Research Article

First Prototype of the Infectious Diseases Seeker (IDS) Software for Prompt Identification of Infectious Diseases

F. Baldassi[1,*], O. Cenciarelli[2], A. Malizia[3], P. Gaudio[1]

[1]Department of Industrial Engineering, University of Rome Tor Vergata, Rome, Italy
[2]International CBRNe Master Courses, University of Rome Tor Vergata, Rome, Italy
[3]Department of Biomedicine and Prevention, University of Rome Tor Vergata, Rome, Italy

ARTICLE INFO

ABSTRACT

The rapid detection of ongoing outbreak – and the identification of causative pathogen – is pivotal for the early recognition of public health threats. The emergence and re-emergence of infectious diseases are linked to several determinants, both human factors – such as population density, travel, and trade – and ecological factors – like climate change and agricultural practices. Several technologies are available for the rapid molecular identification of pathogens [e.g. real-time polymerase chain reaction (PCR)], and together with on line monitoring tools of infectious disease activity and behaviour, they contribute to the surveillance system for infectious diseases. Web-based surveillance tools, infectious diseases modelling and epidemic intelligence methods represent crucial components for timely outbreak detection and rapid risk assessment. The study aims to integrate the current prevention and control system with a prediction tool for infectious diseases, based on regression analysis, to support decision makers, health care workers, and first responders to quickly and properly recognise an outbreak. This study has the intention to develop an infectious disease regressive prediction tool working with an off-line database built with specific epidemiological parameters of a set of infectious diseases of high consequences. The tool has been developed as a first prototype of a software solution called Infectious Diseases Seeker (IDS) and it had been established in two main steps, the database building stage and the software implementation stage (MATLAB® environment). The IDS has been tested with the epidemiological data of three outbreaks occurred recently: severe acute respiratory syndrome epidemic in China (2002–2003), plague outbreak in Madagascar (2017) and the Ebola virus disease outbreak in the Democratic Republic of Congo (2018). The outcomes are promising and they reveal that the software has been able to recognize and characterize these outbreaks. The future perspective about this software regards the developing of that tool as a useful and user-friendly predictive tool appropriate for first responders, health care workers, and public health decision makers to help them in predicting, assessing and contrasting outbreaks.

1. INTRODUCTION

Infectious diseases represent a serious risk for public health; their spread can result in a large impact on people and populations [1]. Infectious diseases are still a fact of modern life: emerging and re-emerging diseases (e.g., Ebola Virus Disease - EVD) cause major problems across the world to people, society, and national and international economies [2]. In addition to the concerns linked to naturally occurring diseases, the potential intentional release of Biological Agents (BAs) with terrorist or criminal intent is very real. Thanks to greatly expanded trade and travel, infectious diseases – either naturally occurring or caused by accidental or intentional release of pathogens, can spread rapidly, resulting in a potentially significant impact on life, major economic crises, and political instability [3].

Information systems play a central role in developing a comprehensive and effective approach to prevent, detect, respond to, and manage infectious disease outbreaks [4,5]. Currently, a large amount of epidemiological data is being collected by laboratories, public health and health care providers, and government agencies at local, regional, national, and international levels [2]. Furthermore, many agencies and institutions have developed information systems to access, analyse, and report outbreaks. As an example, the Center for Disease Control and Prevention (CDC) in the U.S.A. has developed web-based reporting systems for health departments.

In addition to infectious disease-related data sources, the research and public health communities have developed a wide array of analytical and statistical models targeted at analysing diseases data for surveillance and outbreak prediction purposes. Some examples of these outbreaks are:

- the recent EVD outbreak in the Democratic Republic of Congo (DRC), started in 2018;

- the re-emergence of Chikungunya virus in Africa, Europe, and Asia in 2018;

- the emergence of Middle East respiratory syndrome Coronavirus (CoV) in the Middle East in 2012;

- the worldwide H1N1 influenza pandemic in 2009;

- the outbreak of Severe Acute Respiratory Syndrome (SARS) in China in 2003.

All of these recent outbreaks highlight the importance of large-scale infectious diseases surveillance networks. The emergence of infectious diseases is associated with human (i.e., population density, travel, trade, changes in land use) and environmental determinants, and with the combination of them [6–9]. The improved ability to study infectious disease dynamics has made possible – thanks to innovative mathematical models – to construct advanced tools for the early detection (and, in some cases, prediction) of outbreaks. For example, a better understanding of human mobility patterns has enabled simulation of the spread of SARS and novel influenza A/H1N1 diseases [10].

Global and regional surveillance of novel or re-emerging infectious diseases aims to rapidly detect changes in the incidence rate of endemic diseases and promptly recognise and characterise syndromes caused by previously unknown pathogens of epidemic potential. Once an infectious threat is identified, surveillance on the disease spread is paramount to apply and assess preventive and control measures [11].

The goal of infectious disease surveillance is to help in reducing the incidence and prevalence of infectious diseases by providing relevant public health information and knowledge to public health professionals, health care professionals, and decision makers, to promote actions that can result in the timely prevention and control of infectious diseases [11].

In this framework, this study aims to introduce and develop an infectious disease prediction tool – working with an off-line database – built with specific epidemiological parameters of a set of infectious diseases of high consequences. This application would be a useful and user-friendly predictive tool appropriate for first responders, health care workers, and public health decision makers to help them in predicting, assessing, and contrasting outbreaks.

2. MATERIALS AND METHODS

The tool has been developed according to the two steps described below: the database building stage and the software implementation stage.

2.1. Database Building

The key to this first software prototype being useful and workable in practice is the quality and reliability of the database on which the predictions are based. To test this prototype, the database has been populated with the relevant data of 35 pathogens and the relative infectious diseases they cause. The list of 35 pathogens and related diseases is reported in Annex 1. Investigating the current landscape of publicly available data on infectious disease outbreaks, the database is built using data reported through several sources: World Health Organization (WHO)/United Nations (UN) [12], U.S. CDC [13], European Centre for Disease Prevention and Control (ECDC) [14], Global Burden of Disease [15], Program for Monitoring Emerging Disease [16] and scientific literature [17–47]. The period of the data

collected from these different sources is from January to December 2019.

