Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Explainable features responsible for the high or low spread of SARS-CoV-2: Africa in view

Olalekan J. Akintande\textsuperscript{a,∗}, Olusanya E. Olubusoye\textsuperscript{a}, Olaoluwa S. Yaya\textsuperscript{a}, Adeyinka O. Abiodun\textsuperscript{b}

\textsuperscript{a}Department of Statistics, Laboratory for Interdisciplinary Statistical Analysis, Computational Unit, University of Ibadan, Nigeria
\textsuperscript{b}Africa Center of Excellence on Technology Enhanced Learning, National Open University of Nigeria, Nigeria

A R T I C L E   I N F O

Article history:
Received 24 July 2020
Revised 27 February 2021
Accepted 22 July 2022

Editor: DR B Gyampoh

Keywords:
Share of global infection (SGIR)
Low or high spread
SARS-CoV-2
Naïve Bayes
Features interactions
Continents

A B S T R A C T

The low spread of the global pandemic in Africa has raised concerns. Consequently, many commentators have misconstrued concerns suspecting weather, and immunity to be prime reasons. This study investigates the factors associated with the high and low spread of the SARS-CoV-2 (also known as COVID-19) and employs graphical Bayesian models to investigate feature interactions and causality. Through experimentation with the Bayesian framework, we propose that: (i) the proportion of people within the country population who test positive for SARS-CoV-2 and a country’s test capacity cause the rate of spread of the virus [i.e., $P(S|P)$ and $P(S|T)$] (ii) poverty gaps, welfare and freedom of the press directly cause the spread of the virus [i.e., $P(S|E)$, $P(S|W)$, and $P(S|R)$] (iii) Government effectiveness serves as a parent to poverty gaps and welfare [i.e., $P(E|G)$ and $P(W|G)$] and voice and accountability serve as a parent to freedom of the press [i.e., $P(R|V)$]. For the output, we “dichotomized” regions based on the “share of global infection rate” metric (SGIR) that implicitly accounts for a given region’s population, and we find that - out of two hundred and nineteen countries investigated, one hundred and twenty-seven have SGIR $\geq 1\%$, and the majority (44 out 58 - 75.86\%) of Africa countries (as of 12\textsuperscript{th} February 2021) have SGIR $< 1\%$. With Africa in the mirror, the study shows that only 2.2\% of the African population has been tested for SARS-CoV-2 and finds that the low proportion of population tested [i.e., $P(S|P)$] for SARS-CoV-2 is the cause of the low spread [i.e., cases reported] of SARS-CoV-2 in Africa. Similarly, the fragmented socioeconomic statuses [i.e., $P(S|E)$] among citizens leads to socioeconomic distancing, causing socio-class gaps between the rich and poor/average citizens, ensuring low interaction in social space, thus limiting the spread.

© 2022 The Author(s). Published by Elsevier B.V. on behalf of African Institute of Mathematical Sciences / Next Einstein Initiative.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

Introduction

The low spread of SARS-CoV-2 has raised some suspicious about Africa genuine status of the pandemic. And most persons expressing these concerns seems to have a valid argument; a consequence of the continent years of decade’s leadership

\footnote{Corresponding author.}

E-mail address: aojsoft@gmail.com (O.J. Akintande).

https://doi.org/10.1016/j.sciaf.2022.e01301
2468-2276© 2022 The Author(s). Published by Elsevier B.V. on behalf of African Institute of Mathematical Sciences / Next Einstein Initiative. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)
failures and bad image it represents for the continent on the global stage, such as Africa being associated with illness and poorness (see Hund and Mills [13], and Nott et al. [20] for example). For instance, the recent submission by a French scientist on Africa as a SARS-CoV-2, vaccine testing ground affirms how the world views Africa continents. "If I can be provocative, shouldn’t we be doing this study (SARS-CoV-2 vaccine testing) in Africa, where there are no masks, no treatments, and no resuscitation? A bit like as it is done elsewhere for some studies on AIDS. In prostitutes, we try things because we know that they are highly exposed and that they do not protect themselves" [14]. Truthfully, this statement is not only demeaning but puts Africa in a really depressing image.

Meanwhile, a survey result shows that substantial number of people believe that the weather and Africans immune system are responsible for the low cases and death resulting from SARS-CoV-2 in Africa. [3]. But little research regarding the spread of SARS-CoV-2 in Africa exists. Hence, there is a need to explore and investigate the cause of the low reported cases. Rendani and Marwala [38] examine the spreading rate of the pandemic in South Africa. They find that two change-points in the spreading rate of SARS-CoV-2 given the confirmed cases. The first change-point suggests the travel ban and the resultant containment of imported infections. And the second change point, a state-led mass screening and testing programme. Thus, the community level transmissions are slower than the original imported case. Similarly, Anderson et al. [5] suggests that under-reporting of cases of SARS-CoV-2 could be the reason for the low spread in Brazil at the initial stage. Therefore, it was possible that the rate of cases reported in Brazil is 0.07% and thus underestimates the real spreading of the pandemic in the country.

