



# DOT: a crowdsourcing Mobile application for disease outbreak detection and surveillance in Mauritius

Kavi Khedo<sup>1</sup> · Shakuntala Baichoo<sup>1</sup> · Soulakshmee Devi Nagowah<sup>2</sup>  · Zahra Mungloo-Dilmohamud<sup>1</sup> · Zarine Cadarsaib<sup>2</sup> · Sudha Cheerkoot-Jalim<sup>3</sup> · Leckraj Nagowah<sup>2</sup> · Lownish Sookha<sup>1</sup>

Received: 28 December 2019 / Accepted: 29 June 2020 / Published online: 3 July 2020  
© IUPESM and Springer-Verlag GmbH Germany, part of Springer Nature 2020

## Abstract

Early detection of disease outbreaks is crucial and even small improvements in detection can significantly impact on a country's public health. In this work, we investigate the use of a crowdsourcing application and a real-time disease outbreak surveillance system for five diseases; Influenza, Gastroenteritis, Upper Respiratory Tract Infection (URTI), Scabies and Conjunctivitis, that are closely monitored in Mauritius. We also analyze and correlate the collected data with past statistics. A crowdsourcing mobile application known as Disease Outbreak Tracker (*DOT*) was implemented and made public. A real-time disease surveillance system using the Early Aberration Reporting System algorithm (EARS) for analysis of the collected data was also implemented. The collected data were correlated to historical data for 2017. Data were successfully collected and plotted on a daily basis. The results show that a few cases of Flu and Scabies were reported in some districts. The EARS methods C1, C2 and C3 also depicted spikes above the set threshold on some days. The study covers data collected over a period of one month. Once symptoms data were collected using *DOT*, probabilistic methods were used to find the disease or diseases that the user was suffering from. The data were further processed to find the extent of the disease outbreak district-wise, per disease. These data were represented graphically for a rapid understanding of the situation in each district. Our findings concur with existing data for the same period for previous years showing that the crowdsourcing application can aid in the detection of disease outbreaks.

**Keywords** Disease outbreaks · Communicable diseases · Crowdsourcing · Mobile applications

## 1 Introduction

Despite breakthroughs in the field of medicine to prevent and cure different types of diseases, the fear of outbreaks is ever present. This is due to several factors such as the dynamic and volatile nature of viruses and bacteria or the harsh and abrupt changes in weather [1]. To be able to detect or fight these diseases, surveillance is of utmost importance [2]. Disease surveillance is the ongoing scrutiny and monitoring of clinical syndromes that have a significant impact on medical resource

allocation and health policy [3]. Disease surveillance ensures that prompt intervention is done for the control of diseases in order to contain the outbreak and minimize any potential harm [2]. It comprises the continuous methodical collection, organization, analysis and interpretation of disease data [2] and often involves big data [4]. With the COVID-19 pandemic affecting most countries around the world, disease surveillance systems are gaining momentum. Such systems aim at rapidly identifying possible cases of infectious diseases, which can help health practitioners and authorities to take timely decisions. Now more than ever, simple and easily accessible disease surveillance systems are required which can also reach countries with high resource constraints.

Surveillance systems can be categorized into disease-specific surveillance systems, syndromic surveillance systems and event-based surveillance systems [2]. The traditional surveillance systems are disease-specific. Data standardization is required for such systems. Reporting from traditional systems is normally not open to the public due to the fact that individuals might not be able to correctly identify the disease they are

✉ Soulakshmee Devi Nagowah  
s.ghurbhurrin@uom.ac.mu

<sup>1</sup> Department of Digital Technologies, FoICDT, University of Mauritius, Réduit, Mauritius

<sup>2</sup> Department of Software and Information Systems, FoICDT, University of Mauritius, Réduit, Mauritius

<sup>3</sup> Department of Information and Communication Technologies, FoICDT, University of Mauritius, Réduit, Mauritius

suffering from. Syndromic surveillance, on the other hand, is more real-time than the traditional surveillance and requires no laboratory confirmation. Reporting from syndromic surveillance systems can be deployed at large and certain methods can be used to correlate the syndromes to a corresponding disease. Event-based surveillance consists of “*the real-time or near-real time scanning, collection, and analysis of unstructured information from diverse Internet sources (such as news or online discussion platforms) for detecting potential or confirmed health hazards occurring worldwide from reports and rumours*” [2].

With the recent technological advances leading to a decrease in the cost of mobile phones and easy access to social media, event-based surveillance systems are on the rise [5, 6]. People are increasingly using social networks such as Twitter, Facebook and LinkedIn amongst others to share their experiences and opinions on a variety of different things, including healthcare-related topics like health conditions, symptoms, treatments and side-effects. Social media is thus producing massive amounts of data at an unparalleled scale and this is an attractive resource for obtaining interesting healthcare insights and health hazards information [7, 8]. When one’s family member is suffering from a specific disease or in the case of an outbreak, a person tends to search or write about the disease as well as the related symptoms on her/his favorite social network [9–13]. The analysis of structured and unstructured data from these social networks is referred to as *social media analytics*. After the advent of Web 2.0 in the early 2000s, social media analytics has emerged with data centric nature being a key characteristic [14].

Social media analytics spread over several disciplines such as psychology, sociology, anthropology, computer science, mathematics, physics, and economics and even politics [7, 13]. Data can be analyzed using online and offline services [15]. Systems such as DEFENDER [16] and ORBiT [17] make use of social network data such as Twitter. Ramanathan et al. [17] examined the use of the diagnostic data to reveal spatial and temporal patterns of how the 2009 H1N1 pandemic affected the USA. ORBiT was used to identify, quantify and describe spatial and temporal patterns of the 2009–2010 pandemic H1N1 flu within the United States. It was observed that the integration of heterogeneous data sources, including publicly accessible sources, could provide timely and novel information regarding influenza pandemic. As reported by Boulos et al. [18], “*in Cambodia, the Ministry of Health uses GeoChat for disease reporting and to send staff alerts and rapidly escalate response to potential outbreaks, while in Thailand, more than 900 facilities within the Hospital Network exchange information and get alerts to monitor influenza outbreaks in real-time from facilities across the country.*”

However, there are several challenges regarding the analysis of social media data. Twitter, for example, poses several challenges, since streams contain large amounts of meaningless

messages, polluted content and rumors, which negatively affect the detection performance. Moreover, traditional text mining techniques are less suitable for Twitter, because of the short length of the tweets, the large number of spelling and grammatical errors as well as the frequent use of informal, irregular and abbreviated words, and the use of improper sentence structure and mixed language [19]. Moreover, in the context of Big Data Analytics (BDA) from social networks, several issues are raised. Some of these include advances in scale-out systems, data models and high-level abstractions, incremental processing, scalability, resource sharing and classification of social networks [20]. Real-time data analytics poses a number of challenges due to the fact that the resulting actions are usually based on the current and previous situations. In other words, the system needs to react instantly based on what is happening in terms of data being updated. The main challenges are transfer of data events in real-time, incomplete data, analysis of the large volume of data, scalability and infrastructural issues, visualization of the results, and the decision-making [21–24]. The data events can be transferred from the social networks as raw events or as filtered or aggregated events. Real-time analytical services need to deploy fast algorithms that provide alternative options within a bounded time. For effective visualization, data can be grouped together instead of plotting points on a graph, which may be difficult when dealing with extremely large amounts of information or a variety of categories of information. As far as the decision-making process is concerned, it may be automated or not depending on the type of application, for example, for the disease surveillance it will require some level of human inputs. In fact, in most clinical systems, advanced real time analytics has not been incorporated [25].

