Quest Journals
Journal of Software Engineering and Simulation
Volume 7 ~ Issue 7 (2021) pp: 01-09
ISSN(Online) :2321-3795 ISSN (Print):2321-3809
www.questjournals.org

Research Paper

A Time Series Geo-Analytic Model for the Forecast of Air Borne Communicable Diseases

Amoo, A.O., Oyegoke T. O., Awoyelu I. O., and Adagunodo E.R.
[1](Deparmentt. of Computer Science & Engineering Obafemi Awolowo University, Ile-Ife, Nigeria)
[2](Deparmentt. of Computer Science & Engineering Obafemi Awolowo University, Ile-Ife, Nigeria)
[3](Deparmentt. of Computer Science & Engineering Obafemi Awolowo University, Ile-Ife, Nigeria)
[4](Deparmentt. of Computer Science & Engineering Obafemi Awolowo University, Ile-Ife, Nigeria)

ABSTRACT: *Air borne communicable disease can be serious and life threatening. Hence, availability of models predicting its potential outbreak can be very useful in its controlling and eradication strategies. A time series geo-analytic model for monitoring and forecasting Tuberculosis diseases spread was formulated in this paper using Auto-Regressive Integrated Moving Average (ARIMA) model as the forecasting component and a spectral density function as the spread component. This study also presents a real time monitoring framework which guides the development of a surveillance system for air borne communicable diseases which consists of three main modules. The Data collection module for collection of surveillance data from various sources; an analysis module which includes the geo analytic model for forecasting of future cases and a spectral density function for providing the spatial distribution of the spread and an incident notification module for presenting the stakeholders with the results of the analysis of the communicable diseases data collected from reporting centers. Future works is focused on the development of a real time surveillance system for air borne communicable diseases based on the proposed framework*

KEYWORDS: *Communicable disease, disease forecasting, tuberculosis, time series modeling*

Received 07 July, 2021; Revised: 19 July, 2021; Accepted 21 July, 2021
Published with open access at www.questjournals.org

I. INTRODUCTION

In most developing countries, communicable diseases are the most common causes of illness, disability and death. Typical examples of these infectious diseases include Cholera, Ebola Virus, Hepatitis, Influenza Measles, Severe Acute Respiratory Syndrome (SARS), Malaria, Tuberculosis (TB), Typhoid Fever, Diarrhea, Human Immunodeficiency Virus/Acquired Immuno Deficiency Syndrome (HIV/AIDS) and Acute Respiratory Infection (ARIs). The tremendous diversity of microbes in addition to their ability to develop and adjust to changing populations, environments, practices, and technologies has posed ongoing threats to human health and continually challenging the ability to prevent and control these diseases [1]. Communicable disease is classified as the second leading cause of death in humans by the World Health Organization (WHO) with approximately 15 million deaths worldwide every year [2]. Among these communicable diseases HIV/AIDS, TB, and Malaria have been nicknamed the 'big three' because of their important impact on global human health [3].

TB is a severe communicable disease that affects the respiratory part of the human body and is caused by the pathogen Mycobacterium tuberculosis and spreads through air droplets by sneezing and coughing by the infected person [3]. It is a disease of poverty affecting mainly young adults in their most productive years. The disease, if not given an urgent attention can be a threat to life. It has been one of the biggest health challenges worldwide and it is the second major cause of death, especially in developing countries. There was an estimation of about 14 million people worldwide that are infected with active tuberculosis [4]. A total of 1.4 million people died from TB in 2019 (including 208 000 people with HIV). Worldwide, TB is one of the top 10 causes of death and the leading cause from a single infectious agent [5]. With so many efforts and strategy puts in place to combat TB, it still remains a major public health issue with a high global health burden. WHO reports about seven countries accounting for 64% of the burden of TB comprising of India, Indonesia, China, Philippines, Pakistan, Nigeria and South [6].

TB happens to be a new world's most communicable diseases that kill with about 4500 lives being lost per day, and unfortunately, Nigeria is far worse hit by this global epidemic in Africa. Nigeria currently ranks 7th

in the world and 2nd in Africa among the 30 countries with the highest burden of TB, TB/HIV, and multi-drug resistant TB [7].

Tuberculosis is a major public health problem in Nigeria with an estimated prevalence of 616 cases per 100,000. Nigeria ranks first in Africa, and fourth among the 22 high TB burden countries in the world, and no fewer than 460,000 cases of TB are reported annually in Nigeria. The country is considered by WHO as one of the high burden countries and TB incidence continue to be very high. Therefore there is need to continue monitoring and predicting TB incidence in an effort to make control of TB more effective.