At this level of development, the list is not fully exhaustive and it does not include pathogens that cause rare infectious diseases. The list has been built up selecting different categories of pathogens (viruses, bacteria, and parasites) to cover as much as possible the most significant infectious diseases of high consequences.

The epidemiological parameters selection for each pathogen descends from a deep consultation of some of the major scientific papers existing in the literature and other correlated sources (e.g., WHO, CDC, or ECDC official websites, reports and publications). The choice of this kind of collection is to describe as better as possible the main characteristics of each agent and its related disease; in fact, the goal is to define unequivocally with refined parameters each pathogen that has been taken into account to help the software in its searching and prediction process.

Table 1 shows an example of a database string. Principally, it is possible to consult the epidemiological parameters and other significant data of the agent (e.g., Ebola virus) and its related disease (Ebola Haemorrhagic Fever - EHF) that have been considered. In particular:

- the agent name and related disease;

- geographical distribution;

- the signs and symptoms, which are, respectively, any objective and subjective evidence of the disease;

- the vector, that is, any carrier able to transmit a pathogen into a living organism; most carriers regarded as vectors are living organisms, but they could be inanimate media of infection such as dust particles;

- the age group, which is defined as the age susceptibility for a specific agent and the related disease;

- the gender, male or female, that is more susceptible to a specific disease;

- the transmission route, which is, the pathway of causative agents from a source of infection to a susceptible host; regarding this parameter, it has been also reported an additional column called "*the transmission routes more information*" where detailed data on the transmission route are reported;

- the reservoir, which is a living host inside of which a pathogen survives, often (though not always) without causing disease for the reservoir itself;

- the Case Fatality Rate (CFR), which is, the fraction of deaths per case and it defines the fraction of deaths over the entire period of infection, in percentage;

- the transmission rate, which is the rate at which infectious cases cause secondary or new cases in a population with susceptible individuals; it is a constant rate and it has units of inverse time (day^{-1});

- the incubation rate, which is the time elapsed between exposure to a pathogenic organism and when symptoms and signs appear; it has units of inverse time (day^{-1});

- the recovery rate, which is the rate of individuals who recover or die, leaves the infected class at a constant per capita probability per unit of time (day^{-1});

F. Baldassi et al. / Journal of Epidemiology and Global Health. In Press

3

Table 1 | Example of a database string and the epidemiological parameters reported. A 95% confidence interval for all of the detailed numeric parameters has been considered

Agent	Disease	Geographical distribution	Signs and symptoms	Vector/Other way	Age group	Gender	Transmission routes
Ebola virus	Ebola hemorrhagic fever	Central Africa, Western Africa	Fever, severe headache, muscle pain, weakness, fatigue, diarrhoea, vomiting, abdominal (stomach) pain, lack of appetite, unexplained haemorrhage (bleeding or bruising)	No vector	All	Male, female	Blood, body fluids, zoonotic disease, sexual

Reservoir/host/source	CFR (decimals)	Transmission routes detailed information	Transmission rate (day^{-1})	Incubation rate (day^{-1})	Recovery rate (day^{-1})	Infectious mortality rate (day^{-1})
Human, bat, primate, chimpanzees, gorillas, fruit bats, monkeys, forest antelope, porcupines	0.5	Blood or body fluids (urine, saliva, sweat, faeces, vomit, breast milk, and semen), Objects (such as needles and syringes) contaminated with body fluids from a person sick with EVD or the body of a person who died from EVD; Infected fruit bats or nonhuman primates (such as apes and monkeys); Semen from a man who recovered from EVD (through oral, vaginal, or anal sex).	0.284	0.607	0.135	0.135

- the infectious mortality rate, which represents the rate at which infected people die per unit of time (day^{-1}); it is not to be confused with the CFR.

2.1.1. Assumptions and generalization

In the database, some data are absent, due to two main reasons: there are not data available from the scientific literature or other existing sources or – for a specific disease – these data do not make sense (i.e., *Legionellosis* is not a vector borne disease and the vector dataset is not available).

The detailed numeric parameters of the database (CFR, transmission rate, incubation rate, recovery rate, infectious mortality rate) have been selected in accordance with dedicated scientific literature that concern mathematical modelling of infectious diseases. In particular, in the case of CFR, the average data come from infectious disease studies and scientific literature on past outbreaks, and this parameter refers to untreated and unvaccinated individuals, and it is stated in decimals. In case of transmission rate, incubation rate, recovery rate, and infectious mortality rate, these rates are reported as the reciprocals (day^{-1}) of the respective periods (days). And, even if do not shown in Table 1, a 95% confidence interval for all the detailed numeric parameters has been considered.

For simplifying the choice, the successive database parameters have been divided as follows:

- the age group dataset in seven main groups: baby (0–6 years), child (7–12 years), teenager (13–19 years), young (20–35 years), adult (36–65 years), senior (66–99 years), all (0–99 years);
- the transmission routes dataset in 12 options: foodborne, waterborne, vectorborne, zoonosis, contaminated surface, person-to-person, sexual, mother-to-child, faecal-oral, aerosol, blood, and body fluids.

2.2. Software

The software has been named Infectious Disease Seeker (IDS) and it has been developed in the MATLAB™ environment. The first prototype of IDS software has been developed for a still working MATLAB® environmental, however, without requiring an internet connection to work. This is possible because IDS is loaded with a database that includes the epidemiological parameters of 35 infectious diseases (Table 1). Thus, the tool can be useful and applicable in many different situations in the field and remote areas where internet connection and other kinds of communication systems are not always available.

At this stage of development, the prototype is structured as a single user-friendly layout with five tabs, each of them characterized of a specific color (Figure 1).