Besides the work of Rendani and Marwala [38] in South Africa, there is no other work (to the best of our knowledge) that investigated the cause of the low or high spread of the pandemic, in Africa. Consequently, this study investigates the factors associated with the high and low cases of the SARS-CoV-2 pandemic and employs graphical Bayesian models to establish feature interactions.

Related work

Factors justification – low cases of SARS-CoV-2 in Africa

Global actors (such as Bill & Melinda Gates) have expressed concern about the unexplainable causes of the low spread of SARS-CoV-2 in Africa. Other prominent commentators and World leaders have also expressed similar concerns. While Africa is not the only continent with minimal cases of SARS-CoV-2, Africa remains the hotspot of global health concerns due to its poor leadership system, poor health infrastructure, mismanagement, corruption, and health crisis ascribed to the continent over decades.

At the initial stage of the pandemic, sub-Saharan Africa reported the lowest cases of SARS-CoV-2. However, the number of cases began to rise in late March 2020, with confirmed cases increasing across the continent. This number may reflect a shortage of tests (John Hopkins Coronavirus Resource Center [30]). Testing for SARS-CoV-2 is crucial in understanding how many people have the disease and how fast and far it is spreading [22]. SARS-CoV-2 country’s testing capacity and provision of the service could significantly contribute to high and low case reporting of SARS-CoV-2 pandemic globally. Thus, countries with adequate testing capacity are likely to report more cases than those with little or no test capacity. The low spread of SARS-CoV-2 in Africa is due to the lack of testing capacity (WHO, [39]). With a population volume of over one billion people, Africa may need at least fifteen million test kits over the next three months (Africa CDC [33]). Africa countries still account for only 32,000 of the world’s three million-plus infections, while its testing programmes remain in disarray due to limited resources at their disposal ([24], and Dace [31]).

The impending constraints to mass testing in Africa are coordination and access to rural communities due to inaccessible road; densely populated urban areas; limitations on healthcare worker and facilities; distrust of healthcare personnel, and cyberbullying associated with the virus itself (Bruton afn Edward [28], and BBC [35]).

According to Africa (CDC [33]), Nigeria CDC claims to have the capacity to run 1500 tests per day but had only successfully conducted 152 tests as of March 2020. South Sudan (a country of twelve million people) has only eighteen tests conducted as of the first week of April 2020. Mass testing is a big problem in Africa due to rocketing demand for test kits and other medical equipment by Western nations that are more economically stable and buoyant to outbid others and purchase equipment at huge volume (Africa CDC [33]). Thus, one can validly attribute the low spread of cases of SARS-CoV-2 in Africa to lack of or low testing capability.

Another barrier to mass testing for SARS-CoV-2 in Africa are distrust, fear, and bullying attached to the disease and the burden of health response (Bruton and Edward [28]). Unlike in the developed nations where people freely come forward to be tested for disease, Africa communities are suspicious of government authorities, external actors, and health workers about the unknown consequences if, they are positive. Sufferers of Ebola and HIV/AIDS have bullied across Africa (Bruton and Edward [28]). Similarly, the assumption that having the coronavirus pandemic may cause a bully attack will usually result in individuals reluctant to undergo testing (even if available) amid fear of clannish and other reprisals from their communities (Bruton and Edward [28]). Consequently, promote distrust of health workers and disregard any offer to get tested for the virus at all. For instance (Aizenman [29]) reported that the Ebola outbreak in DR Congo was flaring up again. Fake coronavirus stories have led to widespread distrust of health officials as a protester in Côte d’Ivoire target testing centre fearing that the test kits posed a risk to nearby homes (BBC [36]).
Government insincerity, inefficiency, and compromise of freedom of the press may also constitute underreporting or low cases of the SARS-CoV-2 pandemic in Africa. State governors are struggling hard to accept the reality of the pandemic. Thus, resolve to cover up; attributes deaths to malaria, meningitis, hypertension, diabetes, and other illness for the increased mortality seen in days (The Guardian [32]). Measures such as lockdown of public places such as populated marketplaces, religious centres, joints, events centres, institutions, and the banning of social gatherings leading to societal physical distancing and communities’ inflows and interactions significantly reduce the spread of the virus (Rendani and Marwala, [38]). The increase in SARS-CoV-2 cases may be a consequence of the failed implementation of measures and adherence (the USA as a case study). For instance, Ghana records 271 fresh SARS-CoV-2 cases after ease lockdown (CGTN Africa [34]), and the USA adds 184,000 SARS-CoV-2 infections in a day (Schwart, [37]). Unsurprisingly, Kano and other Northern states in Nigeria could be the next epicentre of SARS-CoV-2 in Nigeria after Lagos due to lack of adherence to preventive measures, resulting from religious falsehood, misinformation, and illiteracy.

We have identified three major factors (among others) associated with the low spread of SARS-CoV-2 in Africa. These are health consciousness, economic situation, demographic and geographical factors, and government interference. We highlight the features in Table 1.

