Observing the Unobserved: A Newspaper Based Dengue Surveillance System for the Low-Income Regions of Bangladesh

Nazia Tasnim[1]
Shahjalal University of Science and Technology
embers1010@gmail.com

Md. Istiak Hossain Shihab[1]
Shahjalal University of Science and Technology
istiak@protonmail.ch

Moqsadur Rahman
Shahjalal University of Science and Technology
moqsad-cse@sust.edu

Jillur Rahman Saurav
Shahjalal University of Science and Technology
sauravsust71@gmail.com

Sheikh Rabiul Islam
University of Hartford
shislam@hartford.edu

Mohammad Ruhul Amin
Fordham University
mamin17@fordham.edu

Abstract

Dengue is one of the emerging diseases of this century, which established itself as both endemic and epidemic—particularly in the tropical and subtropical-regions. Because of its high morbidity and mortality rates, Dengue is a significant economic and health burden for middle to lower-income countries. The lack of a stable, cost-effective, and suitable surveillance system has made the identification of dengue zones and designing potential control programs very challenging. As a result, it is not feasible to assess the effect of the intervention actions properly. Therefore, most of the prevention and mitigation efforts by the associated health officials are failing. In this work, we chose Bangladesh, a developing country from the South-East Asia region with its occasional history of dengue outbreaks and with a high out-of-pocket medical expenditure, as a use case. We use some well known data-mining techniques on the local newspapers, written in Bengali, to unearth valuable insights and develop a dengue news surveillance system. We categorize dengue-news and detect the spatio-temporal trends among crucial variables. Our technique provides an f-score of 91.45% and very closely follows the ground truth of reported cases. Additionally, we identify the under-reported regions of the country effectively while establishing a meaningful relationship between complex socio-economic factors and reporting of dengue.

Introduction

Dengue is one of the emerging arboviral (arthropod-borne virus) diseases of the family *Flaviviridae*. Dengue fever can be caused by any of the four genetically related but antigenically distinct dengue virus serotypes. It spreads to humans, primarily through *Aedes Aegypti* and *Aedes Albopictus*, mostly in tropical and subtropical countries (Rigau-Pérez et al. 1998). This vector-borne disease has already established itself as both endemic and epidemic(Spaulding 2016) while becoming a major public health concern for the low to middle-income countries due to its high morbidity and mortality rate. According to WHO, an estimated 400 million people get dengue infection globally each year.

Although the virus infection in its initial stage is often asymptomatic in humans, it can cause a spectrum of clinical manifestations. The infection from one strain doesn't provide life-long immunity from the remaining three strains, and cross-strain infections are often fatal. Currently, there is a lack of effective and authorized vaccine and medicine for curing dengue. Furthermore, government efforts to control dengue transmission through vector control is far from a success, especially in the global south, to stop its rapid spread (Tatem, Hay, and Rogers 2006). Thus, the sustainable control of the dengue outbreak remains a formidable challenge to the public health officials and the policymakers around the world.

With its dense population, socio-economical demography, and environmental parameters, emerging diseases like dengue and chikungunya have become a huge economic and social burden for Bangladesh, a lower middle-income country in southeast Asia (Shepard, Undurraga, and Halasa 2013). In addition, dengue infection has become a severe health issue since its first discovery in Bangladesh in 2000, which had > 5000 cases and 1.7% case-fatality in three of the major cities (Dhaka, Chittagong, and Khulna) (Yunus et al. 2001). In a geographic location like Bangladesh, passive surveillance is insufficient to fight dengue (Low et al. 2015). Moreover, the heterogeneity in the distribution of hospitals and medical services across the country as well as the tendency of seeking medical help only when the health condition is very severe among lower income people (Sharmin et al. 2015) made the passive surveillance data biased towards selective demographics. Therefore, we think that there has been a considerable amount of under-reporting from the hospital-based surveillance, which is eventually misdirecting the dengue prevention strategies, by the government, in recent years.

