Big Data Analytics for Disease Outbreak Forecasting

Tresna Dey

Abstract

Since decades the concept of public health in Bangladesh was characterized by response. Hospitals and clinics have always been the initial to raise a red flag about an outbreak, be it dengue during the monsoon season, influenza during the winter months or chikungunya gradually invading urban areas. By the alert arrival, the wards of Dhaka medical college, the Chittagong general hospital or other facilities in the region were already filled with patients. However, in approaching 2026 in Bangladesh, the mindset is shifting towards the anticipatory to the reactive care.

Together with Big Data Analytics and Deep Learning, the public health authorities are constructing a digital early warning system that does not just follow the outbreaks, it forecasts them. This change is necessary in the country with 170 million people because this movement, the number of people, and the environmental situation contribute to the creation of the interconnected ecosystem where diseases spread rapidly. Big data is no longer a technical instrument in such a situation, but a very important instrument of preventative health.

Conventional surveillance systems that were used in Bangladesh to monitor epidemic diseases are essential but do have shortcomings. They usually depend on manual reporting by hospitals and clinics and data takes days or even weeks to be reported. By the time aggregated information is transmitted to those who will take action on the information, the disease might have become very widespread especially in overcrowded urban slums or in isolated rural areas. On the other hand, predictive analytics incorporates a variety of data streams as they come in. Health logs, weather, people movements, environmental monitoring and even social media trends can be processed concurrently to develop actionable information. Essentially it is aimed at moving a model of responding to the past into one that predicts the future.

A good case study is Dengue fever. A seasonal issue that was only experienced during the monsoon months, dengue has become an issue that poses a threat to the general public throughout the year. The 2025 case and death toll of over 102,000 and 412, respectively in Bangladesh showed that infections are more and more

emerging beyond the conventional July-August high. The use of big data has seen researchers use models like Prophet, XGBoost and hybrid ensemble models- also known as CXGDBoost- to make dengue surge predictions with impressive accuracy. These models currently have forecast accuracies of more than 94 per cent by examining thirteen meteorological variables such as lagged rainfall, temperature minima, humidity, and the speed of the wind. This accuracy will enable the health authorities to know about the outbreak days before it occurs so that their response can be specific and efficient before numbers of patients congest the hospitals. The ongoing projection of the year 2026 shows that in case of transmission throughout the entire year, the case surges in early years will reach up to 25 percent than the past years, and most importantly, they will peak sooner than the trend in history would show.

Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRU) are also Deep Learning architectures that increase the predictive advantage. These architectures are used to identify intricate nonlinear interactions between many streams at the same time compared to the simpler regression models that focus on variables individually. An example is that human mobility information based on anonymized cellular signals can be combined with environmental predictors (e.g. stagnant water pools identified by satellite images or flood-prone areas). Weighted ensemble models based on SARIMA, XGBoost, and recurrent neural models can be used to make predictions that are much better than those made by individual models to give the health authorities up to seven days lead time to get the required resources on board. This window though brief is usually the difference between controlled containment and exponential transmission.

The redefinition of resource planning by big data analytics also takes place during an outbreak. Predictive models are not just a prediction of numbers, they are actionable. In the event that a model projects a twenty-percent respiratory surge in Chittagong, the Ministry of Health will be able to position oxygen cylinders, pre-stock medicines, and train hospital workers on surge capacity. In the same way, DNCC or DSCC can introduce a more cost-effective and effective target by having the vector-control teams target the high-risk areas of Dhaka North City Corporation or Dhaka South City Corporation through a publicity on high-risk areas that should be targeted by fogging campaigns in Dhaka. In such a manner, predictive analytics will turn scattered, lagging data into preemptive actions, which will save lives directly.

Although there is this progress, there are still difficulties. Bangladesh experiences a Data Paradox: immense amounts of data have existed, but they are not distributed between ministries, departments and private operators. The inability of interoperability between the Meteorological Department, the Directorate General of Health Services (DGHS), and telecom providers limits the opportunities of predictive models. Live mobility data, such as that, is priceless in predicting the disease transmission, yet it needs to be made accessible and integrated into a system with the help of clear legal frameworks and data-sharing agreements. There are also policy gaps, and experts identified them at the end of 2025, arguing the need of solid governance frameworks to balance citizens privacy, and health needs of the population. Lacking trust and regulation, big data will be viewed as surveillance tool, and not as a service tool, which can decrease

the participation and compliance of the community.

Moral issues are as well very important. Predictive models have functionality based on individual health information, environmental monitoring and anonymized behavioral inputs. Abuse or leakage of such data may increase social stigma especially in the rural society where there are many misconceptions regarding the disease. Responsible deployment requires the implementation of end-to-end encryption, consent protocols, and clear policies regarding data-use. In addition, model misclassifications should be taken under control. False positives- detecting an outbreak that is not there- may create panic and misrepresentation of resources, whereas false negatives may put communities in danger of uncontrolled outbreak. The algorithmic suggestions should be accompanied by human control in any case.

The future of Bangladesh health system is an anticipatory preparedness. The data repositories would have the ability to integrate clinical records, environmental sensors, satellite imagery, and even community-reported signals and act as centralized data repositories, also known as National Health Data Lakes. Leveraging a centralized access and standardization of data format, policymakers are able to have an overall view of the real-time emerging health threats. It is also important to train new generation of public health data scientists. Such professionals should be able to convert the raw computational output into actionable policies and this is the gap between code and community impact. Also, the data collection should not be confined to the urban areas, the low-tech proxy indicators in the rural chars, the river islands, and the informal settlements will help to make the predictive systems inclusive and fair.

Additionally, community engagement is still one of the pillars of anticipatory health systems. Digital sensing should not be applied only in elite institutions and wealthy districts. Grassroots data can be fed into the predictive engine by community health workers who have mobile reporting applications, neighborhood surveillance networks, and participatory citizen science platforms to provide the data. This makes sure the “digital oracle can take into consideration the whole population, the marginalized and vulnerable populations that the traditional surveillance usually ignores.

Predictive analytics combined with the policy of the public health is a cultural change. Historically, outbreaks of diseases were regarded as unavoidable and responded to symptomatic patients. The use of data-driven approach today enables proactive intervention: it is possible to predict dengue epidemics, avert influenza waves, and prevent the hotspots of cholera. Dynamically, the hospitals may set the staffing and bed distribution and the vegetable-control and vaccination programs are under the guidance of accurate predictions. The outcome is a health system that ceases to be responsive in its mode of existence and becomes proactive.

To sum it up, Bangladesh is also passing into the phase where the sudden epidemics will become a rarer phenomenon. Machine learning and big data offer more than just a view into the future: they offer a

platform to act in concert and in an evidence-based manner. The country is developing a knowledge wall against outbreaks with the help of various data flows, predictive models integration into governance, and investments in inclusive infrastructure and experience. Fevers might still shoot up, but predictive systems provide the health authorities with the ability to anticipate and prepare and respond, making public health less of an active task and more of a life-saving operation. The collaboration of human knowledge and virtual foresight in this new paradigm gives the birth of an entirely anticipatory health system, which gives every Bangladeshi the best opportunity possible to be safe and well and to survive.