Overview of Interactive Chatbot for Modelling, Predicting and Reporting Covid-19 in Mano River Union

Olutola Olaide Fagbolu, Mohamed Sellu & Tolulope Adeola Fagbolu

University of Sierra Leone

ABSTRACT

COVID-19 is a phenomenal pandemic that wreaked havoc and still affecting all facets of the human race globally, a novel virus since December 2019 with a record of millions of confirmed cases and associated mortality of hundredth of thousands in well over 230 countries of the world and these cases rise daily. With the incessant increase in the number of cases in Mano River Union (Sierra Leone, Guinea and Liberia), second wave and new variants. The Basic Reproduction Number (R_0) over a time frame using recorded incident cases of the World Health Organization (WHO) by governments of Sierra Leone, Guinea and Liberia.

Keywords: chatbot, covid-19, NLP parser, RNN health.

Classification: K.8.1

Language: English

London Journals Press

LJP Copyright ID: 975824
Print ISSN: 2514-863X
Online ISSN: 2514-8648

London Journal of Research in Computer Science and Technology

Volume 21 | Issue 1 | Compilation 1.0

Overview of Interactive Chatbot for Modelling, Predicting and Reporting Covid-19 in Mano River Union

Olutola Olaide Fagbolu[α], Mohamed Sellu[σ] & Tolulope Adeola Fagbolu[ρ]

ABSTRACT

COVID-19 is a phenomenal pandemic that wreaked havoc and still affecting all facets of the human race globally, a novel virus since December 2019 with a record of millions of confirmed cases and associated mortality of hundredth of thousands in well over 230 countries of the world and these cases rise daily. With the incessant increase in the number of cases in Mano River Union (Sierra Leone, Guinea and Liberia), second wave and new variants. The Basic Reproduction Number (R_o) over a time frame using recorded incident cases of the World Health Organization (WHO) by governments of Sierra Leone, Guinea and Liberia. The exponential growth method estimates the growth rate of COVID-19 and R_o using R Survival Analysis Packages and functions to report infection rate, mortality rate and offers live information for planning and preventive measures. With WHO speculations that the virus has come to stay with the human populace, there is an urgent need to explore how computing statistics with Natural Language Processing (NLP) will salvage the infection rates, mortality rates, FAQs. NLP parser is used to extract related information from Emergency Department reports that serve as dataset coupled with the death toll and patient counts as of July 24, 2020, to develop an interactive chatbot that gives preventive measures, symptoms, predict R_o, report routine statistical data, FAQs about COVID-19, emergency contacts for all the provinces in Sierra Leone, Guinea and Liberia and general toll-number for Ministries of Information and Communication, Health, etc. This research work is done through intensive and extensive assessments, observations, and information on the case by case of patients to develop the chatbot Covid19Mano. Dialog flow open-source environment is used with PHP for documenting the content of the database for reprogrammed questions, phrases, or words about COVID-19, and NLP parser was integrated with Facebook Messenger and Whatsapp to test the efficiency and accuracy of the chatbot. It offers encoder decoder models for sequence-to-sequence prediction problems in question answering, text and speech translation, and many more magical and exciting features when trained with Recurrent Neural Networks(RNN). The chatbot will enable users to get up-to-date information about the Coronavirus pandemic, its spread in Mano River Union, who to contact, what to do, and many more related challenges such as predictive analytics of infection, transmission, death, and recovery records and consequently model these mathematically. All the information, analytics, graphics would be embedded into government websites of Mano River Union to combat the ravaging fake news, myth, and stigmatization about the COVID-19 and offer the general overview of modelling, reporting, and predicting the pandemic effect on these nations of Mano River Union.

Keywords: chatbot, covid-19, NLP parser, RNN health.

Author α σ: Department of Mathematics and Statistics Fourah Bay College, University of Sierra Leone, Freetown Sierra Leone. e-mail: fagbolu.olutola@usl.edu.sl

ρ: Department of Health Policy and Management, Faculty of Public Health, University College Hospital, Ibadan, Nigeria.