In Bangladesh, the control room of Directorate General of Health Services (DGHS) collects aggregated dengue data of the patients admitted to the hospitals in the Dhaka city on a daily basis. For the districts outside of Dhaka, aggregated data is collected from the upazilla (sub-district) health complexes, and district hospitals including medical colleges (NIPORT 2017). Additionally, DGHS collects data through vector surveys three times a year. Centers for Disease Control and Prevention (CDC) is one of the functional units of the DGHS where Aedes mosquito-borne diseases such as dengue, Chikungunya, and Zika remain as sub-

[1]The two authors contributed equally to this paper.

components under the Malaria Elimination Control Program (MECP)(Mutsuddy et al. 2019).

In a few related surveillance literature, the researchers suggested the use of tweets, RSS feeds or search queries as the data-source for the system. However, countries like Bangladesh don't have many Twitter users. At the same time, most of the people do not use the native language Bengali to search the internet. Therefore, such approaches are not feasible in this use-case. Besides, there is no other reliable data-source in Bengali on the internet other than the local online newspapers that covers local disease news so thoroughly.

As dengue remains a growing threat, both locally and globally, an effective dengue surveillance system is essential to monitor the disease trends correctly, to make timely interventions, and to establish suitable policies to prevent dengue infection. But the absence of timely actions against vectors, lack of fine scale investigations, lack of surveillance, and under-reported data prompt the need to explore an alternative surveillance system. In this paper, we present a surveillance system based on the dengue related news published in Bengali language, in the local online daily newspapers and investigate its potential as a substitute surveillance system. One of the main advantages of incorporating online news sources with traditional surveillance systems is the near real-time acquisition of information that can significantly expand the quantity of data obtained from official records(Woodall 1997). In this paper, our main contributions are:

- A corpus of dengue disease and intervention related news: We collected dengue-related news from 270 online Bengali newspapers from 2017 to 2019, and developed a corpus of dengue disease and intervention related news. We analyzed and explained the corpus, and developed some classifiers;

- An online dengue surveillance system: We design and develop an online dengue surveillance system using the related news published on the disease and its intervention in the local newspaper. By analyzing the collected data from various demographic, environmental and socio-economic perspectives, we identify the local regions that are likely deprived of proper government of dengue intervention plans; and

- Filling the gap: From the analysis of the data, we reveal the gaps in the government and city-corporation intervention activities.

We make the overall approach as simple as possible, so that they can be easily embedded into robust light-weight systems.

Related Work

Newspapers can be a significant source for providing specific information relevant to disease, such as location, environmental factors, population demographics, societal factors etc (Villanes et al. 2018). Hence, data mining techniques can be deployed to perform time-series analysis of disease-specific queries to detect and monitor outbreaks (Ji, Ritter, and Yen 2017)(Zhang et al. 2018). This can be very useful particularly for the developing countries where local news media can provide better surveillance data (Schwind et al. 2014). In a study (Runge-Ranzinger et al. 2008), a systematic review of the gaps in the structure, purpose, and usefulness of dengue disease surveillance in the dengue-endemic countries

was presented. The author recommended a step-wise adaptation of surveillance in combination with active surveillance. Furthermore in another study(Runge-Ranzinger et al. 2014), Runge and his colleagues grouped surveillance systems into 3 groups for endemic countries like Bangladesh. These include, 1) outbreak prediction or detection through electronic event/search query; 2) monitoring dengue infection based on the data received from the passive surveillance; and 3) trend monitoring based on the syndromic surveillance. For example, the European Center for Disease Control (ECDC) uses VBORNET, a combination of multilingual internet scanning for disease keywords, information from entomologists and manual efforts to employ an arthropod vector surveillance for public health (Braks et al. 2011).

Similarly, Google developed a system that analyzes the disease-specific search queries performed by common people on a daily basis for the early detection of Influenza(Ginsberg et al. 2009). Another paper discusses an approach of accurate dengue tracking by combining AutoRegressive model with Google search queries as exogenous variables (Yang et al. 2017). HealthMap - a well-known web based surveillance tool collects and aggregates data from RSS feeds, ProMed alerts, GoogleNews, etc for outbreak monitoring (Freifeld et al. 2008). SourceSeer (Rekatsinas et al. 2015) is another such tool which uses topic modelling to forecast outbreaks. Sentinel, also a syndromic surveillance mechanism, employs CNN and LSTM on twitter data to classify tweets and discover associated information and patterns (Serban et al. 2019). These types of digital surveillance are of great importance to under-developed countries which are often constrained by budget and man-power. It can be used as a proxy to monitor outbreak information, while at the same time being a potential supplement to early outbreak detection mechanisms.