The pillar of the fight against communicable diseases has always monitoring systems for monitoring diseases, pathogens and clinical outcomes. Monitoring and predicting of emerging and recurring infections is particularly important for controlling the spread of communicable diseases among populations [8]. Surveillance of communicable diseases is recognized as a cornerstone of decision-making and public health practice. It is classified as the continuous systematic collection, analysis, interpretation and dissemination of health data for the planning, implementation and evaluation of public health action [9], [10] and [11]. Regular surveillance of diseases is an important tool for organizations that allow timely detection and response to outbreaks [12].

The most important aspects of epidemics is surveillance, early identification of possible outbreaks and patterns useful for controlling disease spread. The first attempt was made by John Snow while modeling cholera epidemic in London. By compiling spatio-temporal data about incidence which were visualized on a map, it led to the identification of the source of the disease as a nearby water pump which helped eradicate cholera in London [13]. By the 20th century, there were already global initiatives on the way for combating epidemic diseases which required effective mechanisms for internal action and readiness around the world.

Knowing future health problems through review of temporal changes and prediction of diseases can play an important role. This ensures that there is continuity in controlling and intervention programs and optimal allocation of resources. Different mathematical and statistical models have been used in different studies to predict tuberculosis incidence and to study its temporal changes and based on the nature of the data and assessment, a particular model was proposed in every study. Some popular methods currently used in prediction for communicable disease morbidity had been linear regression method, gray model method, artificial neural network method, specifically the autoregressive integrated moving average (ARIMA). To make forecast, ARIMA method is applied more widely than other methods, it can take into account periodic changes, changing trends and random disturbances in time series, and it is very useful in modeling the temporal dependence structure of a time series.

Time series is a dynamic field of research which is adopted in order to carefully collect previous observations of time-based data and carefully studying the data in order to formulate an appropriate model that describes the inherent structure of the series. The time series model fitted to the data can then be used to generate future unseen values for the series also called forecasting. Therefore the art of predicting time series can be considered as a process of predicting the future based on an understanding of the past [14]. As a result of the growing importance of Time series in the research community, such as business, economics, finance, science and technology, etc., adequate models for the underlying time series must be properly taken into account. It is clear that a successful forecast of time series depends on the selection of a model with a good fit [15].

The main objective of this paper is to present a time-series based geo-analytic model which will be achieved by using the ARIMA model as the forecasting components for TB infection and thereby extending the use of ARIMA model by introducing a Spectral density function that takes care of the spread of the diseases, thereby providing a means of analyzing the spread of communicable diseases to neighboring locations from a point (or location) of the occurrence (or outbreak) of communicable diseases using time series modeling. The model will also provide a means of using location-based information in combination with time to estimate the spread of communicable diseases to an exposed location from a point of reference. This will provide a means of adopting evacuation and control measures towards reducing the likely spread of communicable diseases to vulnerable neighbouring locations. This geo analytic model will be integrated into the real time monitoring framework which guides the development of a surveillance system for air borne communicable diseases. Following the review of related works on the subject of disease monitoring and surveillance, a number of related papers were observed. Among them included the use of time series to the estimation of future values of diseases under surveillance such as ARIMA and ARIMAX alongside other statistical techniques such as: Bayesian and (Hidden) Markovian models. They review are presented in the following paragraphs.

[16], worked on the time series modelling of Influenza transmissions. The study used autocorrelation and partial autocorrelation plots to determine the appropriate autoregressive integrated moving average (ARIMA) model for influenza transmission. Based on the available data of suspected influenza cases and climate variables, the most appropriate ARIMA(X) model for each region was obtained. The study revealed that the average temperature correlated with influenza cases in both central and southern regions of the study, but average minimum relative humidity played an important role only in the southern region.

[17] Worked on the development of a threshold monitor for TB surveillance data. The weekly counts of newly diagnosed patients with sputum smear-positive pulmonary TB from 2005 to 2011 nationwide was

extracted. Serfling and Hidden Markov Model were applied in presence/absence of linear, seasonal and autoregressive components to detect unexpected incidences of the disease. The models were subsequently evaluated in terms of goodness of fit, and compared in detection of the disease phases. Findings from both adjusted R-square (R2) and Bayesian information criterion (BIC) presented a higher goodness of fit for periodic autoregressive HMM (BIC= -1323.6; R2 = 0.74) than other models.

[18], finds the best model which fits onto the TB occurrence data of Ashanti Region of Ghana. The TB occurrence data spanning a period of January 2001 to March 2013 was obtained from the Ashanti health services. An ARIMA (1, 0) was the best model that fits the data. The mean absolute error (MAE) and the mean squared error (MSE) were used to compare the in- sample forecasting performance of the three selected Models. The authors were able to achieve a forecasting accuracy of 16.3171 and 461.3148 respectively for the MAE & MSE. It was evident from the analysis that TB occurrence in the region can be modelled on AR (1).