2.2.1. Search tab

The "*Search*" tab or green tab (Figure 1A), which is the core of the tool, where the identification of an agent and the related disease reporting some indispensable epidemiological parameters

Figure 1 | Screenshots of four of the five software tabs. (A) Search tab (green tab), (B) disease analysis (blue tab), (C) database (orange tab), and (D) disease information (red tab).

is run. Eight essential parameters ("*Epidemiological parameters available*" section) have been selected because it has been considered that they are the most important factors that also a non-health worker can identify and recognize. The user is able to load these parameters as free text values following the instructions reported in the "*User guide*" tab or black tab (see Subsection 2.2.3 *Others tabs*).

The MATLAB® code developed for regression analysis (logistic regression analysis) permits to determinate if data loaded by users in the "*Epidemiological parameters available*" section are in the database datasets and then recognize the associated correspondences and the related accuracy ratio. As schematized in Figure 2, each time that one parameter that the user loads in the software matches with its related database dataset the software provides a value equal to 1; if not a value equal to 0.

The accuracy ratio is reported in percentage and it is calculated as follows:

$$\text{Accuracy ratio} = \frac{\text{Sum of values}}{\text{Number of all epidemiological parameters available}} \times 100$$

For example of Figure 2, the accuracy ratio is 75%.

Even at this level of development, applying this type of regression analysis, the tool is able to find the possible causative agent and consequently the related disease. Considering the actual phase of software development, the level of accuracy ratio accepted is over 35%.

2.2.2. Disease analysis tab

The "*Disease analysis*" tab or blue tab (Figure 1B), in which expert users can analyse and compare another specific epidemiological parameters (CFR, transmission rate, incubation rate, recovery rate, infectious mortality rate) of two or more diseases.

2.2.3. Others tabs

- The "*Database*" tab or orange tab (Figure 1C), in which the entire database has been loaded and users can consult it directly in the tool.

- The "*Disease information*" tab or red tab (Figure 1D), where the users can find relevant publications (i.e. for WHO and CDC, no internet connection required) and some website links with additional information (this function needs an internet connection).

- The "*User guide*" tab or black tab (this tab is not shown in Figure 1), containing the software user manual.

3. RESULTS

The tool has been tested with the epidemiological data of three outbreaks occurred recently: SARS epidemic in China (2002–2003), plague outbreak in Madagascar (2017), and the EVD outbreak in the DRC (2018). The authors have considered these outbreaks because they had the intention to select pathogens in terms of relevance, and different characteristics as transmission route, type of microorganism (virus, bacteria), and geographical distribution; once the causative agents have been selected, the authors chose the more recent outbreaks. The tool has been tested in two different stages of the epidemic: at the beginning, when there is a lack of data, and during a more advanced stage, when more accurate and detailed data are available. All the data used in these tests have been retrieved from reliable sources [48–55].

3.1. Testing Search Function

Taking into consideration the initial stages of the three outbreaks analysed [48–50], the search function has been tested loading the early data available (Table 2A) in a specific IDS section called "*Epidemiological parameters available*" (Figure 3). Once loaded the data and clicked on "*Search*" button, as shown in Figure 3, the
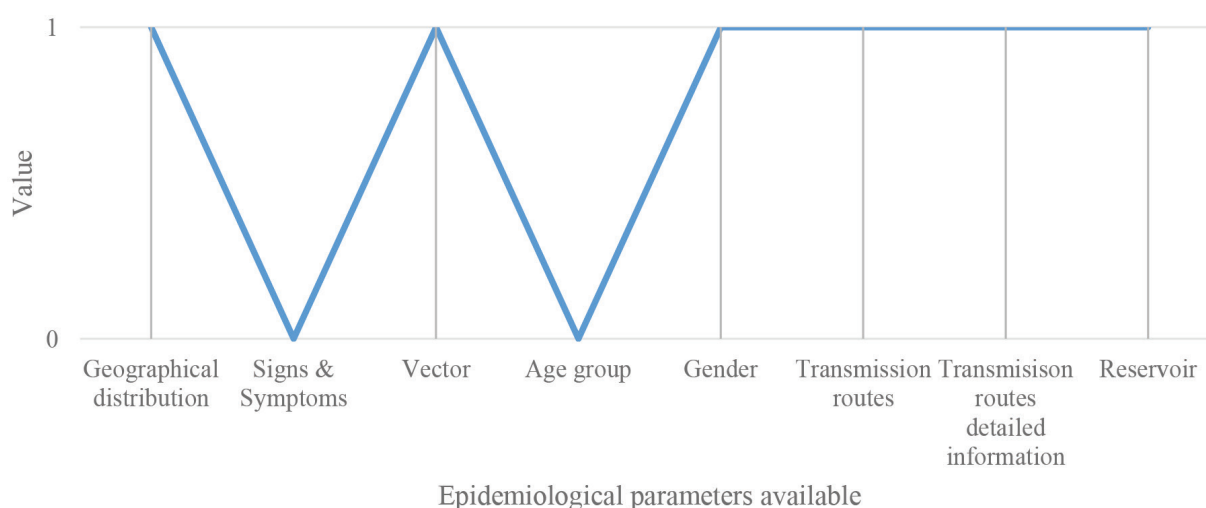

Figure 2 | Schematic representation for describing the values assignment process: value 1 means match, and value 0 means no match. In this specific example, there are six correspondences and two no correspondences.