### Table 1

| S/N | Acronyms | Definition | Measurement |
|-----|----------|-----------|-------------|
| 1   | P        | The overall proportion of the population tested for SARS-CoV-2 in the country | Arbitrary Boolean/binary (high or low). High = ≥ 10% of total population, Low = < 10% of the total population. |
| 2   | T        | Test capacity of the country | Arbitrary Boolean/binary (high or low). High = ≥ 10 out of every 1000 person, Low = < 10 out of every 1000 person. |
| 3   | G        | Government effectiveness | Standard scores between 0 and 100. Arbitrary Re-coded as binary (Effective (EF) or Ineffective (IF)). EF = ≥ 50, IF = < 50. |
| 4   | W        | The welfare of citizen known as Gini | Standard scores between 0 and 100. Arbitrary Recoded as binary (Stable welfare (S W) or Unstable welfare (U W)). SW = countries ranked ≥ 40, UW = countries ranked < 40. |
| 5   | E        | Poverty gaps | Standard scores between 0 and 100. Arbitrary Recoded as binary (Significant (SG) or Insignificant gaps (IG)). SG = < 10, IG = ≥ 10. |
| 6   | V        | Voice and accountability | Standard scores between 0 and 5. Arbitrary Recoded as binary (Low or High). High = ≥ 2, Low = < 2. |
| 7   | R        | Freedom of Press | Standard scores between 0 and 5. Arbitrary Recoded as binary (Low or High). High = ≥ 2, Low = < 2. |

**Theoretical issues - Bayesian network**

Bayesian network (BN) uses Bayes probabilistic reasoning to illustrate intuition in the causal relationships between the features and outcomes of an event [12]. The network graphically illustrates the directed conditional dependencies (arcs) between stochastic variables (nodes) of the event [7]. Given a node, Y represents the spread of infection of a disease, such as in SARS-CoV-2, and node X depicts some likely causes, e.g. test capacity, the directed arc would point from X to Y and interpret as X causes or influences Y. X may also serve as the parent of Y [10]. A predefined set of mutually exclusive states and a node probability table (NP) (Jensen & Nielsen [40]) exists. The NP explains the marginal probability distribution (MPD) across the states. Given that observation is known. The effect would generate both forwards and backwards (i.e., bidirectional) to improve the posterior probability of any dependencies [7].

One major constraint of a BN is that it must be "acyclic" because the algorithm does not handle feedback loops. Thus, form an incomplete graph. That is, it assumes that variables-arc are independent. Fenton and Neil [10]. Bayesian Net is dependable in decision support because it provides evidence of the reasoning behind the conclusions. More so, the Bayes theorem gives a standard algorithm of the interactions between dependent and independent probabilities. It does this by leveraging the reasoning under uncertainty (such as unknown) and inclusion of domain knowledge and make predictions accurately in a sparse population [15,27].

There are several works on Bayesian Network in literature. For instance, Seixas et al. [23] proposed a network to understand dementia and Alzheimer’s diagnosis using predispositional factors, demographics, neurophysiological and symptoms, based on expert (domain) knowledge and supervised learning. They find that the Bayes Net outperforms other classifiers, such as decision tables, in diagnostic accuracy. Bayesian Net has been relevant not only in medicine but also in other areas (such as
agriculture). Bi and Chen [8] proposed a learning Bayes Net algorithm to diagnose crop diseases. The Bayes Net algorithm can work or be used in any study provided that the conditions of application are fulfilled.

**Empirical issues - SARS-CoV-2**

The looming pandemic opened the avenue for pieces of research in various areas of concern. Thus, predictive models and AI-driven approaches exist to help understand (e.g., the spreading and risk of infections) and alleviate the imminent effect of the pandemic on the healthcare systems [17,25,26].

Wynants et al. [26] conduct a systematic review of SARS-CoV-2 predictive learning to assess performance rate. And find that the majority of SARS-CoV-2 prediction algorithms perform between moderate and excellent. However, due to sparse data, the algorithms are biased and overfitting to an extent. To overcome this, Naïve Bayes could suffice since it performs better in the space of a sparse dataset. Similarly, Soltan et al. [25] finds that AI-driven algorithms for classification have high sensitivity and specificity and reduce sorting time from 2 days to 1 h.

The majority of the mathematical model for SARS-CoV-2 consider models inferring parameters of compartmental models such as susceptible-infectious-recovered (SIR) and the susceptible exposed-infectious-recovered (SEIR) that are widely used in infectious disease projections and employed in modeling SARS-CoV-2; for instance [1,2,4,6,16] among many others. Rendani and Marwala [38] perform Bayesian parameter inference of the susceptible-infectious-recovered (SIR) and the susceptible-exposed-infectious-recovered (SEIR) models using MCMC to estimate the mean baseline of spread. They find that the resulting parameter estimates fall in-line with the existing literature in-terms of mean baseline $R_0$ (before government action), mean incubation time and average infectious period. They submit that the initial government action (such as travel ban, school closures and stay-home orders) resulted in a mean decline of 80% in the spreading rate in South Africa.