Dataset Preparation

We collected data from multiple sources. Dengue related records were collected from Access to Information Programme (a2i), CDC, IEDCR and DGHS. Together they run a country-wide program to keep track of dengue patients along with some other tropical diseases (Mutsuddy et al. 2019). As the official dengue reports are not available prior to 2017, we decided to use the similar time period, from January 2017 to December 2019, for collecting the online news. To collect the archived daily news, we used the specialized web crawler of the Bengali Search Engine, Pipilika. To evaluate the proposed system, we performed a time-series analysis of both the officially tracked data and newspaper archive to find the similarities and missing information in the officially published data. We also collected additional data from Bangladesh Bureau of Statistics (BBS) yearbook and The World Bank to analyze patterns associated with newspaper data and the socio-economical situation of different parts of the country.

Furthermore, we follow a systematic approach to filter this massive news corpus for the selection and classification of dengue related news. We selected any news that contain the dengue related words and its chain of infection(CoI) keywords. We performed a thorough preprocessing of the data including the removal of stop words, URLs and special characters to standardize the news dataset. As we're primarily interested in the identification of the degree and deficit of intervention actions in different parts of the country, we separated
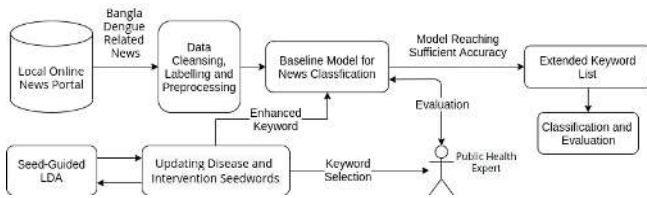
Figure 1: Flowchart of the dengue-news classification method.

the intervention related news and labeled them as a specific class called, *intervention*. Remaining dengue news data were classified in a generic *disease* class, which contains dengue infection and patient related news. After analyzing the dataset under the supervision of a public health expert, we picked our primary keywords sets, for each of the aforementioned classes, based on the CoI keywords. Then, to discover any latent topics within the dengue dataset and expand our keywords sets, we performed a seed-guided topic modeling using the seed words from CoI and guided LDA library. This is the most important part of our methodology. Finally, we separated 1,500 dengue news by random sampling, and labelled them into the mentioned classes. These data were labeled to perform a semi-supervised text classification to differentiate between *disease* and *intervention* related news. The data labelling has been performed by the first two authors of the paper with a continuous consultation of a public health expert.

Methods

In our attempts to develop a surveillance system, we first choose to classify all our available newspaper articles into two classes. One of which is *Intervention Class* containing relevant reports with contexts that indicates cleaning activities and programs, information on transmission and breeding of dengue disease, mosquito breeding spots, unhygienic places suitable for vector habitats, infestations in slums etc. And the other one is *Disease Class* which includes any report containing news of the targeted disease (dengue case count, death ratio, symptoms, daily official reports, virus evolution, epidemiology, serology etc.) that doesn't belong to intervention class.

In Figure 1, we present the flowchart of our classification approach. A major part of the classification task is to extend our initial seed keyword list of *disease* and *intervention*. We use the technique of human-in-the-loop to extend the seed word list, using Guided LDA (Jagarlamudi, Daumé III, and Udupa 2012) and our naive baseline model (a combination of Jaccard and Cosine similarity metrics).

We used the guided LDA model from the guided LDA library (Singh 2017) with 80% seed confidence for the initial seed words. From the top 2,000 keywords of each topic or class (total 4,000), around 60 relevant new terms were chosen, which ultimately expanded the keyword set by $\sim 1.5x$ times. This was done by randomly producing a keyword set of different length for each class, running them on the baseline model, and calculating the accuracy. We continue this process until we achieve a performance threshold for which the last improvement was $< \epsilon$. The final keyword set improved the accuracy of the baseline model from 78% to 82.4% .