Crowdsourcing is an online and distributed sourcing model in which individuals or organizations use contributions from Internet users to obtain needed services or ideas [26]. It allows large-scale and flexible invocation of human input for data gathering and analysis, which introduces a new paradigm of data mining process [27]. Crisis map is a crowdsourcing platform, designed to do information collection, analysis of mass data and display in a straightforward way in real time during a crisis. A few examples of crowdsourcing platforms include: CrowdSource [28], Figure Eight (formerly CrowdFlower) [29], CrowdForge [30], CrowdWeaver [31], the Amazon Mechanical Turk (MTurk) [32] and Crowdbreaks [33]. The types of data mining tasks that can be accomplished using crowdsourcing are classification, clustering, semi-supervised learning and association-rule mining. In the context of healthcare applications, some mobile crowdsourcing applications include HealthMap’s Outbreaks Near Me [34], Sickweather [35], Flu Near You [36], UH-BigDataSys [37], Stop Corona! [38] and Mo-Buzz [39].

A good way to complement a surveillance system is to allow patients to input their medical condition to a centralized system, which is capable of keeping an eye over the country.

A very practical way to allow this is to implement a mobile application accessible to all for sending the data. With the prevailing use of smartphones, it is now easy for anyone to use a mobile application to upload their medical condition to the system. Sensing application domains using mobile phone is being used in healthcare [18]. Schwind et al. [40] present a local media surveillance, which might be related to possible outbreak situations. Google Trends can be used in a similar way such that specific terms associated to the diseases are monitored in a localized region. This data source can complement disease outbreak detection. A script is run to capture the search interest of the respective terms throughout a time period and saved in a MySQL database. This can be combined with data from official sources.

## 2 Objectives

Despite the small size of Mauritius, it takes time to collect and process data, and there is usually a few weeks' time lag before the actual data becomes available. Furthermore, the diseases' estimates are not updated on a regular basis. Early detection of a pandemic is critical because it allows faster communication between policy makers and the public and provides more time to prepare in case of an outbreak. In 2005–2006, the Chikungunya infected a large number of people in Mauritius, causing the death of a few. The country was faced with a major outbreak in February/March 2006 following a minor outbreak in April/May 2005 [41]. Since August 2006, the country has not registered any other Chikungunya outbreak. Similarly, the A/H1N1 flu pandemic caused at least 69 infections and 8 deaths on the island in 2009 [42].

Currently, there is no integrated Health Information System (HIS) in Mauritius, thus the data of daily visits and interventions are not stored electronically on a centralized system anywhere. Hence, in the case of any infectious disease outbreak, pandemic situation is detected when it has already infected many people. News about a high number of cases registered for Flu, Gastroenteritis or Conjunctivitis, for example, are reported in the press only when the Ministry of Health has already registered around a few thousands of such cases and it is already quite late to contain the disease. To date, there is no automated surveillance system for any infectious disease being used in Mauritius although social media is extensively used in Mauritius [43].

The aim of this paper is therefore to present the features of a real-time disease outbreak surveillance system based on a crowdsourcing mobile application. A mobile application has been developed based on some major infectious diseases monitored by the health authorities (identified through the stakeholders' workshop), namely Influenza, Gastroenteritis, Scabies, URTI and Conjunctivitis. The methodology, design principles, implementation details and results are discussed in

subsequent sections. The system provides a real-time visual analysis, which can guide the authorities (Communicable Diseases Control Unit) to take prompt decision and provide the necessary healthcare service and advice to the public in general, thus help in the prevention of an outbreak situation.

It should be noted that the data collected through the mobile crowdsourcing application are unconfirmed cases and will be used solely to provide an indication of a possible disease outbreak as is the case with rumor-based and event-based disease outbreak surveillance system. Data from the real-time disease outbreak surveillance system will have to be correlated with the notifiable disease registers that contain confirmed cases being notified by qualified health professionals. Indeed, the system will be beneficial to the Communicable Diseases Control Unit as a complementary tool to track disease outbreak.

## 3 Methodology

In order to develop the real-time disease outbreak surveillance system, the following methodology has been adopted.

- **Literature Review:** An extensive literature review was conducted to explore the various data analytics techniques used for disease outbreak surveillance, with a focus on event-based surveillance and social media analytics. Related works on healthcare-related crowdsourcing applications have also been reviewed. Existing crowdsourcing platforms and applications designed for disease outbreak surveillance have been investigated.
- **Requirements Gathering:** To have a better understanding of the current methods used for disease monitoring in Mauritius, a stakeholders' workshop was carried out. Participants included representatives from various departments of the Ministry of Health and Quality of Life and the Pharmaceutical Division, WHO representatives, as well as representatives from public and private hospitals. The aim of the workshop was to identify the sources of information used for disease monitoring, the types of analysis carried out and the involvement of medical and pharmaceutical institutions. The workshop was also helpful to finalize the list of diseases to be monitored for a possible outbreak.
- **Analysis and Mapping of Symptoms:** The crowdsourcing application to be developed will monitor the shortlisted diseases, namely Influenza, Gastroenteritis, URTI, Scabies and Conjunctivitis. However, members of the public contributing to the application may not appropriately know the disease they are suffering from. They will therefore be required to only report their symptoms. The mapping between symptoms and diseases is done by the application. Since some people might not suffer from

all symptoms of a particular disease, a probabilistic approach was devised for the correlation. To limit the discrepancies of the symptoms to disease mapping, only submissions with an accuracy of 70% and above were considered.

- **Mobile Application Design:** A user-friendly crowdsourcing mobile application has been designed to allow users to select and submit the different symptoms they are suffering from. Users also have to enter their location upon their first use to easily localize cases of outbreaks, if any. The submissions are clustered into localized regions according to the location of users and the data are analyzed. To avoid duplication of data, a daily restriction of one submission per user is imposed by the application. The extent of the disease outbreak can be visualized district-wise for each of the diseases using the Early Aberration Reporting System (EARS) algorithm. Feature scaling was used to

