Winning the Polio War in Pakistan

Prediction and Visualization of Polio Cases

Toorab Khan[1], Waheed Noor[2], Junaid Babar[3], Maheen Bakhtyar[4]
Department of CS and IT, University of Baluchistan
Quetta, Pakistan

***Abstract*—Polio is one of the most important issues which have caught the global attention. It has been eradicated globally except Pakistan and Afghanistan. Its quiet alarming, where whole world is polio free, still polio cases are emerging from Pakistan. The major motivation behind this research is to study and analyze the past cases (trend analysis) and to predict the number of future cases and obstacles hindering Pakistan to eliminate polio. The areas with peak level of influx could be prioritized for effective tracking, planning and monitoring of vaccination activities and better utilization of human resources for targeted and controlled interventions. It shall provide better management and resource allocation decisions for speedy eradication of this epidemic syndrome. Polio cases are displayed on Google Maps for localization and clustering, and trend analysis is performed for future prediction using linear regression.**

***Keywords*—Prediction; visualization; regression; clustering**

I. Introduction

Poliomyelitis causes lifelong paralysis and once suffered it cannot be cured. However, it can be prevented by getting proper and in time vaccination. Since 1988, 99% of polio cases have been reduced as a whole. In 2011, poliomyelitis existed only in four republics of the world i.e. Afghanistan, Nigeria, India and Pakistan. Then in 2014 only three countries reported confirmed polio cases i.e. Afghanistan, Nigeria and Pakistan [1]. At the current stage, polio is confined to only two countries i.e. Pakistan and Afghanistan. Pakistan caught the global attention of the world in the year 2014 when 306/359 confirmed polio cases were reported. While whole world got polio free, Pakistan reporting such a large figure was such an alarming situation.

II. Background

Developing countries are facing very serious health issues along with severe shortage of resources. One of such serious health threats is named as 'Poliomyelitis' [2]. A British physician Michael Underwood described this disease for the first time in 1789 [3]. Word 'Poliomyelitis' is derived from Greek words i.e. 'polio' (meaning grey), 'myelon' (meaning marrow) that of spinal cord and Latin suffix 'itis' (meaning inflammation) [4].

It is one the most deadly and stable virus. It can stay alive in contaminated food and water for several weeks. Patients of poliomyelitis are mostly in the age between two and five years old in the developing countries where the hygienic conditions are pathetic [5].

The global health officials and state health officials in Pakistan had anticipated an endgame for polio in Pakistan. The upward trend in the number of polio cases after 2005 suggested these predictions and indications as premature. Polio eradication activities had been initiated in Pakistan in 1994 by conducting its first Supplementary Immunization Activities (SIAs) for polio. Acute flaccid paralysis (AFP) surveillance started in 1995.

Pakistan achieved significant progress in polio eradication during 1994-2013. During this time period, the average annual polio cases were reduced by nearly 95% as compared to the pre-vaccine time period.

III. Literature Review

This literature review section has been compiled through surveys, books, profound articles, and the other sources relevant to the current specific issue. This review is intended to produce an outline of sources explored whereas researching the spread and cure practices of this disease. It's unquestionable that this analysis fits inside a bigger field of study.

In year 2012 [6] authors used Google Earth for mapping and dissecting the neighborhood food generation regions in the city of Chicago, Illinois. Its motivation was to highlight the urban regions of Chicago where the food production existed. It was carried out with the help of high satellite pictures taken from Google Earth. To show the agribusiness urban zones, Google Earth was utilized to outline urban horticulture locales which were recently reported. Next, they superficially broke down the pictures/map for those reported nourishment creation destinations and found the holes, for example removed some new destinations of urban agribusiness which were already unreported. At that point they completed ground truthing by visiting those destinations. Based on visual investigation of the recorded, final dataset was made and results were assessed including all unlisted and unreported instances.

In year 2013 [7] a contextual investigation is exhibited in which creators shows a web-based mapping application to show a huge number of nurseries in USA on guide. In this application, a business database is utilized for information stockpiling and the application gives modern functionalities to the clients. Instruments used to build up this application were Google Geocoder along with Google Maps API, the database structure of MS SQL, and dynamic webpages were created with the help of MS aspx.NET.

In year 2014 [8] presented trend analysis of obesity in Canada using linear regression. Number of adults (18 years and above) per year was taken as dependent variable and BMI (Body Mass Index) was taken as independent variable.

In 2015 [9] proposed an energy signature heat balance equation and showed thermal performance line of the building. They correlated energy consumption with weather variables. The basics of simple and multiple linear regression analysis were applied.