Finally, we used the enhanced keyword set as vocabulary

to the tf-idf and count vectorizers and fit the news representation on our dataset. We then separately trained three different models for each representation on these news vectors and evaluated them using standard machine learning metrics. We chose three widely used classification approaches, 1) K-nearest neighbor (KNN), 2) Multinomial Naive Bayes, and 3) Support Vector Machine (SVM). According to our experiments, SVM classifier together with CountVectorizer outperforms other models. For the SVM classifier, we have further used a Binary classification approach with on-average 89.55% F1-score and a OnevsRest classification approach with 91.45% F1-score.

Results

To find the viability of using newspaper surveillance in Bangladesh, we explore the various socioeconomic and demographic factors associated with dengue incidence. We applied the classification methodology described in the previous section on a large corpus of approximately 38,000 news articles. We used the output to explore the relationship between local newspaper reports and aforementioned factors. Thus, we try to shed light on the effect and deficit of existing dengue control mechanism.

Previous studies suggest that the risk of dengue transmission varies over space and time (Banu et al. 2012). Therefore, identification of high-risk areas can be useful for prioritising dengue surveillance and vector control efforts in areas where they are most needed. Bangladesh has 64 districts with 8 divisions (i.e., states). In Figure 2, dengue news reported on local and national online portals and the official dengue count is shown for each of these districts for the year 2019. In the plot, it is visible that the divisional districts, such as Dhaka, Chittagong, Rajshahi, Sylhet, Khulna, Barisal, Mymensingh and Rangpur, tend to have a higher number of news reporting than other districts. These districts are known as the economic hubs of the country. This is likely due to the general practice followed by newspapers, which predominantly focuses on the issues regarding their core urban readers. Additionally, dengue cases and news frequency both have a significantly high Pearson correlation score with the population of each district, but the R^2 value of respective relations are quite low (*69% and 82%*). We choose the number of institutional households and the number of general households per district as a measure for the socioeconomic status of each region, and these two factors have significant R^2 and Pearson value with news frequency (*82% and 91% , 66% and 81%*). This analysis shows us that even though it is usually expected that news frequency will follow the areas with more population, but in reality, the economical importance of an area also plays a major role. Consequently, the high correlation and R^2 score for the dengue cases, and the news count per district tell us that newspapers can be a potential source for surveillance system for the entire country (*88% and 86%*).

In a more granular level, the number of officially recorded dengue cases reached 53,316 in the capital city Dhaka, which was higher than all other districts combined. Dhaka is subdivided into 40 different areas or thanas. According to a recent survey (brac), the south-western parts of the city have a high number of unhygienic slums and dense population with poor living arrangements. These are the areas where we expect a high number of dengue cases. However, from the official IEDCR dengue case distribution map presented in the Figure
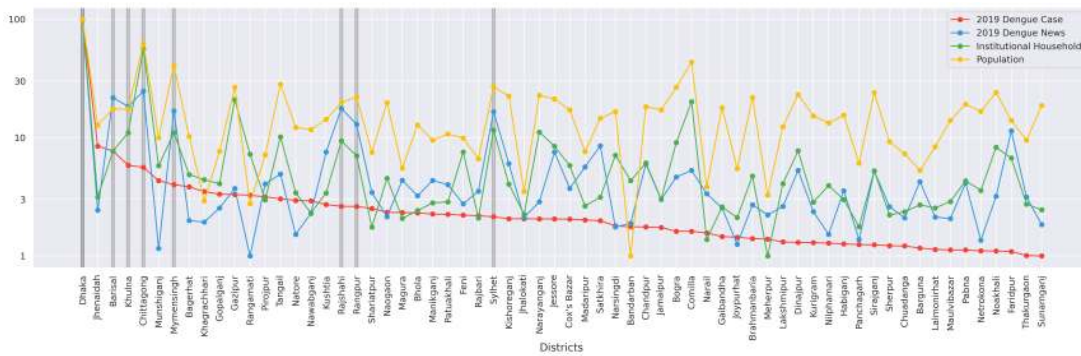
Figure 2: Spatial and socio-economic characteristics of dengue related news.
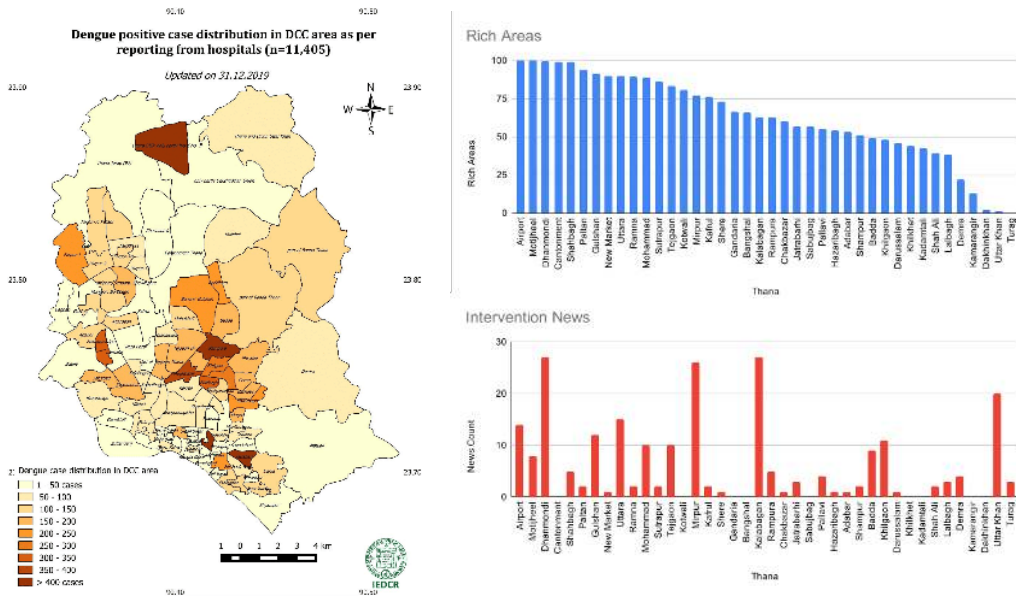

