Factors Affecting Cholera Disease and Its Machine Learning Solution: A Review

AHMAD HAUWA AMSHI[1], RAJESH PRASAD[2]

[1, 2] Department of Computer Science, African University of Science and Technology, Abuja, Nigeria

Abstract— ***The number of African countries reporting indigenous cholera cases as per the report of WHO increased from 24 in 1971 to 30 in 1998 and 36 in 2008, a new high. The latest 2018 cholera epidemic in Nigeria resulted in over 45 thousand cases and a Case Fatality Rate (CFR) of 1. In the year 2021, Nigeria reported a total of 93,362 suspected cholera cases, including 3,283 cholera-related deaths; the highest cases of cholera from the 32 states and the Federal Capital Territory. Cholera is one of the primary causes of morbidity and mortality in Nigeria, from small outbreaks to large epidemics. Because of the limited workforce in the Nigerian healthcare sector and the usage of manual methods, there is an urgent need to build a viable cholera prediction model for early warning mechanisms. This paper investigates the factors affecting the cholera disease within Nigeria and the machine learning solution to it.***

Indexed Terms-- Cholera, Climate change, Machine learning, Logistic regression, and Artificial Neural Network.

I. INTRODUCTION

Cholera is an acute watery diarrheal syndrome and severe intestinal infection caused by the caused by bacterium Vibrio cholerae 01 or 0139 [1], [2]. Though the disease primarily affects the gastrointestinal tract (acute gastroenteritis), the exotoxin produced by Vibrios can cause dehydration, circulatory failure, shock, and electrolyte imbalance due to excessive (and sometimes rapid) fluid and electrolyte loss. Without prompt intervention, the chain of events could lead to acidosis, myocarditis, heart failure, tubular necrosis, and death.

Cholera, generally spread by the fecal-oral route, which is caused by contaminated food or water due to lack of proper sanitation. In developed countries, the majority of cholera cases are transferred by contaminated food, but in undeveloped ones, it is more commonly transmitted through contaminated water and Harvesting seafood such as oysters and mussels in V. cholerae-infected waters might result in food transmission[3].

Seven cholera pandemics have struck the world since 1817. Africa bore the brunt of the global disease burden during the current seventh pandemic. Cholera has gone from an imported to an endemic disease in most African countries in less than a half-century[4]. Between 1970 and 2011, African countries reported 3,221,050 suspected cholera cases as per the report of WHO, accounting for 46% of all cases recorded worldwide. The number of African countries reporting indigenous cholera increased from 24 in 1971 to 30 in 1998 and 36 in 2008, a new high. In 2011, the figure had dropped to 27[1].

The latest 2018 cholera epidemic in Nigeria, which resulted in over 45 thousand cases and a Case Fatality Rate (CFR) of 1.9 %, demonstrates the country's high cholera burden. Furthermore, Nigeria was the only nation in the West African sub-region to record cholera, except Liberia, which reported just two cases [1]. In the year 2021, Nigeria reported a total of 93,362 suspected cholera cases, including 3,283 cholera-related deaths (CFR 3.5 %); The highest cases of cholera from the 32 states and the Federal Capital Territory[5].

In Nigeria, cholera is one of the leading causes of morbidity and mortality, from small outbreaks to large epidemics occurring. In line with global epidemiology, insufficient access to portable water and poor sanitary conditions remains the primary predictors of cholera transmission[5]. Cholera transmission in Nigeria may be aided by several variables, including a lack of access to safe drinking water, an unhygienic environment, natural catastrophes, low literacy, and internal conflicts, which may result in population displacement to IDP camps [6]. Furthermore, persistent armed conflicts, environmental and climatic changes, fast urbanization and population increase, poor emergency or public health responses, and traditional and religious beliefs have all been related to cholera in Nigeria[5].

This paper aims to investigate the factors affecting the cholera disease within Nigeria and the machine learning solution to solve the cholera disease.

