Chapter 26

Yellow Fever in Brazil: Using Novel Data Sources to Produce Localized Policy Recommendations

Shalen De Silva, Ramya Pinnamaneni, Kavya Ravichandran, Alaa Fadaq, Yun Mei, and Vincent Sin

Abstract *Background* Yellow fever is a fatal acute viral hemorrhagic disease. This disease that is spread through the bite of the *Aedes* mosquito is endemic in Africa as well as the Americas, where the tropical climate helps in its transmission. Between January 2016 and March 2018, several territories of the Region of the Americas reported confirmed cases of yellow fever. In view of a global shortage of yellow fever vaccine, it is important to curb the transmission of yellow fever through improved vector surveillance and eliminating mosquito breeding sites. Prompt detection of outbreaks using novel data sources can help in launching immediate responses. *Objective* We discuss modelling disease propagation and case incidence using novel data sources, including Google Trends and Google Streetview. We also provide recommendations for how to contain and manage the outbreak. *Methods* We consider three main methods. First, we look at a traditional vector-borne disease propagation model. We also consider Google Trends data to judge how interest in the disease correlates with incidence. Finally, we propose methods for correlating Google Street View images with incidence to improve policy regarding distribution of vaccines. *Results* In terms of the Google Trends data, we found that we were able to match both peaks with just a basic model, including one week of lag time. Both the traditional vector-borne disease propagation model and the Google Streetview-based computer vision model require further analysis. *Conclusion* Here, we provide a starting point

S. De Silva · R. Pinnamaneni
Harvard School of Public Health, Boston, USA
e-mail: rdesilva@hsph.harvard.edu

R. Pinnamaneni
e-mail: rpinnamaneni@hsph.harvard.edu

K. Ravichandran (✉)
Computer Science and Artificial Intelligence Lab, Massachusetts Institute of Technology,
Cambridge, MA 02139, USA
e-mail: rkavya@mit.edu

EECS, MIT, Cambridge, MA 02139, USA

A. Fadaq · Y. Mei · V. Sin
USF Department of Health Informatics, Tampa, USA

and guidelines for further improving upon existing disease propagation models and using deep learning methods to better predict where disease outbreaks may occur.

Keywords Yellow fever virus · Vaccine · Infectious disease · Google trends · Google maps streetview · Computer vision · Machine learning

Learning Objectives

Issues with resource distribution and treatment of Yellow Fever Virus; Usage of Google Trends Data to model case counts; How one might use Google Maps Streetview Images to predict case counts specific to certain areas.

26.1 Introduction

Yellow fever is a fatal disease, because of its ability to kill patients that contract it within a short period of time. Being an acute viral hemorrhagic fever, yellow fever consolidates its place as a disease with a huge potential for massive human casualties within a short period of time. Its capacity to spread from one human being to another through the bite of the *Aedes* species of mosquitoes makes its spread a point of interest for the healthcare system. The disease is endemic in Africa as well as the Americas, where the tropical climate offers optimal conditions for its transmission. However, it is also a disease against which an affordable and highly effective vaccine is available.

Brazil is one of the countries faced with a huge disease burden as far as the effects of this disease go. It has been hit by yet another Yellow Fever outbreak since 2016. WHO reports indicate as many as 464 cases of the disease in the country and 154 deaths within the period extending from July 1st 2017 to February 16th 2018. The cities of Sao Paulo, Rio de Janeiro and Minas Gerais are among the areas that reported a high number of both cases and deaths from the disease. Because of the Yellow Fever and the Zika outbreaks, Brazil has become a center of interest in research on mosquito-borne illnesses. The research aims to unearth the epidemiology of the disease as well as the measures that can be taken to mitigate it. Interestingly, WHO (2018) reported that evidence suggesting that only *Aedes aegypti* transmit the disease in Brazil is lacking, implying there could be another species of the *Aedes* mosquito, including the *haemagogus*. That the spread could be sylvatic (a forest species) will shift the focus of disease prevention and control. This coupled with the reported low vaccination rates, with none of the areas exceeding a coverage of 30% of the target population, makes Brazil an important case in the control of the disease.