In year 2016 presented an assessment of relative contribution of heritable versus non-heritable factors, they performed a systems-level analysis on healthy twins between. They measured 204 different parameters and found that 58% almost completely determined by non-heritable influences.

Prediction of dengue outbreak was studied in 2016 by [10]. This analysis composes of a comparative set of prediction models including meteorological and lag disease surveillance variables. Generalized linear regression models were used to fit relationships between the predictor variables and the dengue surveillance data as outcome variable. Model fit were evaluated based on prediction performance in terms of detecting epidemics, and for number of predicted cases. An optimal combination of meteorology and autoregressive lag terms of dengue counts in the past were identified best in predicting dengue incidence and the occurrence of dengue epidemics. Past data on disease surveillance, as predictor alone, visually gave reasonably accurate results for outbreak periods, but not for non-outbreaks periods. A combination of surveillance and meteorological data including lag patterns up to a few years in the past showed most predictive of dengue incidence and occurrence in Yogyakarta, Indonesia.

In 2016 [11] have adapted a simple regressive method in Microsoft Excel that is easily implementable and creates predictive indices. This method trends consumption of antibiotic drugs over time and can identify periods of over- and underuse at the hospital level.

In 2017 [12] , developed a dynamic forecasting model for Zika virus (ZIKV), based on Google Trends (GTs). It was designed to provide Zika virus disease (ZVD) surveillance and detection for Health Departments, and predictive numbers of infection cases, which would allow them sufficient time to implement interventions. In this study, they found a strong correlation between Zika-related GTs and the cumulative numbers of reported cases (confirmed, suspected and total cases; $p<0.001$). Then, they used the correlation data from Zika-related search in GTs and ZIKV epidemics to construct an autoregressive integrated moving average (ARIMA) model for the dynamic estimation of ZIKV outbreaks. The forecasting results indicated that the predicted data by ARIMA model, which used the Google trends search data as the external regressor to enhance the forecasting model and assist the historical epidemic data in improving the quality of the predictions, are quite similar to the actual data during ZIKV epidemic.

IV. Methods

The different aspects of this research including data collection, re-arranging and data pre-processing steps were the compulsory requirement for further utilization of the data. The essentials tasks are explained below:

A. Data Gathering

The required dataset for polio cases in Pakistan was got from National Emergency Operation Center Government of Pakistan, Islamabad with proper permission letter. The dataset obtained, contained the record of confirmed polio cases in Pakistan along with the dataset of missed children, untrained team members and refusal cases for last four years i.e. 2013, 2014, 2015 and 2016. It was available in MS Excel spread sheets. Fig. 1 shows the sample snapshot of acquired data.

B. Data Re-structuring

The accumulated information was not organized in the manner, to be used by Google Maps. There were numerous futile fields of data which were not required, so real task was to remove the useless fields and to organize data in legitimate and admissible format. Some portion of the area names were incorrectly spelled in the accumulated information which were amended accordingly.

C. Data Mapping and Clustering

In this stage, the reformatted information was mapped on Google MyMaps, moreover clustering technique was applied on the information. Fig. 2 shows the snapshot of plotted and clustered data of “Polio cases for the year 2014 on Google MyMaps”.

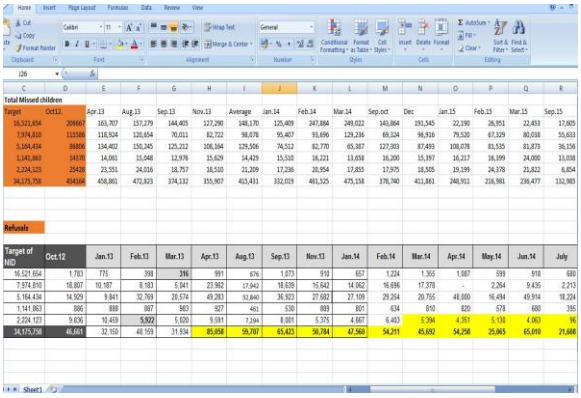

Fig. 1. Dataset by National Emergency Operation Center.

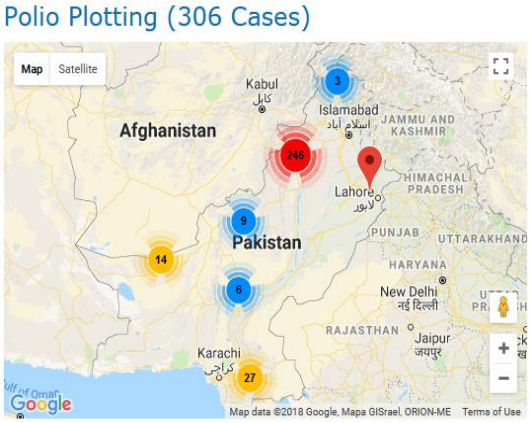

Fig. 2. Polio Plotting for the Year 2014 on Google MyMaps.