In reported cases, government interventions (such as mass screening and testing campaigns) results in a second trajectory change point and driven by the widening of the population eligible for testing, from travelers’ (and their known contacts) to include the community who would have not afforded private lab testing which dominated the initial data. Their study affirms that testing and screening promote SARS-CoV-2 detection and cases leading to an increased case report of the virus (Rendani and Marwala, [38]). As shown in Fig. 4 (in appendix), South Africa has the highest SARS-CoV-2 infections in Africa. This action promotes the statistics of the cases in South Africa.

**Materials and methods**

Following Eq. (1), we assume that any country with a share of global infection rate (SGIR) greater than or equals one per cent (0.01) as having a high SARS-CoV-2 spread otherwise low spread of SARS-CoV-2. That is, 1% or more of the share of global infection by a given country is high. We adopt economics and governmental data from government data 360 of the World Bank at https://govdata360.worldbank.org/ and SARS-CoV-2 related data from Worldometers. To establish the SARS-CoV-2 high or low spread indicator, we compute the share of global infection given country as follows:

$$SGIR = \left( \frac{\text{Country total confirmed cases}}{\text{Country’s population}} \right) \times 100$$

(1)

**Naïve Bayes algorithm**

The goal of this study is to fit a predictive model or a classifier. That is, to classify countries to either low spread or high spread based on the training features. There are many classification algorithms in the literature (see [9,11,18]), and one of the simplest and most effective is the Naïve Bayes classifier [19,21].

Given that $Y$ (output) is a Boolean-valued random variable, and $X$ is a vector containing $n$ Boolean features $\rightarrow X = \{X_1, X_2, \ldots, X_m\}$, denoting ith feature of X. By applying Bayes rule given that $P(Y=\text{y}_i|X)$, we represent this as:

$$P(Y = y_i | X = x_m) = \frac{P(X = x_m | Y = y_i)P(Y = y_i)}{\sum_j P(X = x_m | Y = y_j)P(Y = y_j)}$$

(2)

$\rightarrow y_n$ denotes the nth possible value for Y, $x_m$ denotes the nth vector value for X.

Following the learning goal, the Bayes algorithm assumes that all $X_i$ is conditionally independent of each of the given $X_{m,s}$ to Y. Generally, when X contains m-features holding the conditional independence assumption true, we have

$$P(X_1, \ldots, X_m|Y) = \prod_{i=1}^{m} P(X_i|Y)$$

(3)

1 For the output, we “dichotomized” regions based on the “share of global infection rate” metric (SGIR) that implicitly accounts for a given region’s population. The SGIR computational formula for this study is arbitrary, it is the condition adopted essentially for classification or categorization of the countries and form the output variable.
Observe that when Y and the X’s are Boolean, there are only 2m parameters to define $P(X_i = x_{im} | Y = y_j)$ for the necessary i, j, m, compared to the $2(2^m - 1)$ parameters needed to fit $P(X|Y)$ in the absence of conditional independence. If Y is any discrete observation and X’s are also discrete, Eq. (1) is as,

$$P(Y = y_k | X = x_m) = \frac{P(Y = y_k) \prod_j P(X_i | Y = y_j)}{\sum_j P(Y = y_j) \prod_j P(X_i | Y = y_j)}$$

Model (4) is the base equation for the Naïve Bayes classifier. For any given new features X = {X_1, X_2, ..., X_m}, the algorithm shows how the probability that Y, given any features (X) and the distribution of P(Y) and P(X,Y) from the training sample. Thus, the Naïve Bayes for the most probable explanation (MPE) or maximum posteriori probability (MAP) is as:

$$Y \leftarrow \arg \max_{y_k} P(Y = y_k) \prod_i P(X_i | Y = y_k),$$

simplified as

$$Y \leftarrow \arg \max_{y_k} P(Y = y_k) \prod_i P(X_i | Y = y_k).$$

Conceptual framework of the Naïve Bayes algorithm

For this study, we identify seven features to classify the SGIR. As mentioned in section 1.1, They are health consciousness (test capacity), economic situation/variables (poverty gap and Gini (welfare)), demographic and geographical factors (proportion of population tested), and government interference (government effectiveness, the voice of accountability, and freedom of the press). See Table 1 for feature definitions.

Fig. 1 presents the feature interaction or relationship to the output (S). We propose that T and P directly cause S, i.e., P(S|T) and P(S|P). E and W have the same parent, ‘G’. Thus, P(S|E, G) and P(S|W, G). R has V as a parent, P(S|R, V). And we observe a causal loop of P(S|G) and P(S|V).