[19] developed a five year forecasting model for Malaria incidence among children under the age of five in the Edum Banso Sub-District of Ghana. A secondary monthly malaria incidence data spanning the period of 2008-2013 was used for the model formulation. The Box Jekins ARIMA methodology was employed to obtain the best fitting model by comparing the normalized BIC, MAE and stationary R square values. Comparative analysis showed that the ARIMA (1, 1, 2) had the best performance. The five year forecast showed a linear trend angled diagonally up, which explain that Malaria cases in children under five years will increase steadily over the period of five years.

[20], proposed a time series analysis of TB data from year 2005 to 2017 in China. Monthly TB data from year 2005 to 2017 was used for the analysis. A time series model based on SARIMA and a hybrid model of SARIMA – generalized regression Neural Network (GRNN) model was adopted. The result of the analysis showed that there was a decreasing trend of 3.17% per year and a seasonal variation in the data from 2005 to 2016. The hybrid SARIMA model was found to be more effective than the widely used SARIMA with a mean error rate of 2.56% and 6.07% and determination coefficient of 0.94 and 0.73 respectively. The better performance SARIMA – GRNN model was further confirmed with the forecasting dataset of 2017.

II. FORMULATION OF GEO-ANALYTIC MODEL FOR DISEASE SPREAD

In developing the time series geo analytic model, two different models were proposed. The first model was required for forecasting the future reported cases about TB diseases reported at a given location (also called the point of outbreak) using the auto-regressive integrated moving average (ARIMA) model. Using this model, the future data about TB cases that is expected to be reported a number of periods in the future can be determined.

The second model called the spectral density function was required for estimating the level of spread of TB disease from a known location of disease outbreak to another location some distance away using a probability distribution function. This model was used to estimate the likely spread of air-borne disease to another location a certain distance away and can be used to access the level of exposure of people within the suspected location to the risk TB disease spread. Before formulating the geo-analytic model for this study, stakeholder at different levels of health care service delivery involved with Disease Surveillance and notification were interviewed to ascertain the relevant variables needed for monitoring the spread of infectious diseases.

2.1 Formulation of ARIMA time series model for TB disease forecast

For the purpose of this study, an auto-regressive integrated moving average (ARFIMA) time series model was adopted for making forecast of future cases from collected cases at a given location. The model is also expressed as the $ARFIMA(p, d, q)$ model for which the values of p is used to identify the number of lags for the auto-regressive (AR) part, q is used to identify the number of lags for the moving average (MA) part and d defines the number of differencing performed in order to achieve stationarity.

The data that was examined for this study consisted of the set of reported observations on Tuberculosis, y_t while the climatic data examined formed the auto-regressive component $x_{t,k}$. The data examined consisted of quarterly reports so that t represents the quarter when a report is made. Using the reported dataset collected, an $ARIMA(p, d, q)$ time series model was estimated of the form defined according to equation (1).

$$y_t = \sum_{i=1}^{p} \varphi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \sum_{k=1}^{v} a_k x_{t,k} + a_0 + \varepsilon_t, \quad (1)$$

Where:

$\sum_{i=1}^{p} \varphi_i y_{t-i}$ - Auto Regressive (AR) part of the ARIMA model of y_t at 1 to p lags

$\sum_{j=1}^{q} \theta_j \varepsilon_{t-j}$ - Moving Average (MA) part of the ARIMA model of ε_t at 1 to q lags

ε_t - Error term

$\sum_{k=1}^{v} a_k x_{t,k} + a_0$ - Regression equation of v climatic factors with coefficients a_k

$p \in \square; q \in \square;$ - Positive integer values of the AR and MA orders respectively.
$a_k \epsilon \square;\ \varphi_i \epsilon \square;\ \theta_j \epsilon \square$ - Real number coefficients of the exogenous, AR and MA part.

Using equation (1), the quarterly forecast for *n* quarters was estimated for future cases y_{t+n} for the interval time *t+1* to t+n hence the number of cases returned is for the next *n* quarters after the quarters of the last reported data. The coefficients of the AR part was required for explaining the effect of increasing each time lags cases y_{t-i} $(i \leq p)$ on the present cases y_t. The AR part explains how past observations affect future observations of Tuberculosis data. The coefficients of the MA part was required for explaining the effect of the errors recorded at each time lag e_{t-j} $(j \leq q)$ on the present cases y_t. The MA part explains how past errors affect future observations of Tuberculosis data. Also, the coefficients of the exogenous parts $x_{t,k}$ were required for explaining the effect of increasing each exogenous variable on the present cases y_t. This function was estimated and represented as a function of the observed cases y_t, the observed climatic data x_t and forecasted time t called $f(y_t,\ x_t,\ t)$.

2.2 Formulation of estimated spread of TB disease

For the purpose of this study, a spectral density function based on the survival plot was adopted for estimating the spread of Tuberculosis from one known location of outbreak to another location of interest. The motive is to ensure that given knowledge about the reported cases at a known location, one can estimate the extent of spread at another location which is a distance d away from the point of diseases incidence.

