computing: a key to effective & efficient disease surveillance system,” in *Proc. of Int. Conf. on Advances in Computer Science and Application* 2013.
- [28] Veer Surendra Sai University of Technology. *Internet & Web Technology-Ii Lecture Notes*. Department of Computer Science & Engineering, IT.
- [29] G. Pullaiah College of Engineering and Technology. *Lecture Notes on Web and Internet Technologies*. Department of Computer Science and Engineering.
- [30] J. Choi, et al, 2016, “Web-based infectious disease surveillance systems and public health perspectives: a systematic review,” *BMC Public Health* (2016) 16:1238.
- [31] J. J. Hamilton and R. S. Hopkins, *The CDC Field Epidemiology Manual - Using Technologies for Data Collection and Management*, 2018.
- [32] R. Islam, et al, 2010, “Mobile application and its global impact,” *International Journal of Engineering & Technology IJET-IJENS* Vol: 10 No: 06.
- [33] Y. P. Li, et al, 2013, “Decision support system for the response to infectious disease emergencies based on webgis and mobile services in China,” *PLoS ONE* 8(1): e54842. doi:10.1371/journal.pone.0054842.
- [34] E. Almrot and S. Andersson, 2013, “A study of the advantages & disadvantages of mobile cloud computing versus native environment,” unpublished.
- [35] R. Chunara, 2012, “New technologies for reporting real-time emergent infections,” *Parasitology*, 2012 December; 139(14): 1843–1851. Doi: 10.1017/S0031182012000923.
- [36] N. Beyers, et al, 1996, “The use of a geographical information system (GIS) to evaluate the distribution of tuberculosis in a high incidence community,” *Oepariments of Paediatrics and Child Health and Medical Biochemistry, and Institute for Geographical Analysis*, Volume 86 No.1 January 1996 SAMJ.
- [37] N. Tiwari, et al, 2006, “Investigation of geo-spatial hotspots for the occurrence of tuberculosis in Almora district, India, using GIS and spatial scan statistic,” *International Journal of Health Geographics* 2006, 5:33 doi: 10.1186/1476-072X-5-33.
- [38] B.L. Gleason1, A. Kamara, N. Clemens and D. Kargbo, 2018, “Establishment of an electronic integrated disease surveillance and response system in Sierra Leone,” *International Journal of Infectious Diseases* 79(S1) (2019) 1–150.
- [39] S. Biswas, Anisuzzaman, T. Akhter, M. S. Kaiser and S. A. Mamun, 2016, “Cloud based healthcare application architecture and electronic medical record mining: an integrated approach to improve healthcare system,” *Institute of Information Technology Jahangirnagar University, Dhaka, Bangladesh*.
- [40] M. O. Lwin, et al, 2017, “Flumob: enabling surveillance of acute respiratory infections in health-care workers via mobile phones,” *Front Public Health* 5:49. doi: 10.3389/fpubh.2017.00049.