The paper is organized as follows: Sec. 2 presents the major factors affecting the cholera; Sec. 3 presents machine learning solutions to cholera while Sec. 4 concludes the paper.

II. FACTORS AFFECTING CHOLERA

Rainfall, temperature, and seas surface temperature (SST) all play a role in cholera outbreaks in Africa, such as rainfall and Solar Oscillation Index (SOI) in Ghana; mean temperature, rainfall, and relative humidity in Nigeria; rainfall in Senegal; maximum temperature and rainfall in Zambia; minimum temperature and rainfall in Zanzibar; and mean temperature and SST in southeastern Africa (Uganda, Kenya, Rwanda, Burundi, Tanzania, Malawi, Zambia, and Mozambique)[1].

In Africa, water likely plays a major role in triggering cholera outbreaks, with lagoons and estuaries contributing to coastal disease, and lakes and rivers contributing to inland disease[4]. In general, floods increase the spread of water-borne diseases through disrupting access to or contaminating safe water sources; affecting sanitation conditions; and limiting access to essential health services which is mainly caused by increased rainfall[7].

The key climatic drivers of cholera transmission in Nigeria have been identified as rainfall and temperature[5]. In other contexts, the relationship between cholera epidemics and climatic variables, particularly seasonal tropical rainfall, has been well documented[5]. Two mechanistic models for cholera transmission about rainfall have been proposed: cholera transmission tends to be enhanced due to the high tendency for consumption of contaminated water and worsening sanitary conditions caused by floods; and the ease with which water becomes contaminated by freshly excreted bacteria resulting from washout of open-air defecation sites or overflows from pit latrines during and after rainfall.

Similarly, lack of proper sewage disposal has been identified as another causative factor because people sell and buy food closer to the bins, which in turn contaminates the food and water around the area. A cholera epidemic can occur when people consume the infected foods and water. Since Cholera is spread through the fecal-oral route, either directly from person to person or indirectly through contaminated fluids from an environmental reservoir of varying duration, food, and potentially flies and fomites[1].

According to [4], High population density combined with poor quality informal housing may influence cholera incidence and outbreak amplification by facilitating person-to-person transmission and increasing the burden on inadequate sanitation facilities. Cholera attack rates in Harare ranged from 1.2 cases per 1,000 people in low-density residential areas to 90.3 cases per 1,000 people in an overcrowded suburb in 2008–2009, with similar trends in Ghana and Uganda. Mean temperature and SST in Malawi, Zambia, and Mozambique[8].

Studies in Kenya [1] shows that the potential impact of rising armed conflicts on recurrent cholera transmission is a global phenomenon. Nigeria is facing high levels of insecurity as driven by several factors, such as the Boko Haram/ Islamic State–West Africa Province insurgency in the northeast, attacks by armed groups in the northwest, farmer-herder conflict in the Middle Belt region, and separatist agitation in the southeast which some violent attacks have been targeted towards health facilities and workers [9].

Generally, genetic mutation has been demonstrated to be linked with the emergence of new, virulent, and drug-resistant strains of V.cholerae [4]. For example, that the seventh cholera pandemic became prominent in 1961 after V.cholerae underwent a series of mutations, with suitable niches in the Middle East and gene sources from Makassar to aid the genetic events wherein the 2009 and 2010 cholera outbreaks in Nigeria were linked to multidrug-resistant a typical El

Tor strains. Table 1 presents the summary of the factors affecting Cholera in different Sub- Saharan countries.