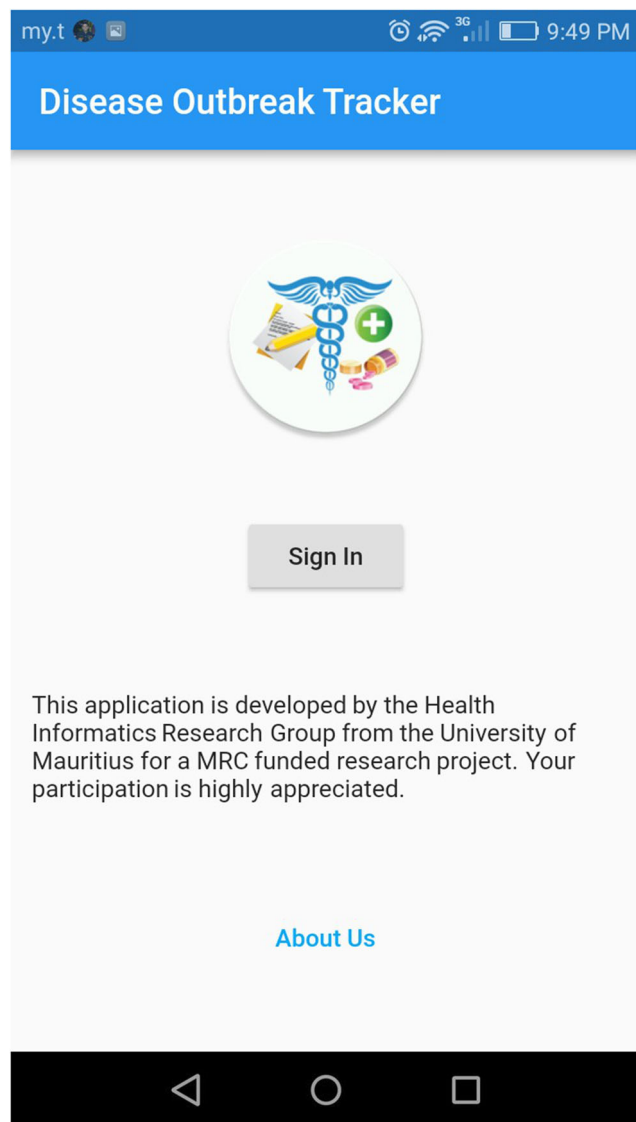


Fig. 1 Main Screen

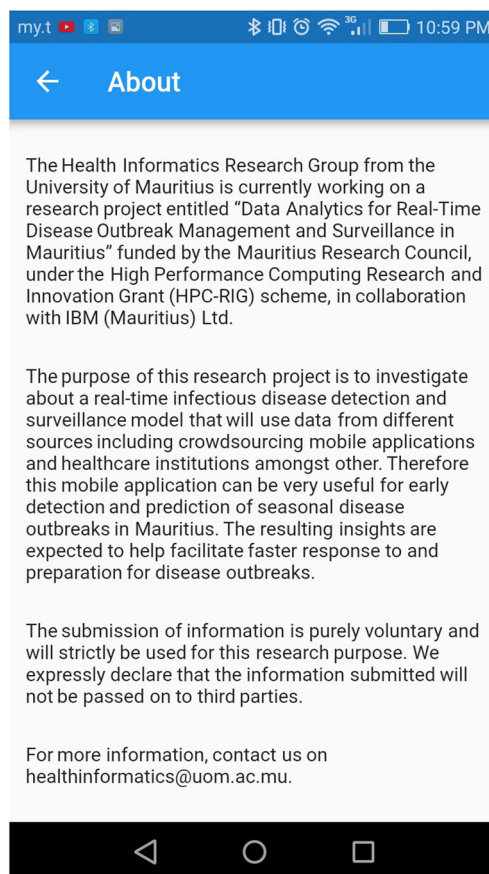


Fig. 2 About Us

standardize the data to a particular scale to allow better interpretation of the recorded data.

- **Application Deployment:** The mobile application is deployed on the Android platform. It is named *DOT* (Disease Outbreak Tracker) and an appropriate, comprehensive logo has been designed for it. The authentication and database services are offered by Firebase, which is an open-source cloud storage service.
- **Application Promotion:** As a pilot phase, *DOT* was mainly advertised to University students through emails, on the university website and on social media and was hosted on Google Play Store. Following feedback received from the first release, a second release was deployed after addressing the issues encountered by users. There were 94 installs with 167 set of symptoms being reported for the period May 2018 to December 2018.

#### 4 DOT crowdsourcing Mobile application

The primary purpose of the *DOT* crowdsourcing mobile application is to enable the general public to report diseases in real time. Since the public might be ignorant of the actual



nature of a disease, the application does not include a list of diseases but a list of symptoms for a set of diseases. As mentioned in the previous section, the stakeholders advised that the application focuses on Influenza, Gastroenteritis, URTI, Scabies and Conjunctivitis. Consequently, the mobile application lists a set of symptoms for these diseases. The mobile application was initially tested among the members of the research group and was then published on Google PlayStore to be made available to the general public.

The main screen of the *DOT* mobile application (Fig. 1) provides a Sign-in feature, a short description of the mobile application and a link to the *About Us* page (Fig. 2).

Google Sign-in (Fig. 3) has been integrated in the *DOT* mobile application and it allows users to authenticate with the Firebase database by using their Google Accounts. The users can select one of the many Google accounts set up on the mobile device or may also proceed by creating a new account. During registration, users also have to select the location (district) they are from. The reporting screen is shown in Fig. 4. It displays the symptoms for the most common infectious diseases in Mauritius: Conjunctivitis, URTI, Scabies and Gastro-Enteritis together with some general regular sickness symptoms. A brief description of the symptoms are also provided to help people in reporting the most appropriate symptoms.

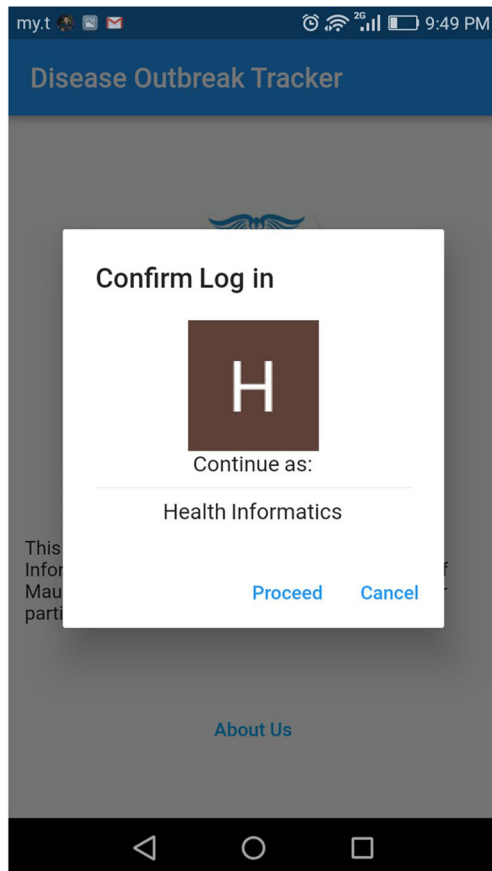


Fig. 3 Google Sign-in

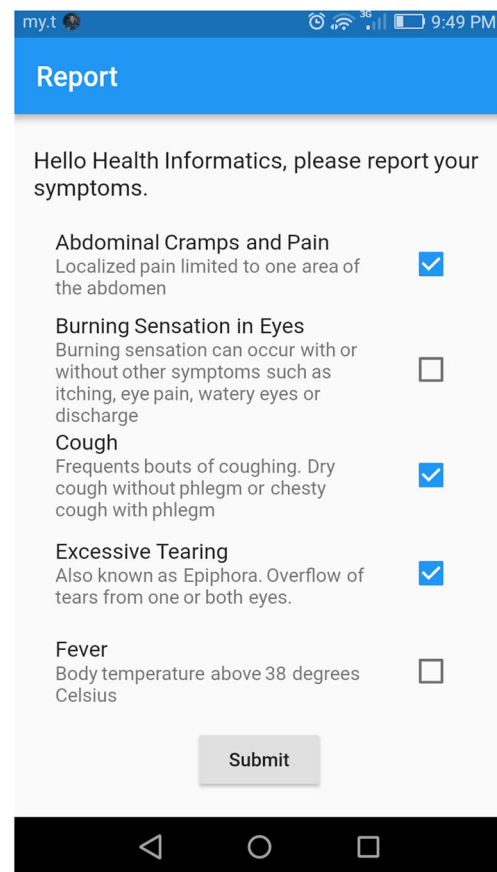


Fig. 4 Report Symptoms

Though the philosophy behind crowdsourcing applications is that users would honestly contribute with accurate data, such applications are, however, always prone to malicious users submitting fabricated data or launching various sorts of attacks [44]. As a means to limit the submission of inaccurate data, only one submission per user is allowed per day. The data are stored on the Firebase database in the format shown in Fig. 5.

All the reported symptoms are saved on the path “Individual Report”. Under this path, the data are categorized based on the submission date and then on the location of the user which the latter chose upon registration. Under a specific location, the Firebase’s unique Id is saved and finally all the submitted symptoms are saved at the end of the tree. This tree enables the application to know whether a submission has already been made for a specific user.

To be able to analyze the submitted information, the data generated from the mobile application is migrated to a MySQL database for easier processing. Every submission is then evaluated to determine the probability of the closest disease. Every disease has its own set of symptoms  $R_i$ , with cardinality  $r_i$  where  $1 \leq i \leq 5$ . The reported symptoms form a separate set, say  $R$ .  $R$  is then intersected with each  $R_i$ .