Table 2 | Epidemiological parameters and other significant data of the three agents. (A) Early epidemiological parameters available for the three outbreak tested [48–50]. (B) The specific epidemiological parameters available, based on mathematical modelling of the three infectious diseases considered [51–55]. A 95% confidence interval for all the detailed numeric parameters has been considered

		SARS China (2003)	**Plague Madagascar (2017)**	**EHF Congo (2018)**
(A)	Geographical distribution	China	Madagascar	Democratic Republic of the Congo
	Signs and symptoms	High fever	Fever, nausea	High fever, haemorrhage
	Vector/Other way	//	//	//
	Age group	//	All	All
	Gender	Male, female	Male, female	Male, female
	Transmission routes	Person-to-person	//	Body fluids
	Transmission routes more information	Droplets, contact	Flea bites	//
	Reservoir	Human	Rodents	//
(B)	CFR (%)	0.17 (17)	0.09 (9)	0.54 (54)
	Transmission rate (day^{-1})	0.75	0.45	0.2
	Incubation rate (day^{-1})	0.083	0.26	0.17
	Recovery rate (day^{-1})	0.125	0.26	0.1
	Infectious mortality rate (day^{-1})	0.006	0.34	0.133

Figure 3 | Outputs of the "*Search*" tab. In section "*Possible agent recognized*", after the initial epidemiological parameters have been filled in, is possible to identify each time the highest accuracy ratio (%): SARS-associated coronavirus (A), *Yersinia pestis* (B), and Ebola virus (C).
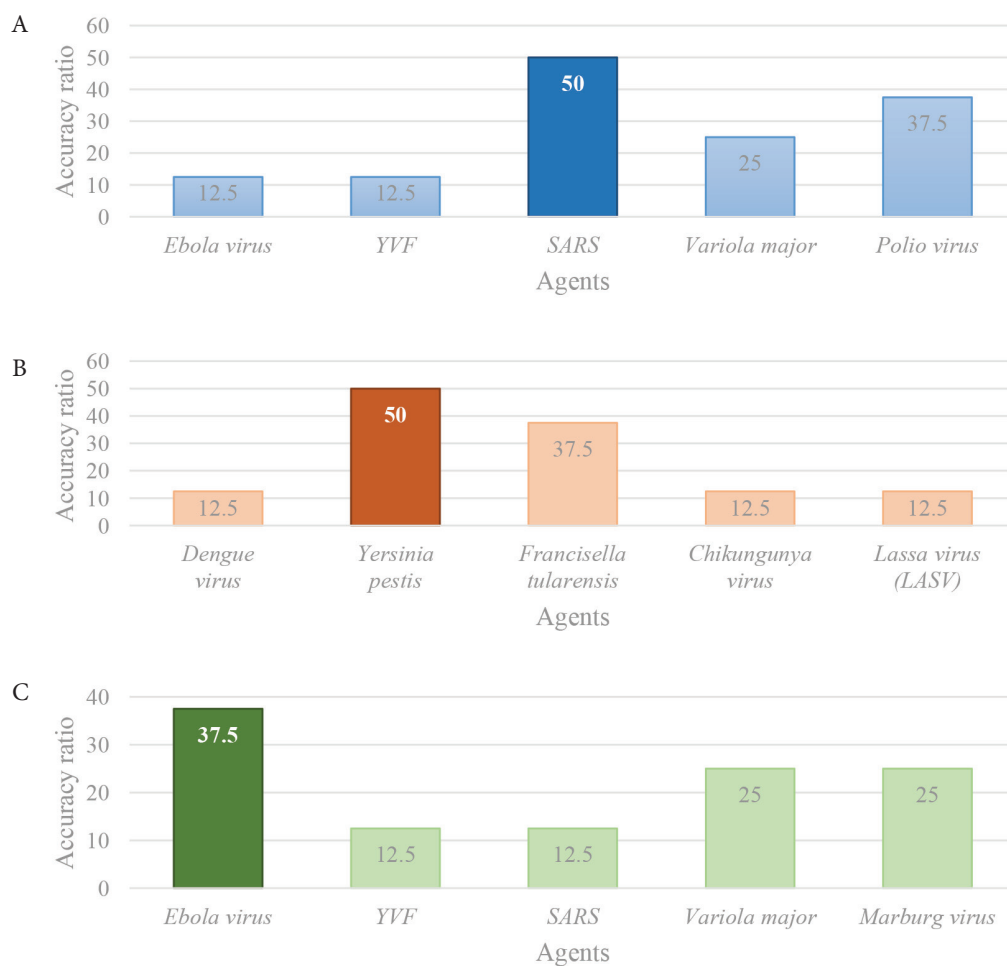
Figure 4 | Schematic representation of the comparison of the highest accuracy ratios. It is obtained comparing the IDS outcomes of the three outbreaks analysed and it distinguishes the more probable agents within the others: (A) the more probable agent is SARS (blue), (B) the more probable agent is *Y. pestis* (orange), and (C) the more probable agent is Ebola virus.

highest accuracy ratios obtained have been, respectively, SARS-associated CoV (Figure 3A), *Yersinia pestis* (Figure 3B), and Ebola virus (Figure 3C). The accuracy ratios are shown in the IDS table called "*Possible agent recognized*" and they are expressed in percentage (%).

It is important to underline that the search function has recognized all of the three agents analysed: it has discerned the more probable agents within the others. In fact, the tool was able to identify SARS-associated CoV and *Yersinia pestis* with an accuracy of 50%, and Ebola virus with an accuracy of 37.5% (Figure 4). It is essential to clarify that, as also shown in Table 2A, the data available at the beginning of these three outbreaks were incomplete, and the presence of these deficiencies has affected the accuracy ratios.

3.2. Testing Disease Analysis Function

Once identified the more probable agent(s) and the related disease(s) and detailed epidemiological data are available, it is possible to proceed with further analysis. In particular, using the disease analysis function ("*Disease analysis*" tab), it has been possible to identify if the disease recognised, behave, and spread with the same dynamic of the equivalent disease loaded in the database. As a result, the detailed epidemiological parameters available, based on the mathematical modelling of the three infectious diseases considered (Table 2B) [51–55], have been compared with the disease parameters of the database (Figure 5). In particular, Figure 5A shows the comparison between SARS disease data from the database (red spots) and the data come from the real epidemic occurred in China in 2003 (blue spots). The same in Figure 5B and 5C, where the comparison is, respectively, from database diseases and plague epidemic (2017) and EVD outbreak (2018).