Figure 3: Top: Plot of official dengue case distribution map of 2019 from IEDCR. Middle: Thanas (sub-district) of Dhaka sorted by wealth percentage obtained from World Bank. Bottom: Intervention news frequency for Dhaka thanas (sub-district) in 2019, sorted as middle plot.

3, we see that dengue cases cluster around the centre (Gulshan, Banani, Baridhara, Tejgaon) which is the economically important portion of the city as seen in Figure 3. From the middle plot of the same figure, we can observe that majority of the intervention actions were also taken primarily for the more wealthy areas where we already expect the dengue case to be less. Uttar Khan, can be considered as an outlier in this case which got sufficient notice from the authority, despite its minor infection rate and economical status. Based on these observations, we can definitively claim that the official dengue control efforts are not properly channeled and many areas get overlooked due to the socioeconomic status.

Figure 4 represents the effect of previous years' intervention reports frequency on the disease report frequency of the following year, in different thanas of Dhaka. We assume that disease related news represents the dengue situation of an area and intervention related news represents the intervention steps taken for that certain area. In the top plot, we see the effect of 2017's intervention actions as reflected by dengue reports in 2018. Bottom plot represents the same theme for 2018-2019. The red arrows show the thanas where it looks like intervention efforts have been useful. In both years, the result is positive for Gulshan, Kalabagan and Khilgaon. On

the other hand, despite high intervention reports, areas like Dhanmondi, Mirpur, Uttara, etc don't see much success. As thana-wise dengue reports are not officially recorded, this could be an effective way to track and monitor the disease on a more granular level.

In Figure 5, we explore the aforementioned argument on a national scale in comparison with the officially available district-wise dengue case counts. We observe that while the divisional districts get well represented in the newspapers, we can also identify which regions didn't receive enough attention from the authority in relation with the reported dengue cases such as: Jhenaidah, Rangamati, Munshiganj and Natore (marked with red arrows in the Figure 5).