26.1.1 Background

According to Goldani (2017), infection with the yellow fever virus could lead to mild and non-specific illness marked by headaches, fatigue, vomiting and other non-specific signs, to a severe illness involving jaundice, fever, chills, headache, bleeding from multiple sites and multiple organ failure. Between 20 and 50% of the people with severe disease die from it in a short time (Goldani 2017). Not enough research is available regarding the determination of the factors that influence the development of severe disease in some individuals and not others (Barnett 2007). Thus, the patterns of occurrence, populations at risk and the control of the disease rightfully warrant focus in research on this disease.

The African continent, Latin and South America have traditionally been known to be the ones reporting endemicity, however a change of patterns is now being reported and the disease is spreading to previously non-endemic countries in the Asia such as China where a case was reported in 2016 (Chen 2016). It has previously bewildered research how despite the high density of the *Aedes aegypti* in the region (as evidenced by the spread of similar viral hemorrhagic viral fevers like Dengue), there haven't been many reported cases of Yellow Fever in Asia (Wickramage 2013). Such an isolated case is by itself an indication of the danger that is caused by the global village that the world is today, where travelling and faster modes of contact between people has heightened the risk of transmission of the disease into areas where it was not reported before. In other words, no one is safe, and the need for vaccination, especially for travelers, cannot be overstated or overemphasized (Pramil Tiwari 2017). While Brazil might be the country facing the outbreak right now and feeling the effect of the threat of the disease, it is the prerogative of the rest of the world to worry about the danger that lies in this outbreak. A high level of alert is specially warranted considering the fact that this particular outbreak hit tourist destinations that were previously spared.

Vaccination remains the mainstay for the management of yellow fever. The 17D-204 YF vaccine is available for use and has been used widely among travelers to endemic areas of Africa and Latin America, to protect them from contracting the disease. The vaccine is effective and efficacious, with a 99% efficacy in preventing the contraction of the disease (Khanna 2013). The vaccine is long-lasting and confers lifelong immunity. Treating travelers alone without focusing on the area where the problem lies is akin to neglecting the real problem. The real solution lies with eradication of the disease in these areas that are considered endemic or hyper endemic such as Brazil. The question, then, is: why has it been difficult to eradicate Yellow fever from Brazil?

The eradication of yellow fever will be dependent on the ability to eradicate *Aedes* mosquito, the main vector for yellow fever as well as Dengue fever, Chikungunya, and other viral hemorrhagic fevers. As a matter of fact, the eradication of *Aedes aegypti* was well documented during the mid-20th century (Kotsakiozi 2017). However, this eradication was not sustained and as the efforts of the eradication campaign ceased, the mosquitoes swiftly re-established themselves within the ecosystem of

Brazil. Since then, efforts for the management of yellow fever have shifted to the vaccination process, which has suffered in recent times due to the paucity in supply. The tropical rainforests cover a large part of the country and animal reservoirs for the virus such as monkeys make it even more difficult to completely eradicate the virus.

Added to this, the rapid expansion of civilization into the forested areas is making the epidemiology ever-changing. The disease has found its way into the urban cycle. Until now, it was only the sylvatic cycle that was reported, with infections occurring in the jungle where the *Aedes heamagogus* infects the monkeys in a cycle that gets to humans when they visit the Amazon and get bitten by the mosquitoes. The migration of the virus southwards and towards cities, where it can be readily received by the *Aedes aegypti* plying the city dwellings and slums, spells more disaster for the millions that reside in these urban areas (Snyder 2018). A refocusing of efforts on this mosquito might once again be necessary in the war against yellow fever.