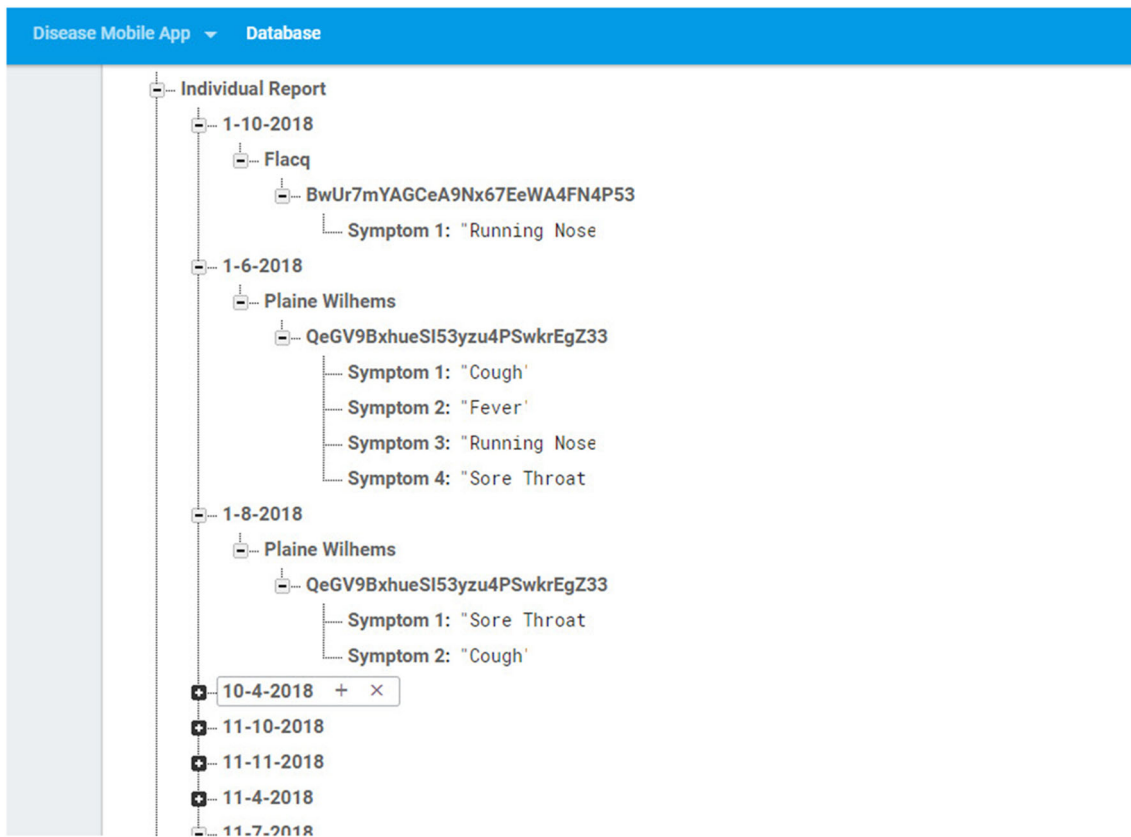


Fig. 5 Firebase Database

Considering the cardinality of each intersected set to be  $x_i$ , the probability  $x$  that the reportee is suffering from any of the diseases is computed as:

$$P(x) = \frac{X_i}{r_i}; 1 \leq i \leq 5$$

Based on the submitted symptoms, the disease having the highest probability is then considered. The data, excluding the symptoms, are then saved in a new table (Fig. 6) along with the corresponding disease. This data in this table is hence ready for applying the surveillance algorithms as shown in the next section.

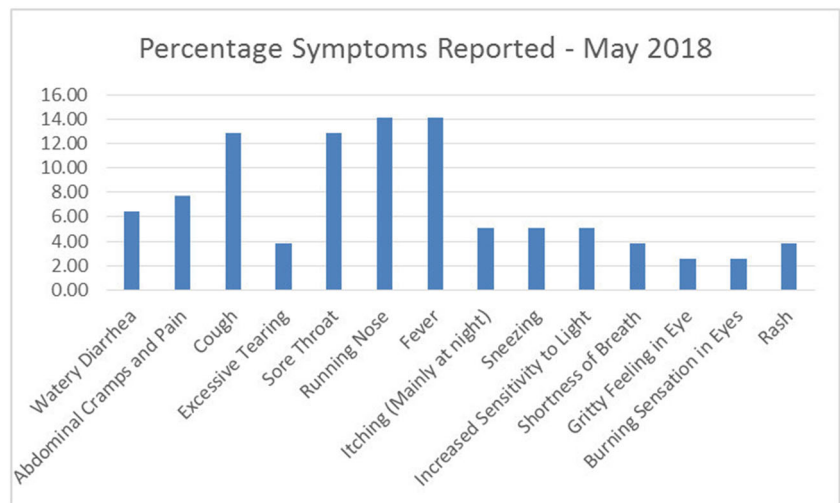
## 5 Disease outbreak analysis methods

EARS is an outbreak detection model which analyses the standard deviation of the data recorded on a daily basis. This model has two key factors, which play an important role in the system. It requires a baseline, which demarcates the situation as an outbreak or not. The baseline is decided upon aspects such as population and risk factors. The second factor is the flow of data. The model requires data from at least the last 7 consecutive days for the computation. Lack of any such data will result in impairment of detecting an outbreak. The criteria for an outbreak is when the current data exceed the baseline. In

id	uid	disease	rdate	location	chance
1	OeGV9BxhueSI53vzu4PSwkrEaZ33	Scabies	2018-04-03	Plaine Wilhems	0.333333
2	zXMiEO90xMYabGsWxENKAJCHKNa2	Coniuctivitis	2018-04-03	Plaine Wilhems	0.5
3	OeGV9BxhueSI53vzu4PSwkrEaZ33	Scabies	2018-04-03	Black River	0.333333
4	zXMiEO90xMYabGsWxENKAJCHKNa2	Coniuctivitis	2018-04-03	Black River	0.5
5	3vPtHWtOKMVAAskhL9KC5WA4T47I2	Scabies	2018-04-03	Black River	0.333333
6	GJ5eXkeHaaPaovIu70uva259RwO2	Scabies	2018-04-04	Plaine Wilhems	0.333333
7	4J54h0l9W7YRSJJbiH9DTfVHxOI3	Flu	2018-05-04	Plaine Wilhems	0.5
8	xEszKIMJchOkrlk8anviaafa0aS2	Flu	2018-05-08	Riviere du Rempart	0.75
9	xEszKIMJchOkrlk8anviaafa0aS2	Flu	2018-05-08	Plaine Wilhems	0.75
10	6mWd9YahIwYDsaOhCnssOa0i3Nv1	URTI	2018-05-08	Plaine Wilhems	0.333333

Fig. 6 Extracted Disease with Probabilities

**Fig. 7** The symptoms reported in May 2018



EARS, there are three detection methods, which are differentiated by their level of rigidity. The three methods are described as follows.

**C1-Mild** provides the least rigid algorithm among the 3 methods. C1 method requires the data of the last 7 consecutive days to determine if this foresees any possible outbreak given the current situation. An alarm is triggered when the current data exceeds three sample deviations from the sample mean, that is,  $C1(t) > 3$ .

**C2-Medium** works in a similar way as C1, except that it leaves a 2-days gap before considering the 7 days required for computational purposes. This gap decreases the possibility of any biasing of the data from the last 2 days. The 7 days taken are more likely to be an “outbreak-free” observed count. An alarm is again triggered when the current count exceeds three sample deviations from the sample mean.

**C3-Ultra** is the method with the most rigidity and is more likely to detect an outbreak first, in comparison to the previous 2 methods. C3 is a combination of historical and current data. It requires the current observed count as well as the C2 values of the previous 2 days. In this case, a possible outbreak is detected when  $C3 > 2$ . A limitation of this method is that there cannot be any gap while the data are being collected. In case (for some reasons), data were not recorded for a particular day, the observed count is considered to be zero, but this will highly affect the detection. A way to counter this is to neglect such days in the timeline and maintain a non-zero continuity.

The EARS method has been used for the analysis of the crowdsourcing data obtained from mobile application for the five diseases being monitored.