Table 1: Factors affecting Cholera in different Sub-Saharan countries

S/N	Country	Major Factors	References
1.	Nigeria	Climate Variability, Population, Poverty, Insecurity, Poor Hygiene	[1], [5], [6], [9]
2.	Kenya	Climate Variability, Poor Hygiene and War	[1], [4]
3.	Mozambique	Weather Conditions (Rainfall), War and Urbanization	[4] [10]
4.	Ghana	Weather Variables, Population and Urbanization	[1], [4],[10]
5.	Tanzania	Climate variables	[3]

III. MACHINE LEARNING SOLUTION TO CHOLERA

There is a need to further investigate the applications of machine learning (ML) in healthcare to realize cutting-edge technologies for the early prediction of cholera incidence. Identification of high-risk areas for deadly infectious and non-infectious disease outbreaks is important so that deadly disease outbreaks can be predicted and detected, and responses to these deadly disease outbreaks can be made more efficient. Health agencies can leverage the use of Machine Learning (ML) technologies to limit the spread of severe infectious disease outbreaks in a variety of ways[11]. Machine learning algorithms can be used to learn datasets containing details about known viruses, animal populations, human demographics, biology, and biodiversity information, available physical infrastructures, cultural/social practices around the world, and disease geolocation to predict disease outbreaks and can also learn integrated multi-source data such as travel schedules, population, logistics, and epidemiological data to predict disease location and rate of spread[12].

Author's in [3] explores the use of machine learning techniques to model cholera epidemics with linkage to seasonal weather changes while overcoming the data imbalance problem. Adaptive Synthetic Sampling Approach (ADASYN) and Principal Component Analysis (PCA) were used to restore sampling balance and dimensional of the dataset. In addition, sensitivity, specificity, and balanced-accuracy metrics were used to evaluate the performance of the seven models. Based on the results of the Wilcoxon sign-rank test and features of the models, the XGBoost classifier was selected to be the best model for the study.

S/N	Methods	Advantages	Disadvantages	References
1	1. Adaptive Synthetic Sampling Approach (ADASYN). 2.Principal Component Analysis (PCA) 3. XGBoost classifier	The selected model is useful in predicting accurately cholera epidemics using future weather variables	The study could not be treated as a time series problem due to the poor quality of data and data collection bias	[3]
2	SARIMA	This study revealed that the relationship between cholera incidence and climatic variables varies with locations and climatic variables.	Not covering temporal variable.	[8]

3	1.Synthetic Minority Oversampling Technique (SMOTE) 2. Random Forest	The model showed promising results when tested on individual districts in coastal India, underlining its potential to perform accurately across a country with large climatological differences evident spatially.	It was less effective in areas where there are fewer cholera outbreaks to both train and test the model. These findings suggest that the model would be more suited for detecting cholera outbreaks in endemic areas but might be less likely to detect more sporadic, epidemic cholera events.	[11]
4	1. logistic regression and Correlate bivariate 2. Artificial Neural Network	This study showed higher temperature (average 27 degree centigrade) had a significant effect on the incidence of cholera. Therefore, the influence of climate conditions in cholera incidence has been well emphasized.	Lack of data on some hygienic, social and demographic indicators has been one of the limitations of prediction. Adding additional hygienic, social and demographic variables would enhance the potential of model.	[12, 13]

IV. CONCLUSION

Making a quick informed decision in response to infectious disease outbreaks is crucial to decrease the damage caused by the impact of disease outbreaks after a disease event is identified.

In a bid to contribute to the mandate of the global health agenda of 2030, there is a dire need to forecast/predict epidemic diseases to provide appropriate preventive actions in disease spread, so that preparedness and emergency response plans can be taken into consideration more comprehensively than at present. The findings of this study would be immensely beneficial to a climatologist, epidemiologist, or public health expert that works with cholera incidence to create preparedness and response plans. It is significant for climatologists because the impact of climate change on cholera incidence may be predicted from this study, as climatologists predict an increase in mean temperature of 1.4–5.8ºC during the next 100 years. An epidemiologist would be helped by the new insights on environmental and climatic linkages of cholera outbreaks. Finally, a health professional can plan for potential coping and adaptation methods for anticipated climate change-related health hazards. This study will also contribute towards the development of a climate-based early warning system for cholera.