26.1.2 Brazil Outbreak (2016–Present)

Between January 2016 and March 2018, seven countries and territories of the Region of the Americas reported confirmed cases of yellow fever—the Plurinational State of Bolivia, Brazil, Colombia, Ecuador, French Guiana, Peru, and Suriname. In Brazil, between 1 July 2017 and 13 March 2018, there were 920 confirmed human cases of yellow fever, including 300 deaths; this figure is higher than what was reported for the same period of the previous year (610 confirmed cases including 196 deaths). Comparing the epidemiological curve in both periods (2016/2017 and 2017/2018) the highest incidence rate is during epidemiological week (EW) 3 of both years. With respect to the probable sites of infection of the confirmed cases, in decreasing order they are the states of Minas Gerais, São Paulo, Rio de Janeiro, Espírito Santo, and in the Federal District (Figs. 26.1 and 26.2).

26.1.3 Yellow Fever Vaccine Shortage

Sanofi Pasteur, the makers of Yellow Fever Vaccine (YF-VAX) announced an anticipated shortage of vaccine through the middle of 2018. The vaccine shortage is assumed to impact both existing vulnerable population in the area of the outbreak and the travelers to the area. The carnival in Brazil saw a large influx of travelers that fueled the spread of the outbreak to other areas. The global shortage of vaccine supply has also affected travel to areas where it is mandatory to show proof of vaccination for entry to the country.

In response to the shortage, Sanofi Pasteur is working with the FDA to offer a European yellow fever vaccine called Stamaril under the Investigational New Drug

Fig. 26.1 Distribution of confirmed human yellow fever cases by epidemiological week (EW). Brazil 2016–2018. *Source* Data published by the Ministry of Health of Brazil and reproduced by PAHO/WHO

Fig. 26.2 Confirmed human cases and municipalities with confirmed yellow fever epizootics. Brazil, 17 January 2018, 22 February 2018, and 13 March 2018

Program. In Brazil, a fractional dose is being administered to provide short-term immunity, but its efficacy is not yet known.

In view of this vaccine shortage, it is important to curb the transmission of yellow fever through improved vector surveillance and eliminating mosquito breeding sites. Epidemic preparedness and response are key to saving lives by preventing outbreaks. Prompt detection of outbreaks using social media and other forms of technology will help in launching an immediate response.

26.1.4 Purpose

In this paper, we discuss modelling disease propagation and case incidence using novel data sources, including Google Trends and Google Streetview. We also provide recommendations for how to contain and manage the outbreak.

26.2 Methods

A vector-borne disease propagation model would traditionally work by looking at an infection status pathway through a population and considering how the infection state may change based on vector transmission, vaccinations and asymptomatic cases. The model diagram (Fig. 26.3) describes the logic of the model as follows: A population of individuals susceptible to infection S_h , are exposed by an infectious vector. These exposed members of the population can go on to become infectious individuals with symptoms or without symptoms. They can both be a source for new vectors to acquire the infection. Those with symptoms may go from a toxic state to either death or recovery, while asymptomatic patients are assumed to recover. Vaccinations carried out on the susceptible population would eliminate risk of those individuals being susceptible to infection. This model does not incorporate other parameters around social determinants of health and environmental factors influencing vectors' exposure to infection or human populations exposure to infected vectors.

26.2.1 Google Trends

Due to the limitations of this modelling system when it comes to social determinants and the massive requirements with respect to granular information, we sought unusual data sources. One novel source of data we used was Google Trends. We used a remarkably simple model for this to extremely high effect. Data was collected by searching 'yellow fever' in Google Trends over the last twelve months and then changing the geographical scope to Brazil.