The survival function of the distribution was considered since it presents the decay of information from a point source using (1.00 or 100%) which gradually decreases to 0 as the distance from the center increases. Consider the region shown in figure 1, the point of disease outbreak (or the epicenter) is the location at which the TB diseases is reported from which is identified by the rectangular coordinates (x_0, y_0). It was observed that diseases spreads outwardly around the region of the epicenter as it moves outward (increasing distance *d*).

The disease spread is said to decays slowly over the area to a location (x_i, y_i) it has covered over time *t*. Given a point of disease incidence and reporting defined by its polar coordinates *(latitude, longitude)*, the values of the distance between the 2 coordinates will be estimated according to equation (2). If the values of the coordinates of the originating point (x_0, y_0) is (θ_1, φ_1) and that of the point of interest (x_1, y_1) is (θ_1, φ_1) then the distance *d* between is determined by:

$$d = R((Sin\theta_1 Cos\varphi_1 - Sin\theta_2 Cos\varphi_2)^2 + (Sin\theta_1 Sin\varphi_1 - Sin\theta_2 Sin\varphi_2)^2)^{1/2}, \quad (2)$$

where:
R is the radius of the earth $\cong 6400\ km$

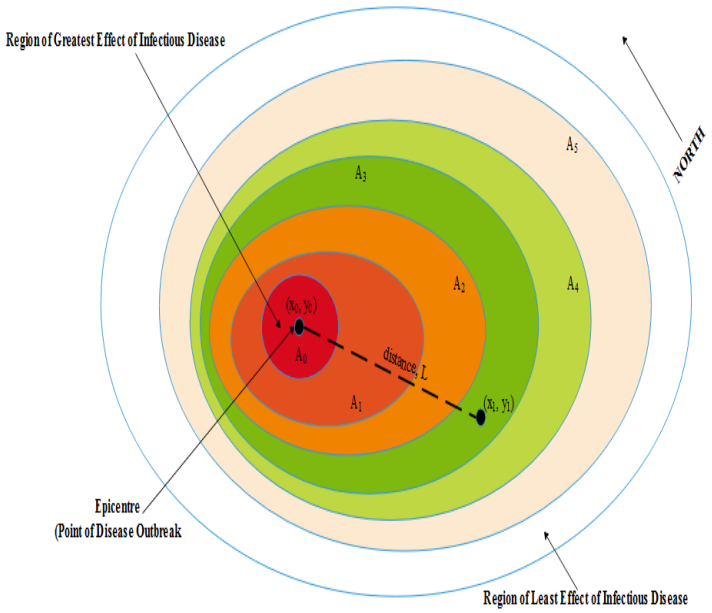


Figure 1: Description of Spread of Disease from Epicenter

Therefore, in order to estimate the spread of the disease from a point of outbreak to a point of interest the values of the forecast made on the *nth* time interval using $f(y_t,\ x_t,\ t)$ alongside the distance *d* are provided to an exponentiated exponential distribution function $S\left(d|f(y_t,\ x_t,\ t)\right)$ as shown in equation (3). The function will be used to estimate the spread as a decay of the last forecast at a known location as distance increases.

$$Spread = S\big(d|f(y_t,\ x_t,\ t)\big) = \Big(1 - \big(1 - e^{-df(y_t,\ x_t,\ t)}\big)^{0.5}\Big) * 100\% \quad (3)$$

2.3 Algorithm for estimating and selecting ARFIMA model of reported data

The algorithm that was adopted for estimating the ARFIMA model is presented in Algorithm I. This was necessary for selecting the best model that fits the dataset. The algorithm required the observed reported cases y_t alongside the exogenous data x_t over a fixed time period. The steps required by the algorithm are presented in Algorithm I

Algorithm I: Estimating ARIMA model from reported Data

INPUT: Time series data consisting of quarterly TB cases and climatic data
Provide time series data y_t alongside the exogenous data x_t
Define the maximum values for lags *p* and *q*

```
1       For (i = 1; i ≤ p)
2           For (j = 1; j ≤ q)
3               Do estimateARIMA(y_t, x_t, i, j);
4                   estimate AIC[ARIMA(i, d, j)];
5               if(AIC[ARFIMA(y_t, x_t, i, j)] = minimum;
6                   Return ARIMA(y_t, x_t, i, j)
7               else next i, next j;
8   Go to 3
9   Until 5
```

OUTPUT: A Fitted ARIMA Model with minimum AIC, error and variance.

The algorithm will load the data following which the maximum values of the orders *p and q* will be stated for which every possible combinations from *p=1 and q=1* to *p=max p* and *q=max q*. The algorithm will then estimates the parameters of the ARIMA model for the different possible combination of the orders generated following which the AIC, log-likelihood and variance will be determined. The ARIMA model with the lowest values for the criteria used will be returned as the appropriate model for fitting the dataset. Conclusions will then be made about the relationships between the AR, MA and exogenous variables on the observed cases y_t based on the returned values of their coefficients.