I. INTRODUCTION

A novel coronavirus has been a global threat since December 30 2019 with a record of many millions of confirmed cases associated with many thousan-

London Journal of Research in Computer Science and Technology

ds of deaths in over 204 countries of the world, however, these cases rise daily (WHO, 2020a). Coronaviruses are family viruses that range from the common cold to MERS (Middle East Respiratory Syndrome) coronavirus discovered in 2012 and SARS (Severe Acute Respiratory Syndrome) coronavirus discovered in 2003. They are circulating in animals and some of these coronaviruses can spill over from animals to humans. SARS-CoV-1 was first identified in China in 2003, caused more than 8000 cases in 33 countries over eight months, it was discovered to be transferred to humans from civet cats. In 2012, WHO was notified of 2,494 confirmed cases of MERS-CoV with 858 associated deaths, first identified in Saudi Arabia and it was transferred to humans from camels. COVID-19 is a zoonotic virus from phylogenetic analyses undertaken with available full genome sequences, bats appear to be the reservoir of the COVID-19 virus, but the intermediary host(s) has not yet been identified (WHO, 2020c). The current outbreak dynamics is human-to-human transmission. Current estimates of the incubation period of the virus range from 1-14 days and these estimates will be refined as more data become available. Understanding the time when infected patients may transmit the virus to others is critical for control efforts. Detailed epidemiological information from more people infected is needed to determine the infectious period of 2019-nCoV, in particular, whether transmission can occur from asymptomatic individuals or during the incubation period (WHO, 2020b).

The World Health Organization strategies through a combination of public health actions, such as rapid identification, diagnosis and management of the cases, identification and follow up of the contacts, infection prevention and control in healthcare settings, implementation of health measures for travellers, raising awareness among the populace, communication of risks, development of chatbots but without any ability to predict the transmission rate, mortality rate and model them but WHO chatbot only gives overview report of cases without capability to model, and predict appropriately (WHO, 2020b).

The novel COVID-19 exhibits non-specific symptoms and the disease's appearance can range from no symptoms (asymptomatic) to severe pneumonia and death. Most asymptomatic infections have relatively rare cases that are asymptomatic on the date of identification that went on to develop the disease. The proportion of truly asymptomatic infections is unclear but appears to be relatively rare and does not appear to be a major driver of transmission. Most people infected with the COVID-19 virus have mild disease and recover. For most people, COVID-19 infection will cause mild illness however, it can make some people very sick and it can be fatal in other people. Older people aged over 60 years, those with pre-existing medical conditions such as cardiovascular disease, chronic respiratory disease, hypertension, or diabetes are at high risk (WHO, 2020a). Approximately 80% of laboratory-confirmed patients have had mild to moderate disease, which inclu- des non-pneumonia and pneumonia cases, 13.8% have severe disease (dyspnea, respiratory frequency ≥30/minute, blood oxygen saturation ≤93%, lung infiltrates >50% of the lung field within two to four days) and 6.1% are critical (respiratory failure, septic shock, and/or multiple organ dysfunction). Disease in children appears to be relatively rare and mild with approximately 2.4% of the total reported cases reported amongst individuals aged less than 19 years. A very small proportion of those aged under 19 years have developed severe (2.5%) or critical disease (0.2%) (WHO, 2020c). Anosmia (inability to smell) and in some people inability to taste have been reported as a symptom of COVID -19 infection (ECDC, 2020).

The importance of global surveillance is to monitor trends in the disease where human-to-human transmission occurs, hurriedly discover new cases in countries where the virus is not circulating, provide epidemiological information to conduct risk assessments at the national, regional, and global levels, provide epidemiological information to guide preparedness and response measures.

A case can be identified when a patient with acute respiratory illness (that is, fever and at least one sign or symptom of respiratory disease, for example, cough or shortness of breath) with no other etiology that fully explains the clinical presentation, a history of travel to or residence in a cou-

ntry, area or territory that has reported local transmission of COVID-19 disease during the 14 days before symptom onset. Specimens from the lower respiratory tract, such as sputum, endotracheal aspirate, or bronchoalveolar lavage, can be collected for investigation. If patients do not have signs or symptoms of lower respiratory tract disease or if specimen collection for lower respiratory tract disease is clinically indicated but the collection is not possible, upper respiratory tract specimens, such as a nasopharyngeal aspirate or combined nasopharyngeal and oropharyngeal swabs should be collected. If initial testing is negative in a patient who is strongly suspected to have COVID-19 infection, specimens should be collected again from multiple respiratory tract sites (such as the nose) and should also include sputum and endotracheal aspirate. Additional specimens may be collected, such as blood, urine, and stool, to monitor the presence of virus and shedding of virus from different body compartments. A confirmed case is a person with laboratory confirmation of infection with the COVID-19 virus irrespective of clinical signs and symptoms (WHO, 2020d, 2020b).