Therefore, the class variable (S) is the SGIR and evidence $X = \{V, P, T, G, R, E, W\}$. Following Fig. 1, the assumption is:

$$P(S, X) = P(S) \prod_{i=1}^{m} P(X_i | S)$$

Methodology

To establish the SARS-CoV-2 high or low spread indicator, we classify countries based on the share of global infection (SGIR) of the pandemic. We "dichotomized" the outcome as follows:

Focus variable = \begin{cases} 
\text{SGIR} \geq 1\% &= \text{SARS – CoV – 2 high spread} \\
\text{SGIR} < 1\% &= \text{SARS – CoV – 2 low spread} 
\end{cases}

Query:

- Marginal inference: $P(S = \text{Low spread}), P(S = \text{High spread})$. 

Fig. 1. Causal structure of the features to the output.
Table 2
SARS-CoV-2 test rate by continent.

| Continent | cases   | test    | pop         | Test rate (%) | Number of countries |
|-----------|---------|---------|-------------|---------------|---------------------|
| Africa    | 3,748,468 | 31,336,570 | 1,369,195,030 | 2.29          | 58                  |
| America   | 48,867,530 | 435,732,000 | 1,024,808,251 | 42.52         | 39                  |
| Asia      | 23,878,316 | 569,483,541 | 4,629,130,661 | 12.30         | 48                  |
| Europe    | 32,019,090 | 472,188,081 | 749,205,571  | 63.03         | 51                  |
| Oceania   | 49,613 | 15,186,109 | 33,547,448 | 45.27 | 11                  |

- Conditional inference: $P(S|P), P(S|T), P(S|E, G), P(S|W, G), P(S|R, V), P(S|G), P(S|V)$.

Therefore,

$$P (V, P, T, G, R, E, W, S) = P(V) P(P) P(G) P(R|V) P(E|G) P(W|G) P(S|P, T, R, E, W),$$

where the MAP inference,

$$Y \leftarrow \arg\max_{y_k} P(S = s_k) \prod_{i} P(P, T, R, E, W | S = s_k)$$ (8)

Causal assumptions

We propose the following possible causal relationship.

1. Test capacity directly causes a high/low spread of SARS-CoV-2.
2. The overall proportion of the population tested directly causes high/low spread of SARS-CoV-2.
3. Government effectiveness is a parent to (causes) poverty gaps and citizens welfare in the country. Therefore, fragmented socio-economic statuses among citizens lead to socio-economic closeness/distance. Thus, causing socio-class gaps between the rich and poor/average citizens. That is, there are huge gaps between the rich and poor/average citizens.
4. Welfare gaps and poverty gaps are related. It also measures the socio-economic distancing or gaps among citizens. It implies that E and W have the same parents.
5. Voice and accountability cause freedom of press and freedom of press causes case reporting leading to the low or high spread of SARS-CoV-2.

Results and discussions

We examine the test rate or capacity so far across the continents and find that Africa has the lowest SARS-CoV-2 test rate (only 2.2% of the Africa population). Unexpectedly, Asia preceded Africa. These two continents have the largest population but have tested few proportions of their respective population so far. We present the summary in Table 2. Out of two hundred and nineteen countries, one hundred and twenty-seven have SGIR ≥ 1% or high SARS-CoV-2 cases. However, 44 out of 58 - [75.86]% of Africa countries; have SGIR < 1% or low SARS-CoV-2 spreads.

1 This ratio varies on all the continents except in Europe. Consequently, 50 out of 51 - [98.04]% of the countries in Europe have (SGIR ≥ 1%) high SARS-CoV-2 cases. However, the rate of infections (high or low) is similar in other continents, in the sense that substantial number of countries within each continent have a low spread (SGIR < 1%). As observed, in America, twelve (12) out of thirty-nine (39) - [33.77]% countries have low spreads. In Asia, twenty-five (25) out of forty-eight (48) - [52.08]% have fewer infections, and in Oceania, ten (10) out of eleven (11) - [90.91]% countries have low SARS-CoV-2 spread. Africa is probably making headlines because of how the world perceives it, as noted in Section 1.0.

To investigate the causal or the conditional probabilities of the features on the output. We find that the low proportion of the population tested for SARS-CoV-2 causes a fewer spread of the pandemic in 78.26 % of countries given a low infection rate. The higher proportion of the country's population tested causes higher infections in 86.61% of countries, given SGIR ≥ 1% (i.e., higher cases of the virus).

Low test capacity causes low spread in 39.13% of countries given SGIR < 1%, and high-test capacity cause high infection in 99.21% of countries given SGIR ≥ 1% (i.e., higher cases of the virus). It implies that a country may have a high-test capacity (60.87%) and still have a low spread. More so, it suggests that most countries with minimal infections have a low testing capacity.

2 The statistics in Table 2 are of SARS-CoV-2 cases reported as of 12 February 2021. The number of countries (column 6) implies the number of countries observed for this study as reported by Worldometers (https://www.worldometers.info/coronavirus/) and based on data availability. It is not necessarily the complete number of countries in the respective continent. There could be more!