The disease analysis function seems to have some little variance between the data of the database and the three diseases. In Figure 6, these variations have been graphically summarized. In particular, the scheme represents the correspondence for each of the five epidemiological parameters (CFR, transmission rate, incubation rate, recovery rate, and mortality rate) between the data from the database and the real cases considered in Table 2B (SARS in blue, Plague in red, and EVD in green). Three different conformity values have been assigned: 1 means that there is a $95 \pm 5\%$ correspondence between the database and the real case; 0.5 means that the correspondence is $50 \pm 5\%$; 0 means that there is not a correspondence between data. A schematic representation is shown in Figure 6.
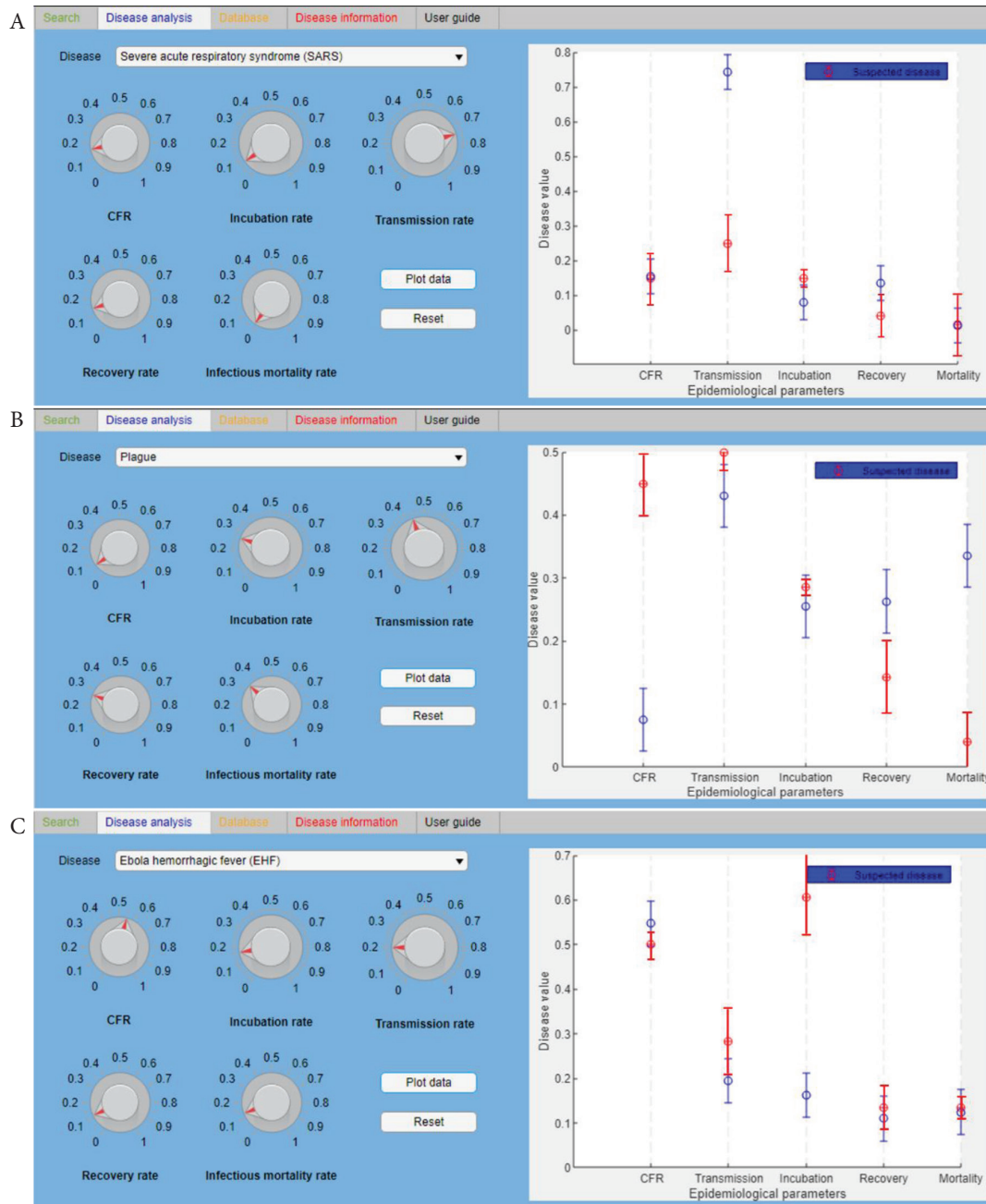
Figure 5 | The "Disease analysis" tab. It shows the comparison between the database (red spots) and the disease data (blue spots), respectively, of SARS (A), Plague (B), and EVD (C). The fixed standard deviation (SD) of blue spots is 0.05 (5%).

The accuracy ratio, that represents in this case the overall correspondence for each disease considered, is calculated as follows:

$$\text{Accuracy ratio} = \frac{\text{Sum of conformity values}}{\text{Number of epidemiological parameters}} \times 100$$

Consequently, the accuracy ratio is 80% for SARS and EVD, and 50% for Plague.

4. DISCUSSION

According to the results of these preliminary tests, the overall outcomes provided by the software are promising. The tool was stressed adequately testing three recent and relevant outbreaks of infectious diseases. In particular, the first step has been to test the capacities of the "*Search*" tab (green tab): the outcomes reveal that the IDS prototype has been able to identify all of the three agents and differentiate them with the rest of the agents loaded in the database (Figure 4).

The second and last step has been to test the "*Disease analysis*" tab (blue tab): the results show a reasonable correspondence between the database datasets and the three real cases considered. This lack of complete correspondence can be due to many reasons. It could be due to the difficulties to merge in the database all the data available in the scientific literature and consequently to find the best values that can represent uniquely the dynamics of a specific

Figure 6 | Parameters correspondence. Schematic representation of the correspondence for each of the five parameters (CFR, transmission rate, incubation rate, recovery rate, and mortality rate) between the data from the database and the real cases considered (SARS in blue, Plague in red, and EVD in green).