It is known that the virus gains access to the body through the eyes, nose, and mouth, so it is advised to avoid touching one's face with an unwashed hand. Washing of hands with soap and running water for at least 30 seconds, or cleaning hands thoroughly with alcohol-based solutions, tissues, or gels. It is also recommended that social distancing (to stay one meter or more away) from people infected with COVID-19, who are showing symptoms, to reduce the risk of infections through respiratory droplets. Practicing voluntary self -isolation by infected people, basic hand hygiene, and respiratory hygiene (sneezing or coughing into a tissue which is immediately disposed of properly or covering one's cough with the elbow) has been recommended to reduce the effect on humans, avoid unprotected contact with live animals. There is no specific treatment for coronaviruses but symptoms can be treated.

In Africa, there are cases of COVID-19 in well over 50 countries with almost six thousand deaths and Mano River Union has thousands of confirmed cases with many casualties hence there is a need to give preventive measures, symptoms, FAQs (Frequently Asked Questions) on COVID-19, Emergency contacts for all Districts in Mano River Union and the general toll-free number of Ministries of Health in Mano River Union, live info on the cases of the virus, quiz game and graphical infection rate, predictive analytics for deaths and recovery on chatbot called Covid19 Mano to reduce mortality or avoid more causality. A chatbot is Artificial Intelligence oriented natural language processor that communicates indistinguishably from human or another participant of communication, the essence of a chatbot is to make users feel that they are communicating with any living human and it becomes imperative in the management of this global outbreak of Coronavirus because of its ability to guide during and after this pandemic with natural language using voice or text. Although several efforts have been put in place to promote health literacy, Most outpatient information for COVID-19 in Mano River Union is provided on mass media, print media, public places, hospitals and other fora despite present publicities, health campaigns, efforts, and methods that are being used to promote health literacy on COVID-19 to the public, the rate of deaths as a result of COVID-19 continue to increase at an alarming rate. The majority of the populace gets to know information about COVID-19 when it reaches their nation, district, city, or town and at that time it might already be too late for them. The versatility of chatbot has found its usage in relatively all domains including customer care, marketing, teaching and learning, training. It handles the day-to-day tasks of office assistants, booking hotel rooms, making reservations, etc. A chatbot is a smart assistant that can be a software tool, algorithm or artificial intelligence that communicate with a human or another participant on a particular topic(s), area of interest, customer's query or satisfactions, or any other specific domain, it can be text-based which follow rule-based approach to respond to queries or voice-based which interact through voice and respond to written commands at times through voice communications or combined both text and voice features (Smutny & Schreiberova, 2020). The technology avails end- users the platform to

interact with digital entities on a user-friendly and renowned social media - Facebook and Whatsapp which are the preferred option in Sierra Leone, Guinea and Liberia because of their wider acceptability among the populace and ease of use. The chatbot on Facebook Messenger will be able to provide accurate and consistent information about COVID-19 in Mano River Union in both English and French Languages, answer any bothering questions, discuss related topics, inform the society about this pandemic and any other related task on COVID-19 such as predictive information about transm- ission, deaths and recovery(Sanni S.A,Rajabu N and Fagbolu O.O, 2018).