It is also important to state that since the ARIMA model will fits a model for every type of data loaded, then for every data there is a unique ARIMA model. Therefore, an ARIMA model fitted from a dataset cannot be adopted for use on another dataset from which it was not estimated. Therefore, for every request of forecast, a new ARIMA model will be estimated from the dataset based on the data stored up till the present time of forecast. Finally, the maximum and minimum forecast will then be returned for the forecast of the ARFIMA model based on the values of the mean forecast *yMEAN* and the standard error SE. Using a 95% confidence interval, the maximum and minimum forecasts will be estimated from the yMEAN and SE according to equations (4) and (5) respectively. These values will be used as the output for the forecast $f\big(y_t,\ x_t,\ t\big)$ required for estimating the maximum and minimum spread according to equation (3.3).

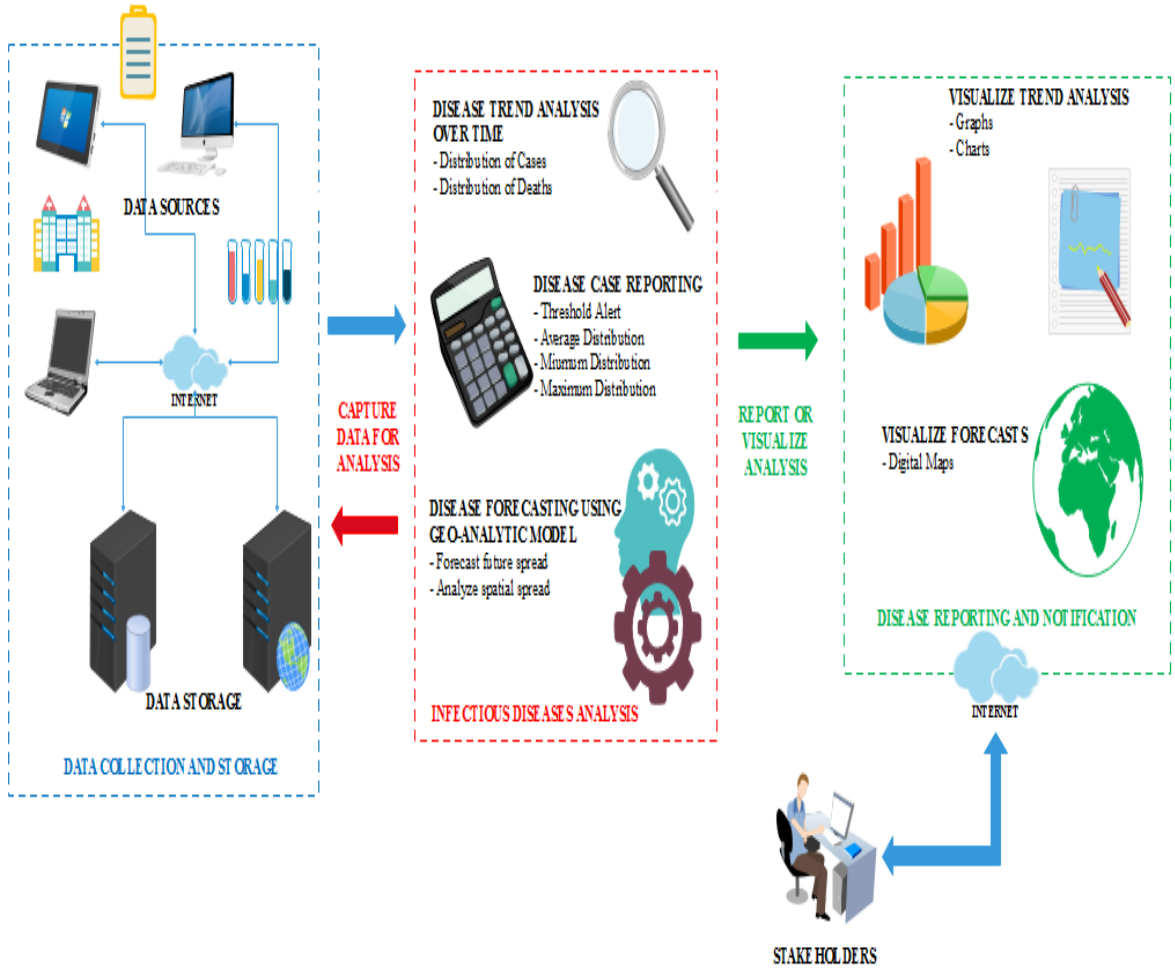
$$Maximum\ Forecast\ = yMEAN + SE \quad (4)$$

$$Minimum\ Forecast = yMEAN - SE \quad (5)$$

II. PROPOSED FRAMEWORK FOR SURVEILLANCE SYSTEM FOR AIR BORNE COMMUNICABLE DISEASE

The proposed prototype real-time disease surveillance system composed of three (3) components, namely: the data collection component, data analysis component and the notification component as recommended by the intending users. Figure 2 shows a description of the framework for the proposed real-time system for monitoring air-borne disease. The data collection component required the collection of reported data from their various sources which may include laboratories, private and public clinics or hospitals consisting of cases and deaths reported alongside other demographic information.