Fig. 26.3 Vector-borne disease propagation model. The Black arrows represent infection status transition paths, the red dashed arrows represent transmission paths, the blue arrow represents the vaccination pathway, the Square compartments represent host classes, the circular compartments represent vector classes, the Red compartments represent infectious classes, and the gray compartments are the simulated weekly reported cases (Z_h) and deaths (Y_h). The model applied the following notations: For human host populations, S_h represents the number of susceptible individuals, E_h is the number of individuals exposed to YF but not yet infectious, A_h represents the asymptomatic (i.e., with clinically inapparent symptoms) cases, I_h the severe infectious individuals, T_h the individuals in the toxic stage, and R_h individuals have either recovered from the disease and/or have been vaccinated (or immunized by vaccination)

The data was downloaded as comma-separated values. The amount of search during certain times during which there was no outbreak was used to normalize noise. Then, the peak of the Google Trends data within a given timespan was matched to the peak in the data derived from the PAHO official case counts and Google Trends data was scaled according to coinciding peaks. Though the peaks in the datasets matched, the rest of the data appeared to differ by about one week, with Google Trends lagging. Thus, we also introduced a lag of one week.

26.2.2 Google Streetview

The PAHO dataset included the map shown in the figure, which depicts the number of confirmed human cases in various parts of the Sao Paulo province. Using WebPlot-Digitizer, we determined the angles and distances of these dots relative to a reference point for which latitude and longitude coordinates were known. We sought to predict the number of cases incident in a certain area from Streetview images of that area. It seems that a correlation could be observed, given that many social determinants of disease transmission (such as infrastructure quality) are evident in Streetview images of that area.

Following this, we extracted Google Streetview images using API calls from each of these coordinates. Around each point, we defined a grid with granularity of 0.0005°

and extracted the Streetview image at that latitude and longitude. We removed all images from areas with no Streetview imagery. We split the dataset into a training and testing split and downsampled the number of images from the area with the most images so that the classes were roughly balanced.

We modeled the problem as a classification problem. In particular, for any given image, we sought to classify it into one of four classes: as coming from a place with (a) 1 case, (b) 2–5 cases, (c) 6–30 cases, (d) 31–106 cases. The data were not granular enough to support regression (i.e., we did not know the exact number of cases in a given area but simply a range). See further details in Discussion.

We normalized the data such that each pixel had a value between 0 and 1, and then we trained an adapted version of VGGnet (Simonyan and Zisserman 2015) (added extra fully-connected layers and ended with 4 classes at the end) on around 100 images for 40 epochs with varying batch sizes.

26.3 Results

26.3.1 Google Trends

Using data from Google Trends anchored by data from PAHO’s official case counts, we developed a method for calculating the number of cases we expect to be incident, discussed in Sect. 26.2. The results of this method are shown in the figure (Fig. 26.4).

We observe that we can match *both* peaks with just a basic model, including one week of lag time (our method involves matching the largest peak, but the second peak is also well-reflected). To test whether this approach is robust, we would have to validate against other outbreaks and test the points at which it breaks (for instance,

Fig. 26.4 **a** (left) map from PAHO showing number of cases in given area. **b** (right) coordinate line added and distances and directions to points of cases marked

Fig. 26.5 Adjusted Google Trends data and PAHO case count data showing similar trends

does this approach work for outbreaks of mosquito-borne diseases in countries with similar internet penetration?) (Fig. 26.5).

26.3.2 Google Streetview

So far, due to limitations discussed in Sect. 26.4, we have not seen many promising results from this data. However, this is less an issue with its potential as a robust data source and more an issue with handling typical issues that occur while initially applying deep learning.

26.4 Discussion

26.4.1 Limitations of the Current Approach with Google Streetview

Deep learning relies on extensive datasets with around thousands of images to learn features that facilitate classification. During the early stages of this project, however, we were only able to use hundreds of images (~300). The reasons for this were related to determining the best way to choose large amounts of images from a given location

and lack of computational power to handle large image-based datasets. We used a structured sampling of a grid centered around our calculated latitude and longitude. One alternative would be defining a radius around that point and conducting random sampling. This would demand some knowledge of the bounds of the locations of these case counts.

There are also various parameters within Google Streetview which could be manipulated to get more images, including pitch and rotation. This would require more thorough examination and possibly experimentation to determine how different the images need to be in order to get good results.