Discussion and Conclusion

The determinant of dengue transmissions are driven by multitude of factors that create the patterns and dynamics of prevalence and incidence of dengue. But, due to the absence of a national mechanism for tracking dengue patients in granular neighborhood levels of Bangladesh, and the entanglement of associated factors have made it very complicated to design and implement a inexpensive surveillance system. The exist-
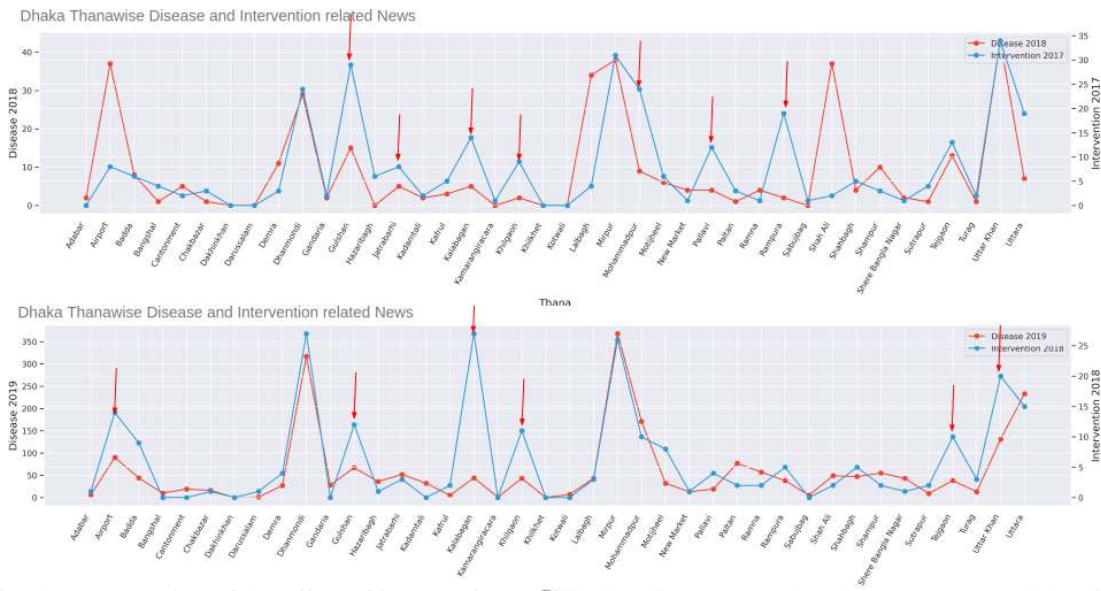
Figure 4: Visual representation of the effect of intervention events (blue line) reported on the newspaper and the disease reports (red line) the following year for the thanas of Dhaka city. The data from 2018 -2019 (bottom) shows that the intervention in 2018 wasn't very fruitful, because the same areas had high disease reporting in 2019. However, the intervention efforts in 2017, seems to have worked partially in 2018 with the red arrows showing in both figures, the representative areas, where intervention measures likely made some effect.
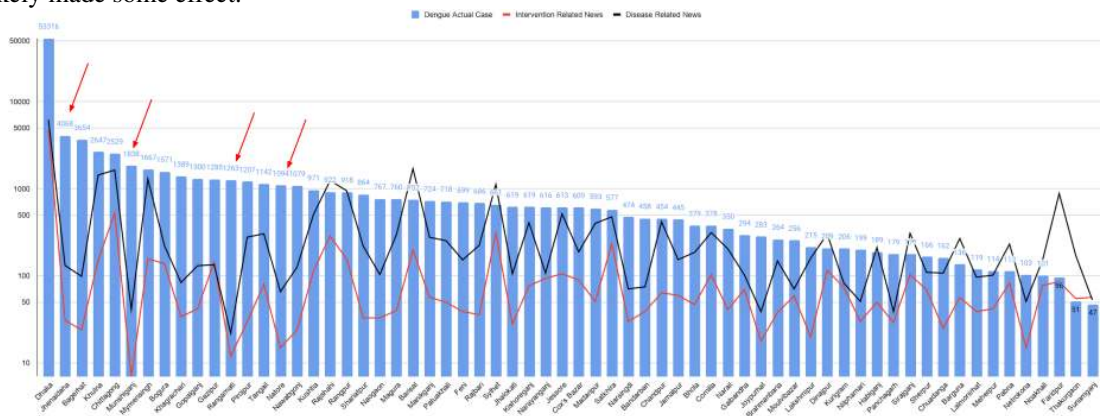

Figure 5: Nation-wise published dengue-related news report in comparison with officially reported dengue case (sorted)

ing health service infrastructure suffers from disease underreporting due to its heavy reliance on passive surveillance.