This research work cut across several interdisciplinary domains of public health informatics, mathematical modelling and natural language processing that design and implement an interactive chatbot to offer adequate, timely information about the COVID-19 outbreak in Sierra Leone, Liberia and Guinea or any other future health and wellness challenge(s) to increase awareness of health literacy and policies, prevention and wellness of human. Emphatic mentions are previously developed chatbots such as ELIZA in 1956, ALICE, Claude, Hex, TAY, Xiaoice, etc. ALICE is based on NLP and pattern matching while relating responses from knowledge records while Claude picks inputs, responds based on its database and finally gives answers. Hex possesses an exemplary feature of being able to introduce new topics after answering any posed question. Xiaoice (literally Microsoft little ice) acts as a seventeen-year-old girl on Weibo and conversation can be established with her (Smutny & Schreiberova, 2020)(ZEMČÍK, 2019). The chatbot will not only comprise content managed technology that runs only on Facebook messenger and Whatsapp but can be deployed on email, SMS and any other chatbot compliant social media with the ability to estimate using gathered data and statistics(Sanni S.A, Rajabu N and Fagbolu O.O). The source of motivation for this research work is the need to provide reliable, accurate and straight- forward information about the growth and spread of this pandemic to curb the ravaging fake news, myth and stigmatization about the virus.

II. STATEMENT OF PROBLEM

Several chatbots for COVID-19 have been developed with various levels of success reported in the literature. However, no available chatbot is supported with adequate estimation, predictive model of infection, transmission and recovery based on the data and statistics rather they are only reporting daily situations. The research work will address the problem by developing an interactive chatbot that will estimate, model and report Basic Reproduction Number (R_0) for COVID-19 over time using incident number cases that are recorded in Mano River Union. It will rapidly proffer a digital time capsule for future researchers of COVID-19 with innovative responses to FAQs, information, and other sundry matters on this pandemic while setting standard measures for any other related challenges.

Developing a light-proof bilingual chatbot to support empirical information of the World Health Organization (WHO), governments in Mano River Union and report, predict and model the information on Facebook messenger and Whatsapp platform for the populace. This will serve as an improvement in previously created WHO chatbot and boost modelling techniques of the pandemic, estimation, and provide an accurate projection of infection, recovery and transmission.

The long-term impact of the project will be on health policy and management, planning, economic and technological application with its benefits. The fundamental benefits being in the health sector, this is expected to permeate other aspects of the society as Mano River Union can plan, project and estimate accurately the rate of recovery, mortality and transmission using mathematical models. Community engagement would further be strengthened by the partnership between the governments of Mano River Union and academia. Aspects of this work will foster study exchanges in both lingua franca, support ongoing undergraduate and postgraduate students' training for long term manpower deve- lopment.

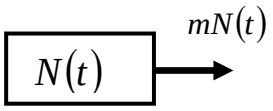
III. RESEARCH OBJECTIVES

The aim is to model predictive analytics of gathered information about COVID-19 in Mano River Union periodically, estimate the rate of growth of the COVID-19 in Sierra Leone (k_s), Guinea (k_g) and Liberia (k_l) and offer situation reports. The study will evaluate numerous factors affecting the detection and prevention of the spread of COVID-19 in Mano River Union, extract relevant information about infected patients from clinical reports to travelling history, contacts and design an interactive chatbot that will combat the consequent effect of inadequate information about this challenging pandemic, predict and model COVID-19 based on the size of the population of individuals that get infected at a particular rate, and increase awareness of health literacy. The specific objectives are-

1) To gather cumulative cases of COVID-19 in Mano River Union with age and sex.
2) To formulate a reality-based model for the chatbot using Recurrent Neural Networks (RNN).
3) To develop interactive chatbot by including accuracy of Natural Language Processing and present its general overview.
4) To test and prototype the chatbot on Facebook messenger and Whatsapp plat- forms.

IV. METHODOLOGY

The source of motivation for this research work is the need to provide reliable, accurate, and straightforward information about the growth and spread of this pandemic to combat the ravaging fake news, myth and stigmatization about the virus. The exponential decline or growth method will be employed to estimate the size of the population of Mano River Union that is Liberia, Sierra Leone and Guinea population assuming that individual gets infected at a constant rate (Vynnycky, E and White, 2010). The model has the following structure:



Here, $N(t)$ is the number of individuals at the time t, m is the rate of growth of the pandemic and so $mN(t)$ is the number of individuals who get infected per unit time. The differential equation for the model is as follows:

$$\frac{dN(t)}{dt} = -mN(t) \quad \text{.....(1)}$$

The assumption that once infected, individuals become infectious at a constant rate, or that once infectious individuals recover and become immune at a constant rate is often incorporated into infectious disease models. This assumption is convenient since the average pre-infectious and infectious periods are unknown.