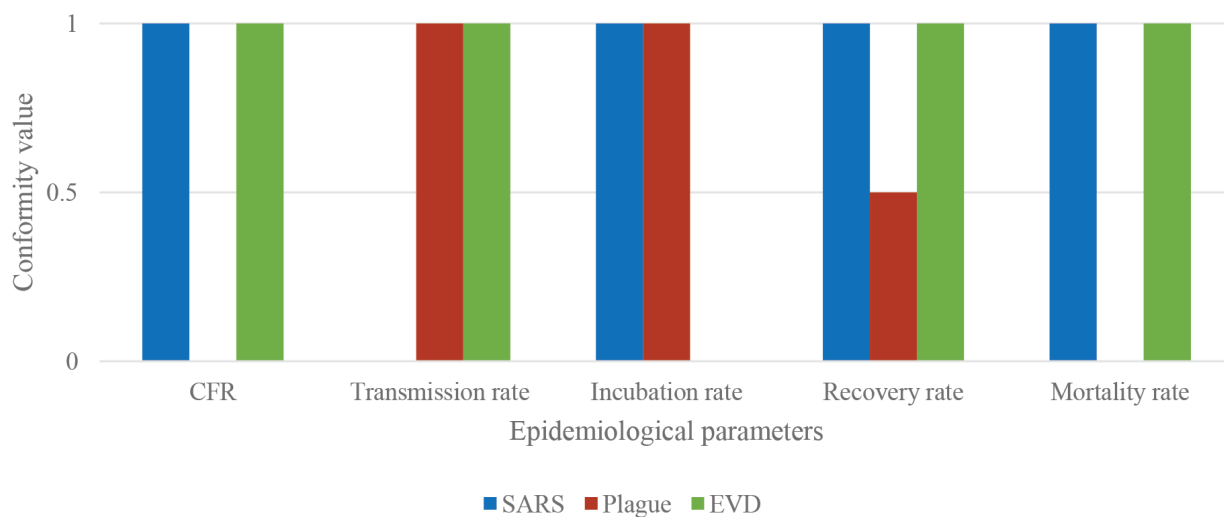disease in a population. As known, the dynamics of a disease can change deeply only, for instance, taking into account the disease spread in two different populations that live in a diverse geographical region [56].

Nowadays, there are important limitations of the current set of tools used in forecasting outbreaks and highlight existing and emerging technologies that have the potential to significantly enhance forecasting capabilities are substantial. The focus on forecasting for outbreak management, specifically the capacity to predict short term (i.e., days to weeks) trends of disease activity or incidence (i.e., the number and location of new cases) in an ongoing outbreak [11,57].

The aim of this project is to insert in this lack of capability and technology a tool such as this first IDS prototype that has revealed itself a useful software to early identify a possible epidemic or outbreak at this level of developing as well. To do this, the IDS takes advantage of the off-line database that represents a pivotal characteristic for working without an internet connection. For this reason and taking into account the outcomes of this study, this first prototype needs, before developing the beta version of IDS, to improve the dependability of the database gathering in it more pathogens, and enhancing and increasing the epidemiological parameters available for each agent in order to increase the precision and sensibility of the agent identification.

5. CONCLUSION

This study had the intention to introduce and develop an infectious disease prediction tool working with an off line database built with specific epidemiological parameters of a set of 35 infectious diseases of high consequences. The analytical results show that the prototype of IDS software can perform this kind of investigation. The future perspective about this software regards the developing of that tool as a useful and user-friendly predictive tool appropriate for first responders, health care workers, and public health decision makers

to help them in predicting, assessing and contrasting outbreaks. Furthermore, the intention is also to enhance the database reliability, crowding it with more pathogens (also rare agents) and extending the epidemiological parameters to permit to the software a better identification process. Then, the last intent is to create a standalone desktop app to run the application without passing or retrieving any arguments to or from MATLAB®. This standalone application will run on the most common operative systems (Windows, Linux®, and Mac) and it does not require a licensed copy of MATLAB®.

CONFLICTS OF INTEREST

The authors declare they have no conflicts of interest.

AUTHORS' CONTRIBUTION

FB study conceptualization and writing (review & editing) the manuscript, FB and OC data curation, formal analysis and writing (original draft), PG funding acquisition and project administration, AM and PG supervised the project, FB and OC formal analysis and writing (original draft) the manuscript.

FUNDING

REFERENCES

[1] McNeill WH. Plagues and peoples. New York: Doubleday; 1976.
[2] Pinner RW, Rebmann CA, Schuchat A, Hughes JM. Disease surveillance and the academic, clinical, and public health communities. Emerg Infect Dis 2003;9;781–7.

[3] Chang M, Glynn MK, Groseclose SL. Endemic, notifiable bioterrorism-related diseases, United States, 1992–1999. Emerg Infect Dis 2003;9;556–64.

[4] Damianos L, Ponte J, Wohlever S, Reeder F, Day D, Wilson G, Hirschman L. MiTAP for bio-security: a case study. AI Mag 2002;23;13–29.

[5] Buehler JW, Hopkins RS, Overhage JM, Sosin DM, Tong V, CDC Working Group. Framework for evaluating public health surveillance systems for early detection of outbreaks: recommendations from the CDC Working Group. MMWR Recomm Rep 2004;53;1–11.

[6] Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, Daszak P. Global trends in emerging infectious diseases. Nature 2008;451;990–3.

[7] Weiss RA, McMichael AJ. Social and environmental risk factors in the emergence of infectious diseases. Nat Med 2004;10;S70–S6.

[8] Lipkin WI. The changing face of pathogen discovery and surveillance. Nat Rev Microbiol 2013;11;133–41.