REFERENCES

[1] J. Deen, M. A. Mengel, and J. D. Clemens, “Epidemiology of cholera,” *Vaccine*, vol. 38. Elsevier Ltd, pp. A31–A40, Feb. 29, 2020, doi: 10.1016/j.vaccine.2019.07.078.

[2] K. Asunduwa *et al.*, “Descriptive analysis of a cholera outbreak in 14 LGAs of Sokoto State – Nigeria, 2018,” *Int. J. Infect. Dis.*, vol. 101, p. 363, Dec. 2020, doi: 10.1016/j.ijid.2020.09.951.

[3] J. Leo, E. Luhanga, and K. Michael, “Machine Learning Model for Imbalanced Cholera Dataset in Tanzania,” *The Scientific World Journal 2019(1):1-12*, doi: 10.1155/2019/9397578.

[4] M. A. Mengel, I. Delrieu, L. Heyerdahl, and B. D. Gessner, “Cholera Outbreaks in Africa,” no. May, pp. 117–144, 2014, doi: 10.1007/82.

[5] E. A. Salubi and S. J. Elliott, “Geospatial analysis of cholera patterns in Nigeria: findings from a cross-sectional study,” *BMC Infect. Dis.*, vol. 21, no. 1, 202, Dec. 2021, doi: 10.1186/s12879-021-05894-2.

[6] G. C. Leckebusch and A. F. Abdussalam, “Climate and socioeconomic influences on interannual variability of cholera in Nigeria,” *Heal. Place*, vol. 34, pp. 107–117, Jul. 2015, doi: 10.1016/j.healthplace.2015.04.006.

[7] K. O. Elimian *et al.*, “What are the drivers of recurrent cholera transmission in Nigeria? Evidence from a scoping review,” *BMC Public Health*, vol. 20, no. 1, 432 Apr. 2020, doi: 10.1186/s12889-020-08521-y.

[8] S. S. Daisy, A. K. M. Saiful Islam, A. S. Akanda, A. S. G. Faruque, N. Amin, and P. K. M. Jensen, “Developing a forecasting model for cholera incidence in Dhaka megacity through time series climate data,” *J. Water Health*, vol. 18, no. 2, pp. 207–223, Apr. 2020, doi: 10.2166/wh.2020.133.

[9] P. E. Idoga, M. Toycan, and M. A. Zayyad, “Analysis of factors contributing to the spread of cholera in developing countries,” *Eurasian J. Med.*, vol. 51, no. 2, pp. 121–127, 2019, doi: 10.5152/eurasianjmed.2019.18334.

[10] A. Zerbo, R. Castro Delgado, and P. A. González, “A review of the risk of cholera outbreaks and urbanization in sub-Saharan Africa,” *Journal of Biosafety and Biosecurity*, vol. 2, no. 2. KeAi Communications Co., pp. 71–76, Dec. 01, 2020, doi: 10.1016/j.jobb.2020.11.004.

[11] A. M. Campbell, M. F. Racault, S. Goult, and A. Laurenson, “Cholera risk: A machine learning approach applied to essential climate variables,” *Int. J. Environ. Res. Public Health*, vol. 17, no. 24, pp. 1–24, Dec. 2020, doi: 10.3390/ijerph17249378.

[12] B. Teri Musa and O. Samuel Olayemi, “APPLICATION OF LOGISTIC REGRESSION MODELS FOR THE EVALUATION OF CHOLERA OUTBREAK IN ADAMAWA STATE NIGERIA,” *Int. J. Math. Stat. Stud.*, vol. 8, no. 1, pp. 32–54, 2020.

[13] Z. Pezeshki, M. Tafazzoli-Shadpour, I. Nejadgholi, A. Mansourian, and M. Rahbar, “Model of Cholera Forecasting Using Artificial Neural Network in Chabahar City, Iran,” *Int. J. Enteric Pathog.*, vol. 4, no. 1, Feb. 2016, 5, doi: 10.17795/ijep31445.