The result of equation (1) solves to give

$$N(t) = N(0)e^{-mt} \quad \text{.....(2)}$$

as illustrated below:

By differentiating the expression $N(t) = N(0)e^{-mt}$ using the rules of differentiation, we obtain the following expression: $\frac{dN(t)}{dt} = -mN(0)e^{-mt}$, since $N(t) = N(0)e^{-mt}$ we obtain our desired result: $\frac{dN(t)}{dt} = -mN(t)$

Attack and Case-Fatality Rates -in general, the Attack Rates for an infectious disease is measured as:

$AR = \frac{\text{Number of who get sick}}{\text{Number of people at risk}}$ in a timeframe: duration of the pandemic.

For the Case-Fatality Rates, where the cases are the people with disease and fatality is how many deaths is measure as:

$CFR = \frac{\text{Number of deaths}}{\text{Number of cases}}$ in a timeframe: duration of the pandemic

To achieve all these objectives, an appropriate flowchart that represents all the requirements were done as depicted in figure 1. The statistical data from different sources were obtained and these sources include WHO, Ministries and Parastatals and District Health Information Desks which are pre-processed by cleaning (removing noise, stop words, and converting to either uppercase or lowercase) furthermore, the research approach would proceed to tokenization if cleaning is accomplished and stemming and lemmatization are employed to split sentences and texts into constituent words afterward vectorization was done through TF-IDF (Term Frequency-Inverse Document Frequency) instead of the commonly used bag of words. The test data created Long Short Term Memory (LSTM), train and save model before allowing any input from the users and finally, responses were generated.