[9] Morse SS, Mazet JAK, Woolhouse M, Parrish CR, Carroll D, Karesh WB, et al. Prediction and prevention of the next pandemic zoonosis. Lancet 2012;380;1956–65.

[10] Hyman JM, LaForce T. Modeling the spread of influenza among cities. In: Banks HT, Castillo-Chavez C, editors. Bioterrorism: mathematical modeling applications in homeland security. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics; 2003, pp. 211–36.

[11] Christaki E. New technologies in predicting, preventing and controlling emerging infectious diseases. Virulence 2015;6;558–65.

[12] World Health Organization (WHO)/United Nations (UN). Available from: https://www.who.int/news-room/fact-sheets & https://www.un.org/en/sections/issues-depth/health/index.html.

[13] U.S. Centers for Disease Control and Prevention (CDC). Available from: https://www.cdc.gov/DiseasesConditions/.

[14] European Centre for Disease Prevention and Control (ECDC). Available from: https://www.ecdc.europa.eu/en/publications-data.

[15] Global Burden of Disease (GBD). Available from: http://www.healthdata.org/gbd/publications.

[16] Program for Monitoring Emerging Disease (ProMED). https://www.promedmail.org/.

[17] Cenciarelli O, Pietropaoli S, Malizia A, Carestia M, D'Amico F, Sassolini A, et al. Ebola virus disease 2013-2014 outbreak in West Africa: an analysis of the epidemic spread and response. Int J Microbiol 2015;2015;769121.

[18] Garske T, Cori A, Ariyarajah A, Blake IM, Dorigatti I, Eckmanns T, et al. Heterogeneities in the case fatality ratio in the West African Ebola outbreak 2013–2016. Philos Trans R Soc Lond B Biol Sci 2017;372;20160308.

[19] Martorano Raimundo S, Amaku M, Massad E. Equilibrium analysis of a yellow fever dynamical model with vaccination. Comput Math Methods Med 2015;2015;482091.

[20] Hui DSC, Chan MCH, Wu AK, Ng PC. Severe acute respiratory syndrome (SARS): epidemiology and clinical features. Postgrad Med J 2004;80;373–81.

[21] Mehedi M, Groseth A, Feldmann H, Ebihara H. Clinical aspects of Marburg hemorrhagic fever. Future Virol 2011;6;1091–106.

[22] Chua KB. Epidemiology, surveillance and control of Nipah virus infections in Malaysia. Malays J Pathol 2010;32;69–73.

[23] Iowa State University Center for Food Security and Public Health, "Hendra Virus Infection". Center for Food Security and Public Health Technical Factsheets.71; 2015. Available from: http://lib.dr.iastate.edu/cfsph_factsheets/71.

[24] Agusto FB, Bewick S, Fagan WF. Mathematical model of Zika virus with vertical transmission. Infect Dis Model 2017;2;244–67.

[25] Yun NE, Walker DH. Pathogenesis of Lassa fever. Viruses 2012;4;2031–48.

[26] National Institute for Communicable Diseases (NICD) - Healthcare Workers Guidelines on Rift Valley fever. Available from: https://www.nicd.ac.za/wp-content/uploads/2018/05/RVF-2018-Guidelines-for-HCWs.pdf (updated January 13, 2020).

[27] Bakach I, Braselton J. A survey of mathematical models of dengue fever. Comput Sci Syst Biol 2015;8;255–67.

[28] Chevalier V, Tran A, Durand B. Predictive modeling of West Nile virus transmission risk in the mediterranean basin: how far from landing? Int J Environ Res Public Health 2014;11;67–90.

[29] Al-Asuoad N, Rong L, Alaswad S, Shillor M. Mathematical model and simulations of MERS outbreak: predictions and implications for control measures, Biomath 2016;5;1612141.

[30] Lai S, Qin Y, Cowling BJ, Ren X, Wardrop NA, Gilbert M, et al. Global epidemiology of avian influenza A H5N1 virus infection in humans, 1997–2015: a systematic review of individual case data. Lancet Infect Dis 2016;16;e108–e18.

[31] Cohen JI. Epstein–Barr virus infection. N Engl J Med 2000;343;481–92.

[32] Bonthius DJ. Lymphocytic choriomeningitis virus: an under-recognized cause of neurologic disease in the fetus, child, and adult. Semin Pediatr Neurol 2012;19;89–95.

[33] Ganesan VK, Duan B, Reid SP. Chikungunya virus: pathophysiology, mechanism, and modeling. Viruses 2017;9;368.

[34] Lambert AJ, Martin DA, Lanciotti RS. Detection of North American eastern and western equine encephalitis viruses by nucleic acid amplification assays. J Clin Microbiol 2003;41;379–85.

[35] Pisano MB, Oria G, Beskow G, Aguilar J, Konigheim B, Cacace ML, et al. Venezuelan equine encephalitis viruses (VEEV) in Argentina: serological evidence of human infection. PLoS Negl Trop Dis 2013;7;e2551.

[36] MacLachlan JH, Cowie BC. Hepatitis B virus epidemiology. Cold Spring Harb Perspect Med 2015;5;a021410.

[37] Saito T, Ueno Y. Transmission of hepatitis C virus: self-limiting hepatitis or chronic hepatitis? World J Gastroenterol 2013;19;6957–61.

[38] Mueller NH, Gilden DH, Cohrs RJ, Mahalingam R, Nagel MA. Varicella zoster virus infection: clinical features, molecular pathogenesis of disease, and latency. Neurol Clin 2008;26;675–97.

[39] Silk BJ, Foltz JL, Ngamsnga K, Brown E, Muñoz MG, Hampton LM, et al. Legionnaires' disease case-finding algorithm, attack rates, and risk factors during a residential outbreak among older adults: an environmental and cohort study. BMC Infect Dis 2013;13;291.