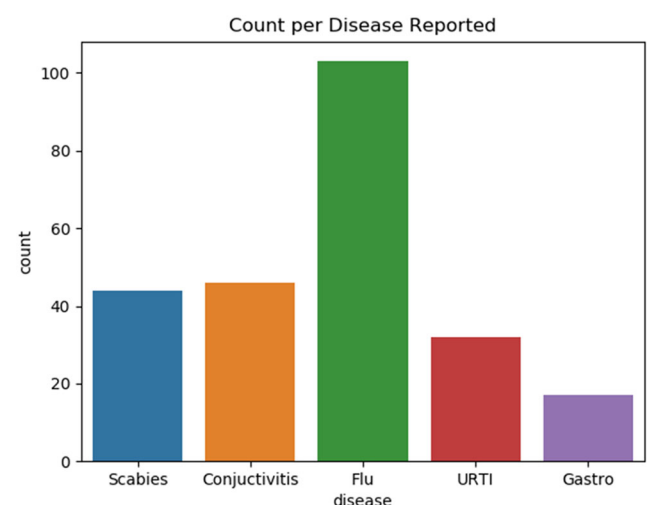
## 6 Results and discussions

Data were collected using the *DOT* crowdsourcing mobile application in May 2018 and were plotted on a daily basis

over the said period, as shown in Figs. 7, 8, 9 and 10. A total of 208 users had subscribed to the system.

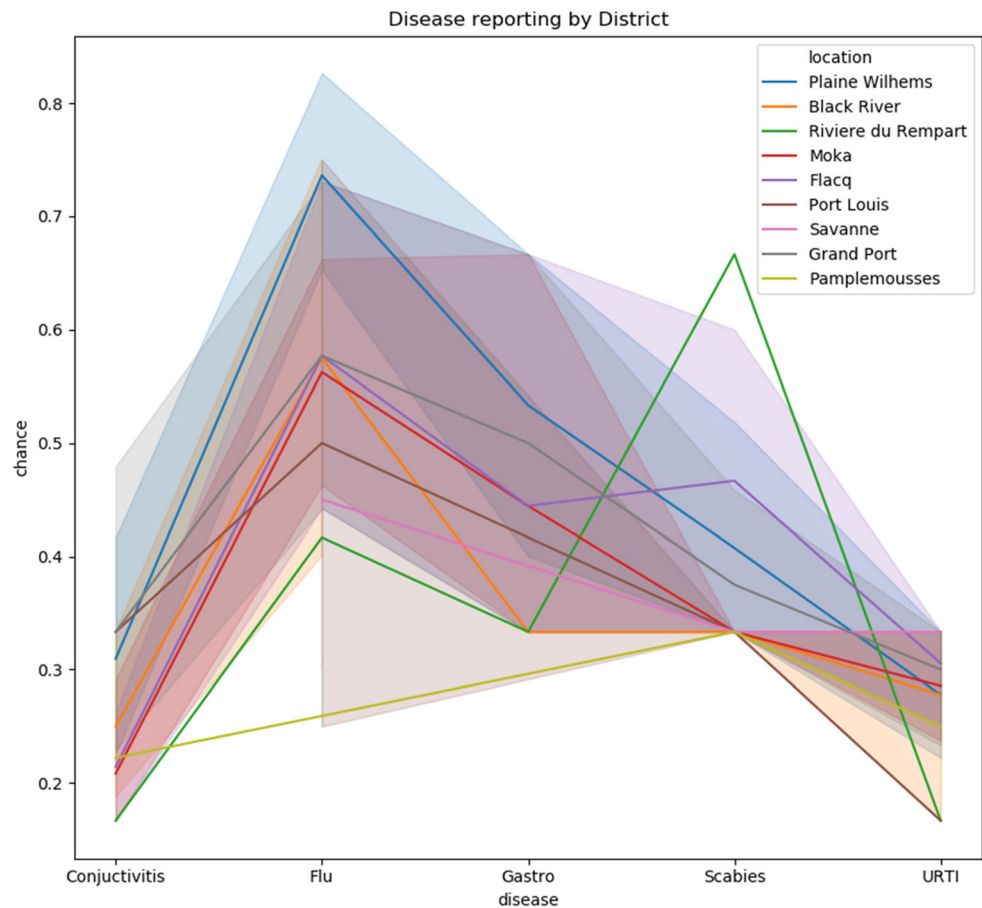
Figure 7 shows the symptoms reported by users. It can be observed that the most common symptoms reported during this period were cough, fever, running nose and sore throat. From Fig. 8, it was observed that 100 users reported Flu symptoms, which amounts to about 50% of the users. Given that these data represent reporting done during the period of May when there is a change in climate (summer to winter), with a higher incidence of Flu like symptoms, this correlation is further supported with the results in Fig. 8.

Most of the reporting for all the districts were Flu as shown in Fig. 9. However, in one of the districts, a number of cases were reported for Scabies as well. Due to the number of cases reported being quite low as well as the duration of the reporting, no conclusion can be drawn in this case. It is expected that as more users will register on the system there will



**Fig. 8** The highest number of reporting was for Flu

**Fig. 9** Disease Reported by Districts



be more reported number of cases for each of the diseases and therefore better disease outbreak analysis can be performed.

Figure 10 depicts the EARS computation for Flu, using data from 5th May to 20th May 2018. The threshold for plausible outbreak is set at 3 for C1 and C2 method, while C3 method has its threshold at 2. On 7th May, all three methods faced a spike, due to a sudden high number of submissions. Since C1 is the method with least rigidity, it was the first one whose alert waved off on 9th May. The alert was waved off one day later for the other two methods.

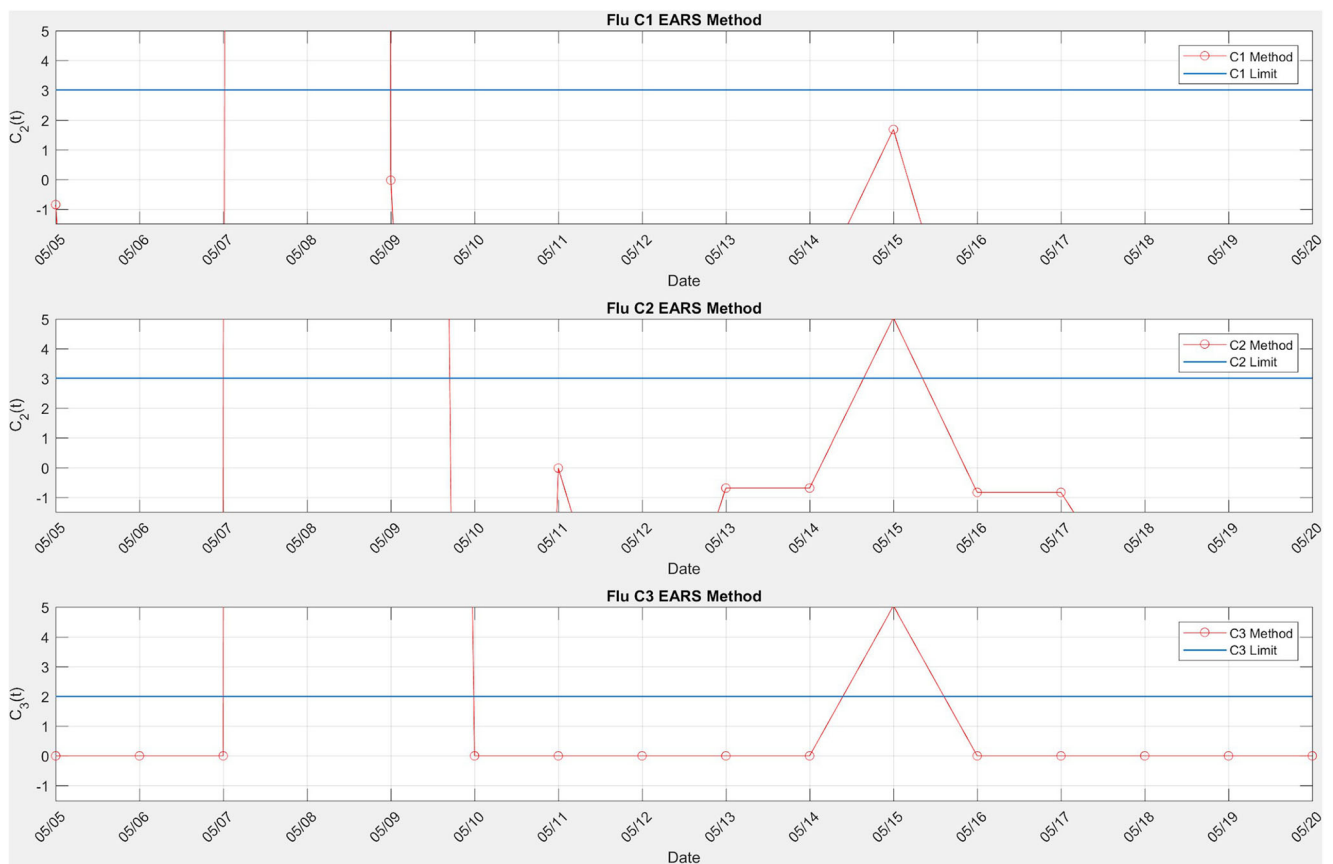
Another period where the rigidity can be evaluated is during 14th May to 16th May. C1 was the only method, which did not detect any outbreak. C3, being the most rigid method, signaled an alert earlier than C2 and lasted for a longer duration. Determining the proper method for surveillance depends on the necessity to get an alert for the observed disease as well as on how reliable the mobile crowdsourcing data are. The characteristics of the disease can also play a part in choosing the correct method as a contagious disease has the ability to spread out at an exponential rate, and hence, needs to be detected and contained at the first sign of plausible outbreak.