In this article, we present a data driven approach to develop a dengue surveillance system based on the dengue related news published in the local newspapers. We have taken several climatic, demographic, and socioeconomic variables into account, and analyzed their individual and combined effects on dengue incidence. We also identified the knowledge and the data gap in these analyses to investigate whether newspapers can be considered as an alternative source of surveillance. We summarize the research findings as follows:

- We can use a readily available machine learning approach to identify the intervention and disease related news with sufficient accuracy
- Dengue news and dengue official report follow similar temporal pattern. However, this correlation isn't strongly visible on the spatial axis (Figure 2,5). This difference allows newspapers to be an alternative source for monitoring the dengue trend across the country.
- Our analysis shows that the intervention actions taken by the health authorities are not being targeted properly, and

thus many of the regions across the country remain overlooked (Figure 5). The existing mechanism is biased and affected by prevailing socioeconomic status. This effect is visible on both the district and thana level (Figure 5 and Figure 3).

In a few related surveillance literature, the researchers suggested the use of tweets, RSS feeds or search queries as the data-source for the system. However, countries like Bangladesh don't have many Twitter users. At the same time, most of the people do not use the native language Bengali to search the internet. Therefore, such approaches are not feasible in this use-case. Besides, there is no other reliable data-source in Bengali on the internet other than the local online newspapers that covers local disease news thoroughly. We also tried to make the overall approach as simple as possible, so that they can be easily embedded into robust and light-weight systems.

Our approach comes with some limitations too. First, we were unable to extend our observation to more than past three years. News archive data before 2017 is not substantial and reliable, particularly in areas outside the Dhaka city. From

such data, we have observed that the correlation between news data and dengue report seems arbitrary in the case of Chittagong city (Port city and most important economic hub of the country), for the year 2018, which can be attributed to a lack of substantial official data for Chittagong in that year. Secondly, we didn't have access to all the official government records and much of the data we used, were collected from third-party sources. However the good news is that the amount of data is increasing and improving significantly year to year. This means the system will only get better, stable and more meaningful as time passes.

In conclusion, Bangladesh, being a lower middle income country, where it is not possible to implement a nationwide infrastructure for dengue surveillance system, a newspaper based surveillance system can be an inexpensive alternative for monitoring dengue and other endemic diseases locally.

References

Banu, S.; Hu, W.; Hurst, C.; Guo, Y.; Islam, M. Z.; and Tong, S. 2012. Space-time clusters of dengue fever in bangladesh. *Tropical Medicine & International Health* 17(9):1086–1091.

brac. Urban slum map. http://urbanslummap.brac.net/slum_map.html.

Braks, M.; van der Giessen, J.; Kretzschmar, M.; van Pelt, W.; Scholte, E.-J.; Reusken, C.; Zeller, H.; van Bortel, W.; and Sprong, H. 2011. Towards an integrated approach in surveillance of vector-borne diseases in europe. *Parasites & vectors* 4(1):192.

Freifeld, C. C.; Mandl, K. D.; Reis, B. Y.; and Brownstein, J. S. 2008. Healthmap: global infectious disease monitoring through automated classification and visualization of internet media reports. *Journal of the American Medical Informatics Association* 15(2):150–157.

Ginsberg, J.; Mohebbi, M. H.; Patel, R. S.; Brammer, L.; Smolinski, M. S.; and Brilliant, L. 2009. Detecting influenza epidemics using search engine query data. *Nature* 457(7232):1012–1014.

Jagarlamudi, J.; Daumé III, H.; and Udupa, R. 2012. Incorporating lexical priors into topic models. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, 204–213.

Ji, X.; Ritter, A.; and Yen, P.-Y. 2017. Using ontology-based semantic similarity to facilitate the article screening process for systematic reviews. *Journal of biomedical informatics* 69:33–42.

Low, S.-L.; Lam, S.; Wong, W.-Y.; Teo, D.; Ng, L.-C.; and Tan, L.-K. 2015. Dengue seroprevalence of healthy adults in singapore: serosurvey among blood donors, 2009. *The American journal of tropical medicine and hygiene* 93(1):40–45.