[40] Respicio-Kingry LB, Yockey BM, Acayo S, Kaggwa J, Apangu T, Kugeler KJ, et al. Two distinct yersinia pestis populations causing plague among humans in the West Nile region of Uganda. PLoS Negl Trop Dis 2016;10;e0004360.

[41] Gürtler L, Bauerfeind U, Blümel J, Burger R, Drosten C, Gröner A, et al. Coxiella burnetii – pathogenic agent of Q (Query) fever. Transfus Med Hemother 2014;41;60–72.

[42] Sulis G, Roggi A, Matteelli A, Raviglione MC. Tuberculosis: epidemiology and control. Mediterr J Hematol Infect Dis 2014;6;e2014070.

[43] Gürcan Ş. Epidemiology of tularemia. Balkan Med J 2014;31;3–10.

[44] Chiyaka ET, Muyendesi T, Nyamugure P, Mutasa FK. Theoretical assessment of the transmission dynamics of leprosy. Appl Math 2013;4;387–401.

[45] Karkey A, Thompson CN, Tran Vu Thieu N, Dongol S, Phuong TLT, Vinh PV, et al. Differential epidemiology of *Salmonella* Typhi and Paratyphi A in Kathmandu, Nepal: a matched case control investigation in a highly endemic enteric fever setting. PLoS Negl Trop Dis 2013;7;e2391.

[46] Costa F, Hagan JE, Calcagno J, Kane M, Torgerson P, Martinez-Silveira MS, et al. Global morbidity and mortality of leptospirosis: a systematic review. PLoS Negl Trop Dis 2015;9;e0003898.

[47] Rouphael NG, Stephens DS. Neisseria meningitidis: biology, microbiology, and epidemiology. Methods Mol Biol 2012;799;1–20.

[48] Xu R, He J, Evans MR, Peng G, Field HE, Yu D, et al. Epidemiologic clues to SARS origin in China. Emerg Infect Dis 2004;10;1031–7.

[49] Nguyen VK, Parra-Rojas C, Hernandez-Vargas EA. The 2017 plague outbreak in Madagascar: data descriptions and epidemic modelling. Epidemics 2018;25;20–5.

[50] Ebola Outbreak Epidemiology Team. Outbreak of Ebola virus disease in the Democratic Republic of the Congo, April-May, 2018: an epidemiological study. Lancet 2018;392;213–21.

[51] Wang W, Ruan S. Simulating the SARS outbreak in Beijing with limited data. J Theor Biol 2004;227;369–79.

[52] Zhou Y, Ma Z, Brauer F. A discrete epidemic model for SARS transmission and control in China. Math Comput Model 2004;40;1491–506.

[53] Tsuzuki S, Lee H, Miura F, Chan YH, Jung S, Akhmetzhanov AR, et al. Dynamics of the pneumonic plague epidemic in Madagascar, August to October 2017. Euro Surveill 2017;22;17-00710.

[54] Meakin SR, Tildesley MJ, Davis E, Keeling MJ. A metapopulation model for the 2018 Ebola virus disease outbreak in Equateur province in the Democratic Republic of the Congo. bioRxiv 2018;465062.

[55] Sule A, Lawal J. Mathematical modeling and optimal control of Ebola virus disease (EVD). Annu Res Rev Biol 2018;22;1–11.

[56] Erraguntla M, Zapletal J, Lawley M. Framework for infectious disease analysis: a comprehensive and integrative multi-modeling approach to disease prediction and management. Health Informatics J 2019;25;1170–87.

[57] George DB, Taylor W, Shaman J, Rivers C, Paul B, O'Toole T, et al. Technology to advance infectious disease forecasting for outbreak management. Nat Commun 2019;10;3932.

ANNEXURE

Annex 1 | List of the 35 pathogens and relative diseases

Agent	Disease
Ebola virus	Ebola Hemorrhagic Fever (EHF)
Yellow fever virus	Yellow fever
SARS-associated Coronavirus (SARS-CoV)	Severe acute respiratory syndrome (SARS)
Variola major	Smallpox
Polio virus	Poliomyelitis
Marburg virus	Viral Hemorrhagic Fever (VHF)
Nipah virus	Nipah virus infection
Hendra virus	Hendra virus infection
Zika virus	Zika virus infection
Lassa Virus (LASV)	Lassa Hemorrhagic Fever (LHF)
Rift valley virus	Rift Valley Fever (RVF)
Dengue virus	Dengue fever
West nile virus	West Nile Disease (WND)
Middle East Respiratory Syndrome Coronavirus (MERS-CoV)	Middle East Respiratory Syndrome (MERS)
Avian (bird) influenza (flu) type A virus	Influenza
Epstein–Barr virus (EBV)	Infectious mononucleosis
Lymphocytic Choriomeningitis Virus (LCMV)	Lymphocytic choriomeningitis virus infection
Chikungunya virus	Chikungunya fever
Eastern Equine Encephalitis Virus (EEEV)	Eastern Equine Encephalitis (EEE)
Western Equine Encephalitis Virus (WEEV)	Western Equine Encephalitis (WEE)
Venezuelan Equine Encephalitis Virus (VEEV)	Venezuelan Equine Encephalitis or Encephalomyelitis (VEE)
Hepatitis A virus (HAV)	Hepatitis A, liver disease
Hepatitis B virus (HBV)	Hepatitis B, liver disease
Hepatitis C virus (HCV)	Hepatitis C, liver disease
Varicella-zoster virus (VZV)	Chickenpox or varicella
Legionella pneumophila	Legionellosis, Pontiac fever
Yersinia pestis	Plague
Coxiella burnetii	Q fever
Mycobacterium tuberculosis	Tuberculosis (TB)
Francisella tularensis	Tularemia
Mycobacterium leprae	Leprosy, Hansen's disease
Salmonella Typhi	Typhi fever
Leptospira spp	Leptospirosis
Listeria monocytogenes	Listeriosis
Neisseria meningitidis	Bacterial meningitis