Clustering of markers (polio cases) is done by using the Google Maps API and a utility library. This API was available in Java and JavaScript for Android and Web respectively. Our system was web based so we used web API written in JavaScript, namely, "MarkerCluster for Map v3" (Fig. 3).

Mapping and clustering were done so as to give comprehension of the epidemic for further breaking down the patterns and making determinations. Fig. 4 shows the snapshots of geo coordinates data, required for plotting on Google Map.

D. Graphical Representation of Data

In this stage, the information was shown with the assistance of charts and tables to make correlations among the information. Correlations were made among polio cases in various areas for 2013, 2014, 2015 and 2016 and afterward looking at polio cases for the referenced four years inside and out. Fig. 5 shows the bar chart epidemic curve bar chart of confirmed Polio cases in Pakistan for the years 2013-2016.

Fig. 3. JavaScript MarkerCluster API Code Snapshot.

Fig. 4. Geo-Coordinated of Lat/Long Provided to Google MyMaps.

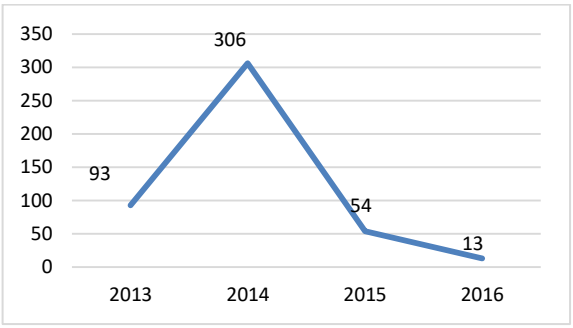

Fig. 5. Epidemic Curve of Polio Cases, 2013-2016.

V. Findings

After evaluating the data visually and graphically, next stage was to discover out the aspects fluctuating polio in Pakistan. After thorough study and discussions with the polio experts in Pakistan, we concluded three basic factors being the reasons behind polio in Pakistan [11].

- Missed Children: Children who are left unvaccinated (unavailability of child due to any reason).
- Refusals: Parents refuse to vaccinate their child. The main reasons behind refusal are misconceptions regarding vaccine that it contains forbidden ingredient or it causes infertility etc.
- Untrained team: It includes the team members who fail to deliver polio vaccines. Reasons include security issues, lack of supervision on team members, untimely payments to the team members, team members who do not know how to enter a record etc.

On the basis of these three dependent factors, we calculated the influence of these variables on effective Polio campaign. Regression Analysis was carried out with comparison to number of Polio cases for each year. After concluding out the factors, trend analysis was carried out of the dependency of the polio cases on the above mentioned three factors using linear regression method.

VI. Results

The relationship between the number of cases for the four years and the missed children for the respective years was analyzed and estimates for the next three years were made using linear regression method. Although the methods could be used to predict future Polio cases for any number of years. MS Excel was the tool used for this purpose. The linear regression equation is.

$$y = a + bx$$

Where Y is independent variable, x is dependent variable, a is y-intercept and b denotes the slope of projected line.

In our prediction model, the number of polio cases depends upon the effective vaccination coverage therefore missed children, refusal cases and untrained team members data are represented by x (on x-axis) as being independent variable and polio cases are represented by y (on y-axis) as being dependent variable. The below listed data was provided by National Emergency Operation Center Government of

Pakistan, Islamabad. The given data for x variables and y variables were fed to MS Excel analysis tool pack function of Regression and further Linear Regression was selected among various regression techniques. The data was fed category wise separately for missed children, refusal cases and untrained team members respectively. The Regression function returned the following regression equations against each category of input data. The regression function also returned the predicted y variables values for future years. The predicted values are plotted on bar chart in this paper "Table I".

The predicted values are plotted on bar chart, with comparison of confirmed Government sources and WHO unconfirmed sources later in this paper.

E. Prediction Results

The results of all three predictive variables are compiled together for better understanding. The trend analysis shows that government is making low progress in Polio data compilation. The progress slow because of many challenges such as the adoptability in remote areas where polio cases still exists. People in remote area of Pakistan are not cooperating with the polio workers and the Government. The second issue is the infrastructure. People are living on the small units spread over the peaks of the mountains where polio workers can't reach easily nor contact them.