In the data collection component, data is collected from the reporting centers to the system using mobile and web based devices such as laptops, PCs, smartphones and so on. The data collected from each reporting center will be stored in the cloud-base repository. The information will be stored by mapping the geographical-based information which is the coordinate of the reporting centers represented to a point in the ward from which data is reported. This makes it easy for the system to associate data with point location represented on a digital map.



The analysis component is the next component of the framework which required the application of various statistical analytic techniques by stakeholders for the analysis of the disease data collected from the reporting centers. The various analysis to be performed included the distribution of the reported cases and deaths which were organized as the number of reported cases (or deaths) over a fixed time interval (weekly, monthly or yearly) and with respect to the demographic information of the reported individual cases. The beginning and end of the analysis is provided alongside the location for which the required information is related to. Also another part of the analysis component included the geo-analytic model which is required for the forecasting of future cases (or deaths) at a location of disease incidence reports and for estimating the spread at neighbouring locations a distance away from point of incidence.

The notification components will be required for presenting the stakeholders with the results of the analysis of the disease data collected from the reporting centers. The trend analysis of the data distribution of the cases (or deaths) will be presented graphically using bar charts and graphs or alternatively using a frequency distribution table. The results of the forecasting will also be presented using the Google digital Map to present a spatial view of the spread of infectious diseases based on the forecast required for a given time of interest. The results of new disease reports from reporting centers will be provided in real-time so that other stakeholders in neighbouring and related locations can view over their dashboard.

3.1 Data flow diagram (DFD) of disease surveillance system

Data flow diagram (DFD) was used to specify a graphical representation of the flow of data through the proposed disease surveillance system by identifying the flow of data from objects via process which translate into a storage location. It was a way of specifying the structural requirements of the system in terms of the system users by showing the kind of information presented as input and processed into output and where the data is coming from and going to, and to where the data will be stored.

Figure 3 shows a description of the flow of reported data within the proposed system which will start from the creation of users by the system administrator followed by the collection of data by the collation officer located on-site such as the data collation officer, investigators and the stakeholder. The data collation officer collates and submits information regarding the number of cases of diseases based on the observations that were made at

the site of data collation. The data will be stored in the database for future analysis by the investigators or stakeholders for facilitating decision making. The surveillance data about diseases which was stored by the data collation officer also contained location-based information about the data reported from specific locations. This contained details about the physical location of disease incidence alongside information that describes the individual affected such as gender, age group, education and alongside personal and clinical information about the individual affected. The disease surveillance data will be stored for analysis using the geo-analytic model to determine the future forecast of diseases and to estimate the extent of the spread of disease to various surrounding locations from the point of outbreak. The ARFIMA model will be used to estimate the future forecast of disease reports for the next *n* periods in the future which will be collected by the spectral density function for the estimation of the spread of disease to neighbouring locations based on the value of distance recorded. The information stored about reported cases of the disease alongside their demographic variables will be viewed as tables, charts and graphs while the forecast can be presented as a time series plot and the estimated spread displayed on the digital map.