3 The values printed are conditional probabilities for the experimented graphical interaction. It computed based on the causal structure of the features to the Output (Figure 1). For convenience, we use H[L for all the categories [H – High & L – Low]. Invariably, the features (arbitrary) categorization is in Table 1.
Expectantly, governmental variables should be bidirectional. For instance, Government ineffectiveness causes a low SARS-CoV-2 spread in 77.17% of countries, given that SGIR < 1%. Government effectiveness causes high SARS-CoV-2 cases in 57.48% of countries, given that SGIR ≥ 1%. This means that, if a government is “assumed” efficient (e.g., a country having a robust/dependable healthcare services), citizens may rely ‘wholly’ or have high confidence on the government. And thus, report and follow strict measures for treatments, leading to more case reporting. And if the government is “assumed” unreliable (e.g., a country not having a reliable healthcare system), citizens may depend less or have little or no confidence on the government, and thus self-medicate or adopt other in-house treatment means, leading to unreported cases. Hence, government effectiveness or ineffectiveness could cause high and low in various countries having a high or low spread of the virus. Voice and accountability serve as how seriously we should take information received about the country. Expectantly, this should also be bidirectional. A country (being) ranked low on voice and accountability cause the low SARS-CoV-2 - spread of the pandemic in 73.91% of countries given SGIR < 1%. And cause high SARS-CoV-2 cases in 53.54% of countries given SGIR ≥ 1%.

As expected, the close poverty gap (socio-economic closeness) causes a high SARS-CoV-2 - spread in 88.98% of countries, given that SGIR ≥ 1%. And the high poverty gap (socio-economic distancing) causes low SARS-CoV-2 cases in 51.09% of countries, given that SGIR < 1%. Similarly, welfare closeness causes a high SARS-CoV-2 - spread in 77.95% of countries, given SGIR ≥ 1%. And welfare distancing causes low SARS-CoV-2 cases in 47.83%, given that SGIR < 1%. Case reporting is associated with freedom of the press. Hence, it causes high SARS-CoV-2 infection in 74.02% of countries, given that SGIR ≥ 1%. And that it causes low SARS-CoV-2 cases in 50.0%, given that SGIR < 1%. Fig. 2 provides the result of the features casual interaction.

Concluding remarks

Contribution of study

Our study answers the puzzle behind the cause of the low or high spread of SARS-CoV-2. We find that; out of two hundred and nineteen countries investigated, one hundred and twenty-seven have SGIR ≥ 1%, and the majority of Africa countries have SGIR < 1%. With Africa in view, the cause of the low SARS-CoV-2 cases in Africa are:

(1) The low proportion of the population tested for SARS-CoV-2. As observed, only 2.2% of Africa 1.37 billion population is tested for SARS-CoV-2 as of February 2021 (Table 2).
(2) Government ineffectiveness in Africa significantly contributes to the low SARS-CoV-2 cases in Africa. Most importantly, government ineffectiveness contributes to high poverty gaps and poor welfare of citizens in Africa. Essentially, the
Table 3
Feature importance.

| Acronyms | Feature                        | Importance |
|----------|--------------------------------|------------|
| P [tp]   | Prop. of the pop. tested       | 100.00     |
| E [pov]  | Poverty gaps                   | 44.59      |
| T [tc]   | Test capacity                  | 42.65      |
| G [govef]| Government effectiveness      | 38.30      |
| V [voice]| Voice and accountability      | 28.12      |
| W [welf] | Welfare                        | 16.56      |
| R [press]| Freedom of press              | 0.00       |

Table 3 is the numerical value plotted in Fig 3. Referencing (as in Eq. (1)) the SGIR < 1%, the 100% implies, it applies to all the ninety-two countries observed. Thus, the proportion of the population tested for SARS-CoV-2 is responsible 100% for the low SARS-CoV-2 cases. Poverty gaps cause low SARS-CoV-2 among 44.59% [41 out of the 92] of the countries. Etc.

Fig. 3. Order of importance of features.

sparse interactions (in social space) between the rich and poor in Africa limits the spread of the virus in Africa. Thus, the level of rich and poor interaction is ‘nearly’ non-existent.

(3) Another factor is the country’s SARS-CoV-2 test capacity. It supports the WHO [39] submission that the low spread of SARS-CoV-2 in Africa is due to low SARS-CoV2 testing capability.

Conclusion

The study did not consider some features such as lockdown, travel ban and other governmental interventions. These features may play some significant role in the causes of low or high spread (Rendani and Marwala [38]). Thus, the predictive performance of this study learning classifier stands at [Accuracy = 0.8121; Kappa = 0.6151]. Table 3 and Fig. 3 gives the summary and order of the important features.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
Appendix

Fig. 4

Fig. 4. SARS-CoV-2 spread in four continents.