The initial trial of the crowdsourcing mobile application shows the potential for early detection and prediction of seasonal disease outbreaks. The resulting insights are expected to reduce the response time in case of a pandemic, as well as help in tracking the spread of an infectious disease in Mauritius. Since the system will be completely automated and the output of analysis will be updated in near real time, it is expected to detect disease outbreaks significantly faster than the traditional disease surveillance system that collects public health data from sentinel medical practices.

Historical data were also obtained from the Communicable Diseases Control Unit (CDCU), Ministry of Health, for 2017. Figure 11 provides a snapshot of the number of reported cases for URTI, starting from January 2017 to December 2017.

If we compare the number of cases reported as possible flu cases in May 2018 compared with the cases of URTI in May 2017, it is clear that there is a correlation. Around that period, there is a change in temperature from warm to slightly colder, hence a rise in flu-related diseases and URTI. According to [45], seasonal exposure to cold air causes an increase in the incidence of URTI due to cooling of the nasal airway.





**Fig. 10** Data visualization for the flu data collected using the mobile application, based on the EARS methods

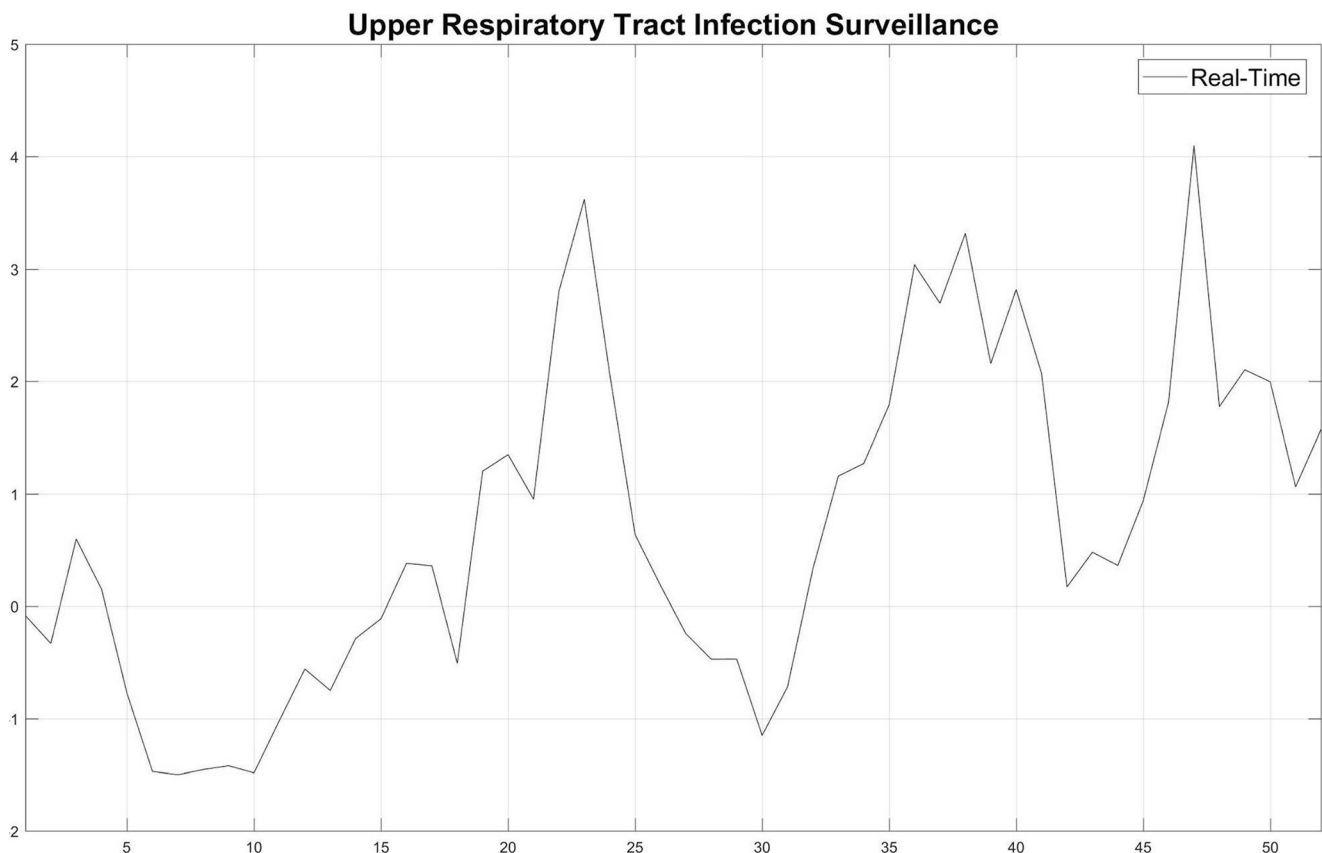
The close correlation between the crowdsourcing data and the historical data from the Ministry shows the reliability of the real-time disease outbreak surveillance system. It is expected that as more users register on the system there will be more submissions of reported diseases cases and the system will provide more accurate results. As future works, other sources of information such as pharmacy sales for specific drugs related to the diseases being monitored and absenteeism in schools and workplace will be integrated in the real-time disease outbreak surveillance system for better detection of disease outbreak.

The proposed real-time disease outbreak surveillance system is intended to be used by the Communicable Diseases Control Unit as a complementary tool to track disease outbreak in Mauritius. Analytical data from the system will only be used as an indication of a possible outbreak and will allow the health authorities to have advance notice of outbreaks that may eventually be correlated with daily confirmed cases of specific diseases.

A comparative evaluation was carried out with respect to existing health-related surveillance systems identified in the section 1 based on different criteria as highlighted in Table 1.

From Table 1, it can be observed that applications like *Mo-Buzz* and *Stop Corona!* restricted their surveillance to only one type of disease, namely Dengue and COVID-19 respectively as compared to *DOT*. Like *Stop Corona!*, *DOT* provides the user with a list of different symptoms to choose from when reporting health status while in *Sickweather* and *Crowdbreaks*, the main source of information is social media data with more emphasis on Twitter. However, in Mauritius there is limited health-related information being shared on Twitter with only few active users on. The main source of information in *DOT* is a crowdsourcing application, which is more suited to the context of Mauritius. In terms of visualizations, *DOT* limits itself to hotspots in Mauritius while *FNY*, *Stop Corona!* and *Outbreaks Near Me* built on *HealthMap* show a worldwide view of hotspots for specific diseases, with different colors indicating different outbreak status. Compared to *DOT*, *UH-BigDataSys* makes use of IoT sensing and focuses on diseases related to air quality only.