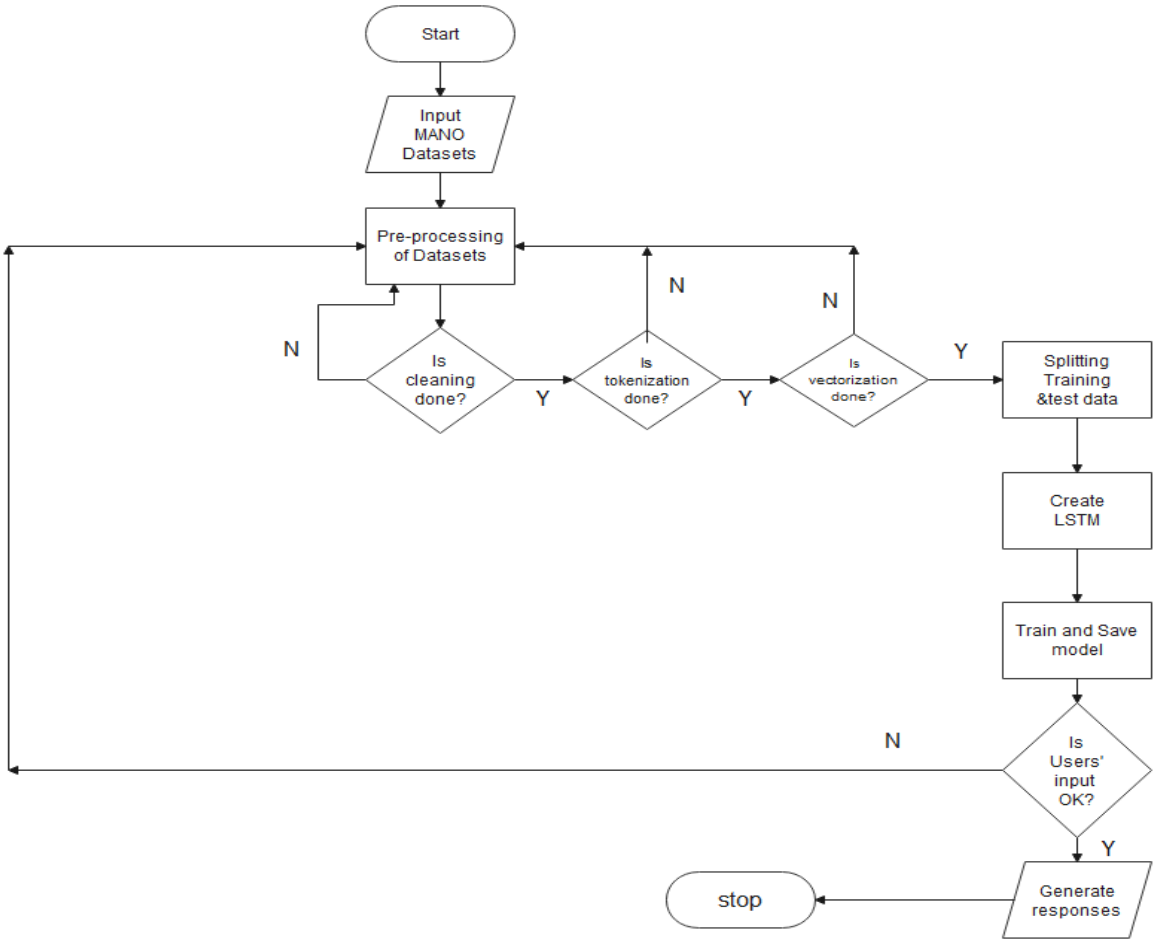


Figure 1: Flowchart for the Design methods

As seen in figure 2, the rectangular shape and flowlines depict vector and functions respectively as in matrix multiplication, input vectors are in red, output vectors are blue and RNN's states are sequence output, sequence input, both sequence output and input and synced sequence input and output (Karpathy, 2015)

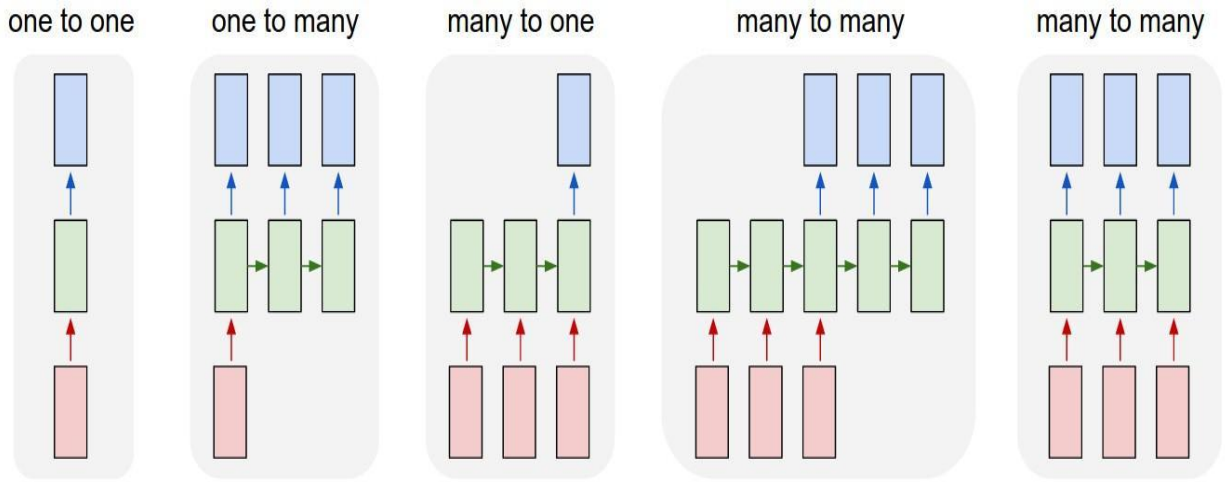


Figure 2: Sequences of Recurrent Neural Networks (RNN)(Karpathy, 2015)

The reality-based model will rely on rules, statistical methods, or templates of encoder- decoder models to situate interactive chatbot as an example of RNN for sequence-to-sequence prediction problems in question answering, text and speech translation, video captioning, and many more.

The programming tools needed are Facebook, Twitter, WhatsApp, Python, R, AWS Aurora, Node.js, Express JS, AWS AI/ML. The collected data and derived mathematical models will be used as a data set to train chatbot (Covid19Mano) over a specific period.