Another option for dealing with the limited dataset is to use transfer learning. We could pretrain the model using the Places dataset,¹ which might help tune to features relevant to images of places, and then we could use the pretrained weights to tune to our problem. An alternate option would be training the network against census data about socioeconomics (e.g., median income bracket) and then fine tuning to case incidence. A potential issue here would be that by formalizing the connection between census data and case counts, we are studying the explicit connection between those two variables and not necessarily Streetview images.

26.4.2 Implications of the Current Approach with Google Streetview and Google Trends

If the deep learning model can be adequately trained by providing it with a large number of case information datasets matched accurately to images of different areas, the approach could help identify areas at higher risk of Yellow Fever cases. The high-risk areas could then be studied for common features including infrastructural or social determinants that lead to increased vector density and hence higher disease transmission. This can help municipal authorities create local policies and take action to improve the high-risk areas. Along with the Google Trends data, this can also help the local bodies stay prepared for an outbreak. Identifying the early phases of an outbreak can assist the authorities to launch an immediate response in terms of vaccination and preventing further spread.

26.5 Conclusion

Yellow fever is a fatal disease for which a vaccine exists, but a global vaccine shortage means it cannot reach many people in regions susceptible to yellow fever. Brazil is one such country where yellow fever outbreaks have taken many lives, with 154 deaths and 464 cases reported in the most recent outbreak in 2017–18. According to PAHO data, cases in large metropolitan regions like greater Sao Paulo and Rio

¹<http://places.csail.mit.edu/>.

De Janeiro appear to occur in various pockets spread around the region. We set out to build a disease projection model that attempted to complement traditional prediction models by incorporating additional parameters that may be informed by social determinants of health. We utilised Google trend data for simple mapping to PAHO case data, and deep learning techniques using Google Streetview image data to find associations with the case data. We believe the methodology has merit. In this paper we have proposed means of expanding the data set acquired in the study, including using more efficient techniques of extracting Streetview images. As such, this paper provides a starting point and guidelines to further explore this novel means of improving upon existing disease propagation models and use deep learning to better predict where disease outbreaks may occur.

References

- Barnett, E. (2007). Yellow fever: Epidemiology and prevention. *Clinical Infectious Diseases*, 850–856.
- Chen, Z. (2016). A fatal yellow fever virus infection in China: Description and lessons. *Emerging Microbes & Infections*.
- Goldani, L. (2017). Yellow fever outbreak in Brazil, 2017. *Brazilian Journal of Infectious Diseases*.
- Khanna, R. V. (2013). Yellow fever vaccine: An effective vaccine for travelers. *Human Vaccines and Immunotherapeutics*.
- Kotsakiozi. (2017). Tracking the return of aedes aegypti to Brazil, the major vector of the dengue, chikungunya and zika viruses. *PLoS Negl*.
- Pramil Tiwari, R. A. (2017). Knowledge and attitude of travellers reagrding yellow fever vaccination. *Indian Journal of Pharmacy Practice*.
- Simonyan, K., & Zisserman, S. (2015). Very deep convolutional networks for large-scale image recognition. *ICLR*.
- Snyder, M. (2018, March 8). *Brazil battles record yellow fever outbreak*. Retrieved from Outbreak Observatory: outbreakobservatory.org.
- WHO. (2018, February 27). *Yellow fever-Brazil*. Retrieved from Emergencies, Preparedness, Response: www.who.int/csr/don/27-february-2018-yellow-fever-brazil/en/.
- Wickramage, S. B. (2013). Is there a risk of yellow fever virus transmission in South Asian countries with hyperendemic dengue? *Biomed Research International*.

Shalen De Silva worked on data extraction and translation of data from published forms to machine-readable ones and writing, editing.

Ramya Pinnamaneni developed the context and background for the outbreak and analyzed policy implications for the findings of quantitative methods and writing, editing.

Kavya Ravichandran explored the use of computer vision techniques applied to this problem, analyzed Google Trends, and writing, editing.

Alaa Fadaq helped review the literature to contextualize the problem.

Yun Mei developed the transmission model.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.