References

[1] M. Adeniyi, M. Ekum, I. Ogunsanya, J. Akinyemi, S. Oke, M Matadi, Dynamic model of COVID-19 disease with exploratory data analysis, Scientific Africa, 9. Elsevier, 2020.
[2] N. Ahmad, COVID-19 modeling in Saudi Arabia using the modified susceptible-exposed infectious-recovered (SEIR) model, Cureus 12 (9) (2020) e10452, doi:10.7759/cureus.10452.
[3] O.J. Akintande, O.E. Olubusoye, Dataset on how misinformation promotes the immune perception of COVID-19 pandemic in Africa, Data Brief, 2020 Elsevier.
[4] S.A. Alanazi, M. Kamruzzaman, M. Alruwaili, N. Alshammari, S.A. Alqahtani, A &Karime, Measuring and preventing COVID-19 using the SIR model and machine learning in smart health care, J. Healthc. Eng. (2020). Hindawi. https://downloads.hindawi.com/journals/jhe/2020/8857346.pdf.
[5] R.M. Anderson, T.D. Hollingsworth, R.F. Baggaley, R. Maddren, COVID-19 spread in the UK: the end of the beginning? Lancet (2020), doi:10.1016/S0140-6736(20)31689-5.
[6] D.K. Bagal, A. Rath, A. Barua, D. Patnaikd, Estimating the Parameters of the Susceptible Infected-Recovered Model of COVID-19 Cases in India During Lockdown Periods, Elsevier Public Health Emergency Collection, 2020.
[7] Butcher, R., & Fenton, N. (2020). Extending the range of symptoms in a Bayesian network for the predictive diagnosis of COVID-19, medRxiv. Accessed online February 15, 2021. Available at https://www.medrxiv.org/content/10.1101/2020.10.22.20217554v1.full.pdf.
[8] C. Bi, G. Chen, D. Li, Y. Liu, Y. Chen, Bayesian networks modeling for crop diseases, Computer and Computing Technologies in Agriculture IV. CCTA 2010. IFIP Advances in Information and Communication Technology, 344. Springer, Berlin, Heidelberg, 2011.
[9] R. Caruana, A. Niculescu-Mizil, An empirical comparison of supervised learning algorithms, in: Proceedings of the 23rd International Conference on Machine Learning, 2006.
O.J. Akintande, O.E. Olubusoye, O.S. Yaya et al.

Scientific African 17 (2022) e01301

[10] N.E. Fenton, M. Neil, Risk Assessment and Decision Analysis with Bayesian Networks, Chapman and Hall/CRC. Boca Raton, Florida, 2018.

[11] T. Hastie, R. Tibshirani, J.H. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction: With 200 Full-Colour Illustrations, Springer, New York, 2001 [Jerome H.JSBIN 0-387-95284-5. OCLC 46809224.]

[12] P. He, B. Akhgir, H. Arabnia, Counter cyber attacks by semantic networks, in: Emerging Trends in ICT Security, Morgan Kaufmann, Massachusetts, 2014, pp. 455–467.

[13] Hund, W., and Mills, C. (2016). Comparing black people to monkeys has a long, dark simian history. The conversation: academic rigor, journalistic flair. Available at https://theconversation.com/comparing-black-people-to-monkeys-has-a-long-dark-simian-history-55102. Accessed October 10, 2020.

[14] Jean-Paul Mira (2020), "Coronavirus: France racism row over doctors' Africa testing comments", interviewed reported by BBC News, 3 April 2020, https://www.bbc.com/news/world/europe-52157722.

[15] C.E. Kahn, L.M. Roberts, K.A. Shaffer, P. Haddaww, Construction of a Bayesian network for mammographic diagnosis of breast cancer, Comput. Biol. Med. 27 (1) (1997) 19–29.

[16] P. Kumari, H.P. Singh, S. Singh, SEIAQRDT model for the spread of novel coronavirus (COVID-19): a case study in India, Appl. Intell. (2020) Springer Link.

[17] C. Menzi, A.M. Valdes, M.B. Freidin, C.H. Sudre, L.H. Nguyen, D.A. Drew, et al., Real-time tracking of self-reported symptoms to predict potential COVID-19, Nat. Med. 26 (1) (2020) 1037–1040 [Online].

[18] McCallum, Andrew. (2019). Graphical models, lecture2: Bayesian network representation. Available at https://people.cs.umass.edu/~mccallum/courses/gm2011/02-bn-rep.pdf.

[19] Mitchell, T (2017). Generative and discriminative classifiers: Naïve Bayes and logistic regression, available at https://www.cs.cmu.edu/~tom/mlbook/NBayesLogReg.pdf, accessed February 16, 2021.

[20] J.C. Nott, G.R. Gliddon, S.G. Morton, L. Agassiz, W. Usher, H.S. Patterson, Type of Mankind, Lippincott, Grambo & Co, Americana, Harvard University Press, Philadelphia, 1854.

[21] O'Reilly (2019). Naive Bayes, available at https://www.oreilly.com/library/view/data-algorithms/9781491906170/ch14.html#naive_bayes_training_phase, accessed, February 16, 2020.