TABLE. I. INPUT DATA TO PREDICTION SYSTEM

Description	Year	X Variable	Y Variable	Regression Equation
Missed Children Data	2013 2014 2015 2016	431022 409838 577236 323410	93 306 52 20	y = 158 – 0.00009248x
Refusal Data	2013 2014 2015 2016	53118 45815 56613 48847	93 306 52 20	y= 37-6.467077982x
Untrained Team Data	2013 2014 2015 2016	13319 12554 9116 7734	93 306 52 20	y= 1016.8 – 0.017593x

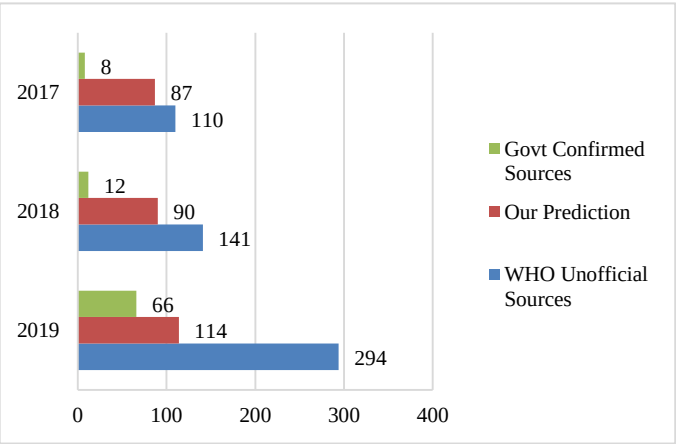

Fig. 6. Future Prediction of Polio Cases.

Our prediction model found the following conclusions:

- We identified an increase in the number of polio cases in future despite the eradication claims made the Government and other Health Organizations.
- The record keeping of confirmed Polio cases by Government agencies does not depict true picture.
- The government still needs to spend more resources on proper vaccination coverage and training of health workers.
- Our prediction model trend analysis is upheld by the unconfirmed sources cited above which are taken from WHO HQ unofficial sources [12].

VII. CONCLUSION AND DISCUSSION

In this research, online portal is created to visualize and predict the Polio cases. Although the same prediction model had been used in many similar trend analyses as presented in literature review, the following short comings handicapped the accuracy and precision of our model: (1) The system relied upon the available data maintained by the Government sources. Which may not reflect true population. (2) The high peak of y variable for the year 2014 i.e. 306 cases may pull the trend line irrationally.

Future work include to make visulization of other related diseases, and inlcude all countries rather only Pakistan.

REFERENCES

[1] S. Kanwal, A. Hussain, S. Mannan and S. Parveen, "Regression in polio eradication in Pakistan: A national tragedy," J Pak Med Assoc, vol. 66, 2006.

[2] T. Ahmad, S. Arif, N. Chaudry and S. Anjum, "Epidemiological Characteristics of Poliomyelitis during the 21st century (2000-2013)," in Proceedings of the International Conference on Software Testing Analysis & Review, 2000.

[3] N. H. D. Jesus, "Epidemics to eradication: the modern history of poliomyelitis," Virology Journal, PubMed NCBI, 2007.

[4] A. Artenstein, "Vaccines," New York: Springer, 2010.

[5] B. Seytre and M. Shaffer, The death of a disease, New Brunswick: Rutgers University Press, 2005.

[6] J. R. Taylor and S. T. Lovell, "Mapping public and private spaces of urban agriculture in Chicago through the analysis of high-resolution aerial images in Google Earth," Landscape and Urban Planning, vol. 108, 2012.

[7] S. Hu and T. Dai, "Online Map Application Development Using Google Maps API, SQL Database, and ASP.NET," International Journal of Information and Communication Technology Research, vol. 3, 2013.

[8] L. K. Twells, D. M. Gregory, J. Reddigan and K. W. Midodzi, "Current and predicted prevalence of obesity in Canada: a trend analysis," CMAJ Open, vol. 2, 2014.

[9] N. Fume and R. Biswas, "Regression analysis for prediction of residential energy consumption," Renewable and Sustainable Energy Reviews, 2015.

[10] L. Aditya, L. Lutfan, L. H. Yien, H. Asa, K. Hari and R. Joacim, "Prediction of Dengue Outbreaks Based on Disease Surveillance and Meteorological Data," PLOS One, 2016.

[11] J. Ahmad, S. Anjum and S. Razzaq, Interviewees, Factors fluctuating polio in Pakistan. [Interview]. 2015.

[12] WHO, "Wild virus type 1 reported from other sources," Data in WHO HQ, 19 Nov 2019. [Online]. Available: http://polioeradication.org/wp-content/uploads/2016/08/Weekly-GPEI-Polio-Analyses-wpv-20191119.pdf.