Mutsuddy, P.; Tahmina Jhora, S.; Shamsuzzaman, A. K. M.; Kaisar, S.; and Khan, M. N. A. 2019. Dengue situation in bangladesh: An epidemiological shift in terms of morbidity and mortality. *Canadian Journal of Infectious Diseases and Medical Microbiology* 2019.

NIPORT, I. 2017. Bangladesh health facility survey 2017. https://dhsprogram.com/pubs/pdf/SPA28/SPA28.pdf.

Rekatsinas, T.; Ghosh, S.; Mekaru, S. R.; Nsoesie, E. O.; Brownstein, J. S.; Getoor, L.; and Ramakrishnan, N. 2015. Sourceseer: Forecasting rare disease outbreaks using multiple data sources. In *Proceedings of the 2015 SIAM International Conference on Data Mining*, 379–387. SIAM.

Rigau-Pérez, J. G.; Clark, G. G.; Gubler, D. J.; Reiter, P.; Sanders, E. J.; and Vorndam, A. V. 1998. Dengue and dengue haemorrhagic fever. *The Lancet* 352(9132):971–977.

Runge-Ranzinger, S.; Horstick, O.; Marx, M.; and Kroeger, A. 2008. What does dengue disease surveillance contribute to predicting and detecting outbreaks and describing trends? *Tropical Medicine & International Health* 13(8):1022–1041.

Runge-Ranzinger, S.; McCall, P. J.; Kroeger, A.; and Horstick, O. 2014. Dengue disease surveillance: an updated systematic literature review. *Tropical Medicine & International Health* 19(9):1116–1160.

Schwind, J. S.; Wolking, D. J.; Brownstein, J. S.; Consortium, P.; Mazet, J. A.; and Smith, W. A. 2014. Evaluation of local media surveillance for improved disease recognition and monitoring in global hotspot regions. *PLoS One* 9(10):e110236.

Sharmin, S.; Glass, K.; Viennet, E.; and Harley, D. 2015. Interaction of mean temperature and daily fluctuation influences dengue incidence in dhaka, bangladesh. *PLoS neglected tropical diseases* 9(7):e0003901.

Shepard, D. S.; Undurraga, E. A.; and Halasa, Y. A. 2013. Economic and disease burden of dengue in southeast asia. *PLoS Negl Trop Dis* 7(2):e2055.

Singh, V. 2017. Guidedlda: Guided topic modeling with latent dirichlet allocation. https://github.com/vi3k6i5/GuidedLDA.

Spaulding, S. 2016. Dengue fever: Endemic to epidemic. https://yaleglobalhealthreview.com/.

Tatem, A. J.; Hay, S. I.; and Rogers, D. J. 2006. Global traffic and disease vector dispersal. *Proceedings of the National Academy of Sciences* 103(16):6242–6247.

Villanes, A.; Griffiths, E.; Rappa, M.; and Healey, C. G. 2018. Dengue fever surveillance in india using text mining in public media. *The American journal of tropical medicine and hygiene* 98(1):181–191.

Woodall, J. 1997. Official versus unofficial outbreak reporting through the internet. *International journal of medical informatics* 47(1-2):31–34.

Yang, S.; Kou, S. C.; Lu, F.; Brownstein, J. S.; Brooke, N.; and Santillana, M. 2017. Advances in using internet searches to track dengue. *PLoS computational biology* 13(7):e1005607.

Yunus, E. B.; Bangali, A. M.; Mahmood, M.; Rahman, M. M.; Chowdhury, A.; and Talukder, K. 2001. Dengue outbreak 2000 in bangladesh: From speculation to reality and exercises. *Dengue Bulletin* Dengue Bulletin. 2001 Dec;25:15–20.

Zhang, Y.; Ji, X.; Ibaraki, M.; and Schwartz, F. W. 2018. Mining information from collections of papers: Illustrative analysis of groundwater and disease. *Groundwater* 56(6):993–1001.

Şerban, O.; Thapen, N.; Maginnis, B.; Hankin, C.; and Foot, V. 2019. Real-time processing of social media with sentinel: A syndromic surveillance system incorporating deep learning for health classification. *Information Processing & Management* 56(3):1166–1184.