As future enhancements, real-time physiological indicator, as captured in *UH-BigDataSys*, could be considered as potential new sources of data for *DOT*. Additionally, like in *Mo-Buzz*, the posting of health alerts and personalized messages to people living in specific geographical zones, based on the identified hotspots can be considered. Similar to *Outbreaks*



**Fig. 11** URTI Records for January 2017 to December 2017

*Near Me*, *DOT* can be enhanced to advise users based on their current locations and be platform independent. Furthermore, *DOT* can be extended to include symptoms for COVID-19 and other diseases. The ability of *DOT* to integrate admissions data from healthcare institutions can considerably improve the reliability of disease surveillance. None of the applications analyzed considered such sources of data.

## 7 Conclusion

The aim of this work is to develop a crowdsourcing mobile application for Disease Outbreak Surveillance for Mauritius since our background study has shown that there exists no such application in the country. The application is intended to monitor the number of cases reported for chosen communicable diseases across the island as well as in specific districts. To gain a better understanding of the disease outbreak surveillance in Mauritius, a stakeholders' meeting was held with representatives of various health institutions of Mauritius. This exercise provided insight into the analysis currently being performed for outbreak detection as well as which diseases required monitoring in Mauritius. The communicable diseases that were chosen are

Influenza, Gastroenteritis, URTI, Scabies and Conjunctivitis. Next, the list of symptoms associated to these diseases had to be created as people might not know which disease they are suffering from but are aware of the symptoms they have. This list was included in the application and users had to select all symptoms that apply. *DOT* was successfully implemented, tested amongst a small group of people and then was deployed on Google Play Store. All the data collected were anonymized.

Once symptoms data were collected using *DOT*, probabilistic methods were used to find the disease or diseases that the user might be suffering from. The data were further processed using the EARS algorithm to find the extent of the disease outbreak district-wise, per disease. These data were represented graphically for a rapid understanding of the situation in each district. Our findings concur with existing data for the same period for the previous years showing that the crowdsourcing application can aid with disease outbreak detection. One major issue of the project was the low-adoption rate of the application. However, in a real-life scenario if the application is deployed on behalf of decision-makers and is promoted on a more massive scale, adoption rate will be higher and the application can only be more efficient at detecting disease outbreak as early as possible. In the future, *DOT* can be extended

**Table 1** Comparative Evaluation

Application	Domain	Data Collection and type	Analytics/ Methods/ Algorithms	Results/ Visualization
Crowdbreaks [33]	Health trends, flu or other infectious diseases	Tweets from Twitter using a list of vaccine-related keywords collected between July 2018 and January 2019	FastText classifier for sentiment analysis	Real-time predictions of vaccine sentiments, sentiment trends over time
HealthMap's Outbreaks Near Me [34]	H1N1, typhoid, Swine Flu, Cholera, COVID-19 and other food-borne bacterial outbreaks	Health-related reports from a multitude of web based sources: GoogleNews, ProMed-Mail, World health organization reports, Geo Sentinel, OIE (the World organization for animal health), EuroSurveillance, Baidu News, Soso info and the Wildlife Data Integration Network	Uses Bayesian methods to classify News into breaking News, context, warning, old News and non-disease News	GIS map of hotspots - Regional Disease Map View
SickWeather [35]	Flu, Norovirus, Pink Eye, Whooping Cough, and 20 other illnesses	Social media monitoring (Facebook, Twitter) and user reporting (symptoms, illness, location)	Machine learning models are used to measure the rate of real-time input compared to an extensive archived data to predict the rate of illnesses	The application tracks all the illnesses on a map and provides alerts of sickness that may be present in a specific region (sick zone) along with forecasts. Users are able to get alerts about nearby sicknesses and daily forecasts on active illness.
Flu Near You (FNY) [36]	Influenza-like illness (ILI), H1N1	Participants report health information (their symptoms and that of their household members) weekly via website, Facebook or mobile application [46]	ILI rates are computed by dividing the number of weekly participants by the total number of FNY participant reports in that same week at each spatial resolution [47]	Reports are generated and mapped to allow one to be aware when the flu is around. GIS Map is used to provide regional disease Map view.
UH-BigDataSys [37]	Health advices for respiratory disease, sleep quality control and outdoor travel	Air quality data collection through: meteorological site data, mobile device crowdsourced data and IoT sensing data. Real-time physiological indicator (collected through wearable devices)	Spring framework for BDA	Dashboard for trend analysis (Air Quality Index and Physiological Status). On mobile: Air Condition Data, Physiological Data and Healthcare Guidance through GIS
Stop Corona! [38]	COVID-19	Participant information such as gender, age, location, whether person has travelled abroad, whether person has been in contact with another person suffering from COVID-19 and whether person is suffering from disease or not.	System aggregates all information to show on a map of the world	Heatmap of disease hotspots
Mo-Buzz [39]	Dengue	Participant information (symptoms, mosquito bites, breeding sites) is reported through interactive forms and social media (twitter). Other sources include weather data and human and vector population.	Predictive surveillance based on a computer simulation. GPS-based location sensor to track location.	Information is disseminated in the form of hotspot maps to health officials and the public. Health alerts and personalized messages are sent to people living in specific geographic zones.
Disease Outbreak Tracker (DOT)	Influenza, Conjunctivitis, URTI, Scabies and Gastro-Enteritis	Data are recorded through crowdsourcing applications (symptoms) and via submissions from clinic (admissions data)	EARS methods (C1,C2,C3)	Dashboard showing: Spatial visualization of hotspots in Mauritius and trends with respect to number of infections over time and per district

to include symptoms for COVID-19 and other diseases. Other data sources such as pharmacy sales for specific drugs related to the diseases being monitored and absenteeism in schools and workplace can also be integrated.

**Funding information** This work was supported by the Mauritius Research Council [MRC/HPCRIG-A03].

The content is solely the responsibility of the authors and does not necessarily represent the official views of the MRC.

## Compliance with ethical standards

**Disclosure of interest** The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Ethics compliance** Consent was obtained from all users of the application upon registration on the *DOT* application and prior to using the application. All data collected were anonymous and no personal data of users were saved.