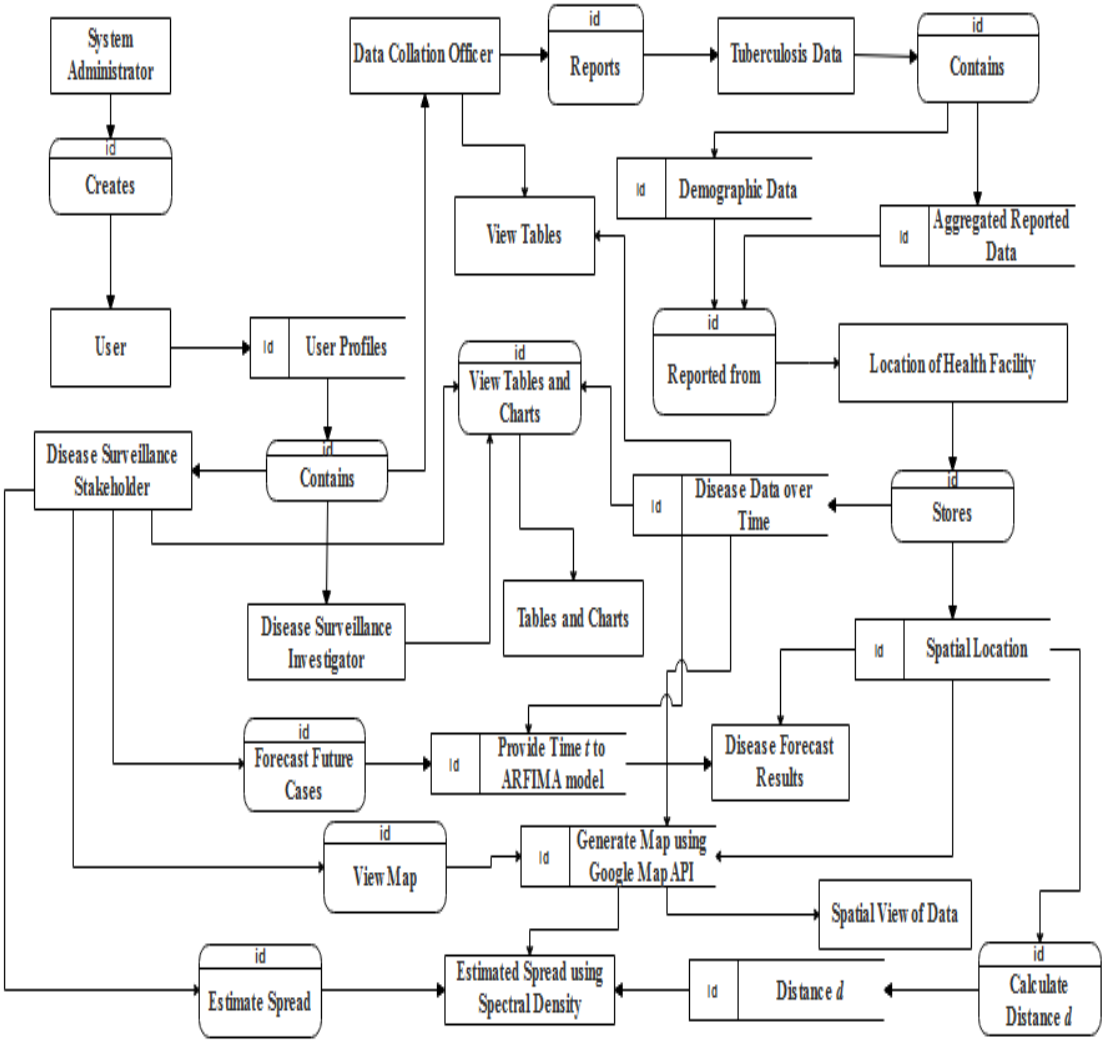


Figure 3 Data Flow Diagram of Information through the Disease Surveillance System

IV. CONCLUSION

This paper has identified that a great number of human lives are been lost to communicable diseases across the globe and the availability of timely reports of outbreaks and epidemics of such diseases in developing nations like those in sub-Saharan is still a serious challenge. Tuberculosis was also discovered to be the new world's most communicable diseases that claim about 4500 lives per day. This has led to the need of the development of a time-series based geo-analytic model that provides a means of analyzing the spread of communicable diseases to neighboring locations from a point (or location) of the occurrence (or outbreak) of communicable diseases alongside for the development of a real time framework which employs the use of cloud services for the storage, analysis and forecast of the spread of communicable diseases. The study will assist in providing a means of using location-based information in combination with time to estimate the spread of

communicable diseases to an exposed location from a point of reference. It will also help in providing a means of adopting evacuation and control measures towards reducing the likely spread of communicable diseases to vulnerable neighbouring locations. The real-time communicable diseases surveillance system which uses a geo-analytic model will provide a means via which communicable diseases data can be captured, stored, analyzed and accessed easily from a single cloud-based repository by a variety of computing devices ranging from PCs to smartphones thereby removing the gap facing accessibility to information by stakeholders. The disease mapping capacity of the system will also allow stakeholders to monitor the distribution of communicable diseases across various reporting locations thereby providing accurate report of the public health status of the nation.

This paper presented a proposed framework that can be used for the development of a real time air borne communicable diseases surveillance system. The framework depends on three components namely: the data collection module, analysis module and incidence notification modules. The Data collection module for collection of surveillance data from various sources; an analysis module which includes the geo analytic model for forecasting of future cases and a spectral density function for providing the spatial distribution of the spread and an incidence notification module for presenting the stakeholders with the results of the analysis of the communicable diseases data collected from reporting centers.