[22] Patel, O., and Kalebi, A. (2020). The challenges facing Africa countries testing for COVID-19, Pasha 60, available at https://theconversation.com/pasha-60-the-challenges-facing-african-countries-testing-for-covid-19-135442, accessed April 22, 2020.

[23] P.L. Sexas, B. Zdrozney, J. Lale, A. Conic, D.C. Muchaluat Saade, A Bayesian network decision model for supporting the diagnosis of dementia, Comput. Biol. Med. 51 (1) (2014) 140–158.

[24] South China Morning Post –SCMP (2020). Africa's coronavirus cases double in 2 weeks but do low rates of testing hide an even bigger crisis, available at https://www.scmp.com/news/world/africa/article/3081965/africas-coronavirus-cases-double-2-weeks-do-low-rates-testing, accessed April 30, 2020.

[25] D. Soltan, A. Kouchaki, S. Zhu, T. Kiyashe, D. Taylor, T. Hussain, Z.B. et al., (2020) Artificial intelligence-driven assessment of routinely collected healthcare data is an effective screening test for COVID-19 in patients presenting to hospital, medRxiv.

[26] L. Wynants, B.V. Calster, G. Collins, R.D. Riley, G. Heine, E. Schuit, et al., Prediction models for diagnosis and prognosis of COVID-19: a systematic review and critical appraisal, BMJ 369 (1) (2020) 1037–1040 [Online].

[27] H.T. Zheng, BY. Kang, HG. Kim, P.C.G. da Costa, An ontology-based Bayesian network approach for representing uncertainty in clinical practice guidelines et al. (eds), Uncertainty Reasoning for the Semantic Web I, 5327, Springer, Berlin, Heidelberg, 2008 URSW 2006, URSW 2007, URSW 2005. Lecture Notes in Computer Science.

[28] B. Bruton, N. Edward, The barrier to mass testing for COVID-19 in Africa, Atlantic Council’s Africa Center, 2020, available at https://www.atlanticcouncil.org/blogs/africafocus/barriers-to-mass-testing-for-covid-19-in-africa/ accessed April 20, 2020.

[29] N. Azzenman, Ebola Flares up Amid Attacks on Health Worker in Congo, 2019, available at https://www.npr.org/2019/12/20/790077784/ebola-flares-up-amid-attacks-on-health-workers-in-congo, accessed April 20, 2020.

[30] Johns Hopkins Coronavirus Resource Center, Johns Hopkins Coronavirus Resource Center, 2020 [https://coronavirus.jhu.edu/BCC. (2020a). Kenyan nurse: “I was shunned over coronavirus fears”, available at https://www.bbc.com/news/world/africa-52630804, accessed May 14, 2020.

[31] H. Dace, Mass Community Testing is crucial to the Covid-19 response: Here’s how we get there, Tony Blair Institute for Global Change, Covid-19, 2020, available at https://institute.global/policy/mass-community-testing-crucial-covid-19-response-heres-how-we-get-there accessed April 22, 2020.

[32] A. The Guardian, “This in Nigerian City raise concerns over undetected Covid-19 outbreaks, 2020, available at https://www.theguardian.com/world/2020/apr/28/nigerian-authorities-deny-wave-of-deaths-is-due-to-coronavirus-19, accessed April 28, 2020.

[33] Africa CDC – Centre for Disease Control and Prevention, The barrier to mass testing for COVID-19 in Africa, 2020, available at https://www.atlanticcouncil.org/blogs/africafocus/barriers-to-mass-testing-for-covid-19-in-africa/, accessed April 20, 2020.

[34] CGTN Africa, Ghana’s total COVID-19 cases surpass 1500, 2020, available at https://africa.cgtn.com/2020/04/26/ghanas-total-covid-19-cases-surpass-1500/, accessed April 28, 2020.

[35] BBCKenyan nurse: “I was shunned over coronavirus fears, 2020a, available at https://www.bbc.com/news/world/africa-52630804 accessed May 14, 2020.

[36] Centre for Democracy & Development - CDDs African Blood Resistant to Coronavirus?, 2020, available at https://www.cddwestafrica.org/is-the-african-blood-resistant-to-coronavirus/, accessed March 31, 2020.

[37] M. Schwartz, U.S. Adds 184,000 Coronavirus Cases In 1 Day, With No End In Sight, 2020, Available at https://www.wfdd.org/story/us-adds-184000-coronavirus-cases-1-day-no-end-sight, accessed April 30, 2020.

[38] R. Mbuwah, T. Marwala, Bayesian inference of COVID-19 spreading rates in South Africa, PLoS ONE 15 (8) (2020) e0237126.

[39] WHOtransmission of SARS-CoV-2: implications for infection prevention precautions, 2020, available at https://www.who.int/news-room/commentaries/detail/transmission-of-sars-cov-2-implications-for-infection-prevention-precautions accessed August 2020.

[40] F.V. Jensen, T.D. Nielsen, Bayesian Networks and Decision Graphs. Information Science and Statistics series, 2nd ed., Springer-Verlag, New York, 2007 ISBN 978-0-387-68281-5.