**Conflict of interest** The authors declare that they have no competing of interests.

## References

- Semenza JC, Menne B. Climate change and infectious diseases in Europe. *Lancet Infect Dis*. 2009;9:365–75. [https://doi.org/10.1016/S1473-3099\(09\)70104-5](https://doi.org/10.1016/S1473-3099(09)70104-5).
- Abat C, Colson P, Chaudet H, Rolain JM, Bassene H, Diallo A, et al. Implementation of syndromic surveillance systems in two rural villages in Senegal. *PLoS Negl Trop Dis*. 2016;10:e0005212. <https://doi.org/10.1371/journal.pntd.0005212>.
- Isere EE, Fatiregun AA, Ajayi IO. An overview of disease surveillance and notification system in Nigeria and the roles of clinicians in disease outbreak prevention and control. *Niger Med J*. 2015;56:161–8. <https://doi.org/10.4103/0300-1652.160347>.
- Manogaran G, Lopez D. Disease surveillance system for big climate data processing and dengue transmission. In: Management Association IR (ed) *Web services: concepts, methodologies, tools, and applications*. IGI Global; 2019; pp 490–509.
- Choi J, Cho Y, Shim E, Woo H. Web-based infectious disease surveillance systems and public health perspectives: a systematic review. *BMC Public Health*. 2016;16:1238. <https://doi.org/10.1186/s12889-016-3893-0>.
- Fung IC-H, Tse ZTH, Fu K-W. The use of social media in public health surveillance. *Western Pac Surveill Response J*. 2015;6:3–6. <https://doi.org/10.5365/WPSAR.2015.6.1.019>.
- Gandomi A, Haider M. Beyond the hype: big data concepts, methods, and analytics. *Int J Inf Manag*. 2015;35:137–44. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>.
- Stieglitz S, Mirbabaie M, Ross B, Neuberger C. Social media analytics – challenges in topic discovery, data collection, and data preparation. *Int J Inf Manag*. 2018;39:156–68. <https://doi.org/10.1016/j.ijinfomgt.2017.12.002>.
- Batrinca B, Treleven PC. Social media analytics: a survey of techniques, tools and platforms. *AI Soc*. 2015;30:89–116. <https://doi.org/10.1007/s00146-014-0549-4>.
- Chae J, Thom D, Bosch H, et al. Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition. 2012 IEEE Conference on Visual Analytics Science and Technology (VAST). IEEE; 2012; pp 143–152.
- Fan W, Gordon MD. The power of social media analytics. *Commun ACM*. 2014;57:74–81. <https://doi.org/10.1145/2602574>.
- Ribarsky W, Xiaoyu Wang D, Dou W. Social media analytics for competitive advantage. *Comput Graph*. 2014;38:328–31. <https://doi.org/10.1016/j.cag.2013.11.003>.
- Stieglitz S, Dang-Xuan L. Social media and political communication: a social media analytics framework. *Soc Netw Anal Min*. 2013;3:1277–91. <https://doi.org/10.1007/s13278-012-0079-3>.
- Zeng D, Chen H, Lusch R, Li S-H. Social media analytics and intelligence. *IEEE Intell Syst*. 2010;25:13–6. <https://doi.org/10.1109/MIS.2010.151>.
- Lutu PEN. Data analytics to support social media marketing : challenges and opportunities. 2016.
- Thapen N, Simmie D, Hankin C, Gillard J. DEFENDER: detecting and forecasting epidemics using novel data-analytics for enhanced response. *PLoS One*. 2016;11:e0155417. <https://doi.org/10.1371/journal.pone.0155417>.
- Ramanathan A, Pullum LL, Hobson TC, et al. ORBiT: Oak Ridge biosurveillance toolkit for public health dynamics. *BMC Bioinformatics* 16 Suppl. 2015; 17:S4. doi: <https://doi.org/10.1186/1471-2105-16-S17-S4>
- Kamel Boulos MN, Resch B, Crowley DN, Breslin JG, Sohn G, Burtner R, et al. Crowdsourcing, citizen sensing and sensor web technologies for public and environmental health surveillance and crisis management: trends, OGC standards and application examples. *Int J Health Geogr*. 2011;10:67. <https://doi.org/10.1186/1476-072X-10-67>.
- Louis A. Natural language processing for social MediaAtefeh Farzindar and Diana Inkpen (NLP technologies Inc., Université de Montréal; and University of Ottawa)Morgan & Claypool (synthesis lectures on human language technologies, edited by Graeme Hirst, volume 30), 2015, xix+146 pp; paperbound, ISBN 978-1-62705-388-4; ebook, ISBN 978-1-62705-389-1; doi: 10.2200/S00659ED1V01Y201508HLT030, \$55.00 Computational Linguistics 2016; 42:833–836. doi: 10.1162/COLI\_r\_00270.
- Tan W, Blake MB, Saleh I, Dustdar S. Social-network-sourced big data analytics. *IEEE Internet Comput*. 2013;17:62–9. <https://doi.org/10.1109/MIC.2013.100>.
- Eklund JM, McGregor C. Real-time online health analytics for interplanetary space missions. 2014 IEEE Aerospace Conference. IEEE. 2014; pp 1–10.
- Jirkovsky V, Obitko M. Semantic heterogeneity reduction for big data in industrial automation. *ITAT*. 2014;1214:1–10.
- Khan M, Wu X, Xu X, Dou W. Big data challenges and opportunities in the hype of Industry 4.0. 2017 IEEE International Conference on Communications (ICC). IEEE. 2017; pp 1–6.
- Van Sinderen M, Oude Luttighuis P, Folmer E, Bosems S. Enterprise Interoperability: 5th International IFIP Working Conference, IWEI 2013, Enschede, the Netherlands, march 27–28, 2013. Proceedings. doi: <https://doi.org/10.1007/978-3-642-36796-0>.
- Carrière R, MacDonald A, Chan Y. Past, present and future: the outlook from mid-career nurse Informaticians. *Nurs Leadersh (Tor Ont)*. 2016;28:8–17.
- Brabham DC. Crowdsourcing as a model for problem solving: an introduction and cases. *Convergence: The International Journal of Research into New Media Technologies* 2008; 14:75–90. doi: <https://doi.org/10.1177/1354856507084420>.
- Xintong G, Hongzhi W, Song Y, Hong G. Brief survey of crowdsourcing for data mining. *Expert Syst Appl*. 2014;41:7987–94. <https://doi.org/10.1016/j.eswa.2014.06.044>.
- CrowdSource. <https://www.crowdsourcing.com/>. Accessed 3 Sep 2019.



29. CrowdFlower. <https://www.figure-eight.com/>. Accessed 3 Sep 2019.
30. Kittur A, Smus B, Khamkar S, Kraut RE. Crowd Forge. Proceedings of the 24th annual ACM symposium on User interface software and technology - UIST '11. ACM Press, New York, New York, USA, 2011; p 43.
31. Kittur A, Khamkar S, André P, Kraut R. Crowd Weaver: Visually managing complex crowd work. Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work - CSCW '12. ACM Press, New York, New York, USA, 2012; p 1033.
32. Amazon Mechanical Turk. <https://www.mturk.com/>. Accessed 3 Sep 2019.
33. Müller MM, Salathé M. Crowdbreaks: tracking health trends using public social media data and crowdsourcing. *Front Public Health*. 2019;7.
34. Outbreaks Near Me. HealthMap, <https://www.healthmap.org/outbreaksnearme/>. Accessed 21 May 2020.
35. Sickweather. <http://www.sickweather.com/>. Accessed 17 May 2020.
36. Flu Near You. <https://flunearyou.org/> Accessed 21 May 2020.
37. Chen M, Yang J, Hu L, Hossain MS, Muhammad G. Urban healthcare big data system based on crowdsourced and cloud-based air quality indicators. *IEEE Commun Mag*. 2018;56(11):14–20.
38. Stop Corona!. <https://stopcorona.app/>. Accessed: 21 May 2020.
39. Lwin MO, Vijaykumar S, Fernando, et al. A 21st century approach to tackling dengue: Crowdsourced surveillance, predictive mapping and tailored communication. *Acta Trop*. 2014;130:100–7.
40. Schwind JS, Wolking DJ, Brownstein JS, PREDICT Consortium1, Mazet JAK, Smith WA. Evaluation of local media surveillance for improved disease recognition and monitoring in global hotspot regions. *PLoS One*. 2014;9:e110236. <https://doi.org/10.1371/journal.pone.0110236>.
41. Ghamy AB, Goorah SSD, Caussy BS, Cheeneebash J, Ramchurn SK. Clinical complications of chikungunya in Mauritius. *Int J Infect Dis*. 2008;12:e328. <https://doi.org/10.1016/j.ijid.2008.05.878>.
42. (2009) ECDC RISK ASSESSMENT Pandemic H1N1 2009. 1–23.
43. Jha KK. Impact of Social Networking Sites on the Mauritian Youth : A Study of their Life Styles and its Effectiveness in Nation Building. 2016; 114–124.
44. Ren J, Zhang Y, Zhang K, Shen X. Exploiting mobile crowdsourcing for pervasive cloud services: challenges and solutions. *IEEE Commun Mag*. 2015;53:98–105. <https://doi.org/10.1109/MCOM.2015.7060488>.
45. Eccles R. An explanation for the seasonality of acute upper respiratory tract viral infections. *Acta Otolaryngol*. 2002;122:183–91. <https://doi.org/10.1080/00016480252814207>.
46. Smolinski MS, Crawley AW, Baltrusaitis K, Chunara R, Olsen JM, Wójcik O, et al. Flu near you: crowdsourced symptom reporting spanning 2 influenza seasons. *Am J Public Health*. 2015;105(10):2124–30.
47. Baltrusaitis K, Brownstein JS, Scarpino SV, Bakota E, Crawley AW, Conidi G, et al. Comparison of crowd-sourced, electronic health records based, and traditional health-care based influenza-tracking systems at multiple spatial resolutions in the United States of America. *BMC Infect Dis*. 2018;18(1):403.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.