V. FUTURE WORKS

Future works in line with this study are focused at the identification and collection of the most relevant data relevant to the monitoring of air borne communicable diseases spread alongside the development of the time series geo analytic model which uses an Auto-Regressive Integrated Moving Average (ARIMA) model as the forecasting component and a spectral density function as the spread component. Finally, the real time air borne communicable surveillance system will be implemented based on the framework proposed for this study using a combination of Visual C#, HTML, PHP, CSS, PHP, Java and XML for implementing the business logic of desktop devices and mobile devices containing the geo-analytic model developed using R program, the Google Map® API for providing the digital map interface and Google's Firebase® Cloud Services for storing and accessing surveillance data in real-time. This study will contribute to knowledge by developing a geo-analytic model for the analysis and forecasting of the spread of communicable diseases using a combination of the auto-regressive fractionally integrated moving average (ARFIMA) and spectral density model based on cases of a known location, time of forecast and distance from a point of interest. In addition, the study will provide a real-time system for supporting the interoperability of desktop and mobile devices required for reporting surveillance data for the prompt notification of communicable disease spread.

REFERENCES

[1]. Bansal, S., Chowell, G., Simonsen, L., Vespignani, A. and Viboud, C. (2016). Big Data for Infectious Disease Surveillance and Modeling. *The Journal of Infectious Diseases 214*(4): 375 – 379.

[2]. Morens, D. M. and Fauci, A. S. (2013). Emerging Infectious Diseases: Threats to Human Health and Global Stability. *PLoS Pathology* 9(7): e1003466.

[3]. Bourzac, K. (2014). Infectious disease: Beating the big three. *Nature* **507,** S4–S7.

[4]. Snow K J, Sismanidis C and Denholm J. (2018). The incidence of tuberculosis among adolescents and young adults: A global estimate. Eur Respiratory Journal 51: 1702352

[5]. World health organization (WHO) (2020) Global Tuberculosis Report 2020. Available from https://www.who.int/tb/publications/global_report/TB20_Exec_Sum_20201014.pdf Accessed 25th March, 2021.

[6]. World Health Organization (WHO) (2017). WHO African Region Communicable Diseases Cluster Annual Report.

[7]. World health organization (WHO) (2018) World Tuberculosis Report 2018. Available from http://www.who.int/ Tuberculosis/publications/world-TB -report-2018/WMR - 2018 - annexes.pdf. Accessed 25th March, 2020.

[8]. Guerrisi, C., Turbelin, C. and Blanchon T. (2016). Participatory Syndromic Surveillance of Influenza in Europe. *Journal of Infectious Diseases 214*(4): 386 – 392

[9]. Teutsch, S. M., and Churchill, R. E. (2000). Principles and Practice of Public Health

[10]. Oshitani, H., Ailan, L., Roces, M.C., Sian, D.T., Ken, C. and Kiedrzynski, T. (2005). Implementing the new International Health Regulations in the Pacific – Challenges and Opportunities. *Pacific Health Dialogue Journal 12*(2): 135 – 143.

[11]. Woolhouse, M.E., Rambaut, A. and Kellam, P. (2015). Lessons from Ebola: Improving Infectious Disease Surveillance to inform Outbreak Management. *Scientific Translation Medicine 7*: 307 - 312.

[12]. World Health Organization (2003). WHO Recommended Standards for Surveillance of Selected Vaccine-Preventable Disease Geneva, Switzerland World Health Organization.

[13]. McLeod, K.S. (2000). Our Sence of Snow: The Myth of John Snow in Medic al Geography. *Journal of Social Science and Medicine 50*: 923 – 935.

[14]. Raicharoen, T., Lursinsap, C. and Sanguanbhoki, P. (2003). Application of Critical Support Vector Machine to Time Series Prediction. *In Proceedings of the International Symposium on Circuits and Systems ISCAS* 5: 741 – 744.

[15]. Zhang, G.P. (2007). A Neural Network Ensemble Method with Jittered Training Data for Time Series Forecasting. *Information Sciences 177*: 5329 – 5346.

[16]. Chadsuthi, S., Iamsirithaworn, S., Triampo, W. and Modchang, C. (2015). Modeling Seasonal Influenza Transmission and Its Association with Climate Factors in Thailand Using Time-Series and ARIMAX Analysis. *Computational and Mathematical Methods in Medicine 20:* 1 – 8.

[17]. Rafei, A., Pasha, E. and Orak, R.J. (2015). A Warning Threshold for Monitoring Tuberculosis Surveillance Data: An Alternative to Hidden Markov Model. *Tropical Medicine and International Health 20*(7): 919 – 929.

[18]. Kwame A. G., Amoakoh G. A. and William O. D. (2014). Mathematical Modelling of the Epidemiology of Tuberculosis in the Ashati Region of Ghana. *Journal of Advances in Mathematics and Computer Science.* 4 (3).
[19]. Senyefia B. A. (2017). Non Seasonal ARIMA Modeling and Forecasting of Malaria cases in Children under Five in Edum Banso Sub – District of Ghana. *Asian Research Journal of Mathematics 4(3): 1 – 11*
[20]. Wang H., Tian C.W., Wang W.M. and Luo X. M. (2018). Time Series Analysis of Tuberculosis from 2005 to 2017 in china. *Epidemiology and Infection. Volume 146, Issue 8, PP 935 – 939.*