4.1 Design activities and output

Activity	Expected Outcome
Collection of data and frequency of infections in population as described in terms of incidence and prevalence	Data and statistics based on projection for the susceptible population for a particular time during the pandemic, mathematical equations, and subsequent models
System Specification	Study of requirement and technical documentation of interactive chatbot.
System Design	System architecture, user-computer interface, web-interface, algorithms for the chatbot processes.
System Implementation	Formatting data and statistics results and creation of program data files, programming using Python with other open-source tools and applications as well as Java Android platforms.
System evaluation	Report on analysis of gathered cumulative cases in Mano River Union questionnaire administered to patients, health workers, and cross-sectional and sample population.

4.2 Activity indicators

1) Demonstrated design of Chatbot.
2) Documentation of the mathematical model and system specification.
3) Presentation of system architecture, user-computer interface, web-interface, algorithms for the chatbot processes.
4) System testing (running chatbot on different platforms).
5) Deployed chatbot for use by selected stakeholders within Mano River Union to determine the real-life and practical impacts and the chatbot (Covid19Mano) hosted on the web.

V. LIMITATIONS

Despite the inherent advantages and contribution to social, scientific, and health-related knowledge, the following are some of its limitations that have made the chatbot to be at its overview stage.

1) The training cost is exorbitant and without training the chatbot would only be generative and whenever is in a dilemma for a certain query, it will generate a response by recalling fixed answers without evidence as prevalent in the retrieval-based approach.
2) Training and retraining are imminent once new data are added and by considering the robust nature of COVID-19.

CONCLUSION

The data collection procedure involved primary and secondary data. The primary data were collected using both qualitative and quantitative techniques. The quantitative technique involved the administration of structured questionnaires and observation while the qualitative technique involved conducting in-depth interviews. Pre-tests on chatbot were administered by the researchers using tests containing questions selected from FAQs. The Natural Language Understanding (NLU), semantic frame, and Dialogue Management (DM) were used to train the chatbot. R Survival Analysis Packages and functions reported infection rate, mortality rate and offer live information for planning and prevention measures. This chatbot with RNN architecture was developed, evaluated with datasets from Mano River Union and remember the contexts, queries and keep them to generate responses, the model is

basic without any complexity and offers high accuracy for large data such as Covid19 information in Sierra Leone, Guinea and Liberia. Sequence input and sequence output provided a bilingual feature, for example, text or datasets can be in the English language and the responses are displayed in French and vice versa, most especially for Guinea as a francophone nation. In the future, Bidirectional Recurrent Neural Networks (BRNN) and attention mechanism can be deployed in the development of interactive chatbot with recommender systems on what to do, food to eat, and many more precautionary approaches would be part of the chatbot.

REFERENCES

1. ECDC, E. C. for D. P. and C. (2020). Question and Answers on Covid-19. European Centre for Disease Prevention and Control.
2. Karpathy, A. (2015). The Unreasonable Effectiveness of Recurrent Neural Networks. Web Page, 1–28. http://karpathy.github.io/2015/05/21/rnn-effectiveness/.
3. Sanni S.A, Rajabu N and Fagbolu O.O Interactive Chatbot to support Interns during Internship Program, 8th East African Education Quality Assurance Forum (EAQAN), May 7-10, 2018. Rwanda.
4. Smutny, P., & Schreiberova, P. (2020). Computers & Education Chatbots for learning : A review of educational chatbots for the Facebook Messenger. Computers & Education, 151(June2019), 103862. https://doi.org/10.1016 /j.compedu.2020.103862.
5. Vynnycky, E and White, R.. (2010). An Introduction to Infectious Disease Modelling. Oxford University Press.
6. WHO, W. H. O. (2020a). Coronavirus disease 2019 (COVID-19) Situation Report – 70 (Vol. 70, Issue March).
7. WHO, W. H. O. (2020b). Novel Coronavirus (2019-nCoV) Situation Report - 7 (Issue 7).
8. WHO, W. H. O. (2020c). Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19) (Issue February).
9. WHO, W. H. O. (2020d). The First Few X cases and contacts (FFX) investigation protocol for coronavirus disease 2019 (COVID-19) (Issue February).
10. ZEMČÍK, M. T. (2019). A Brief History of Chatbots. DEStech Transactions on Computer Science and Engineering, aicae. https://doi.org/10.12783/dtcse/aicae2019/31439.