Social Media Misinformation Detection Using NLP

Tresna Dey

Abstract

One of the posts in a large Facebook group during the dead of a Friday morning in 2025 stated that a routine childhood vaccine was based on an experiment by foreigners. The post was shared over 15,000 times on both WhatsApp and Messenger in the first six hours spreading through group chats in villages and market towns. By afternoon dozens of rural Sylhet parents had resolved not to attend their appointment in the local Upazila Health Complex. The rumor was not substantiated by any laboratory. It was not supported by any scientific authority.

Nevertheless, its pace and emotional appeal made it even more effective than any official press release. This is the morphology of the contemporary infodemic - a kind of parallel epidemic of misinformation, which diffuses even faster than pathogens themselves. With Bangladesh being projected to progress through 2026, the battle against 2026 is no longer played in clinics and hospitals; it is played in comment threads, forwarded messages, and promoted newsfeed as it is perceived visually on a mass scale.

Falsehood in Bangladesh has become an infrastructural health threat to the populace. However, a report published in 2026 by Dismislab showed that false or misleading content on the Internet increased by 30 percent in 2025 alone according to the fact-checkers. A large portion of this growth was in the form of health related rumors, including fake cures, or conspiracy theories about immunization campaigns. In a nation where social media has been heavily embraced, and smartphones are common even in isolated locations, digital rumors have the ability to reach the entire nation within a few hours. Its impact is real: tackling treatment late, vaccine scepticism, mass purchasing of untested drugs, and undermining of trust in the medical community.

There is no abstract cost of misinformation. In the later parts of the pandemic period, there were viral rumors of miracle treatment and home treatments. There were rumors at an early stage that the leaves of Thankuni (Indian pennywart) when consumed and with particular prayers recited could cure the viruses. The story was interwoven with the traditional knowledge of the herbs with the undertones of religion,

giving it an emotional authority. The myth caused unplanned meetings in certain areas which defied the government imposed health guidelines compromising the social distancing. Worse still, there were also other counterfeit claims which propagated the intake of unregulated or toxic substances. In the period of the COVID-19 crisis, a study carried out in a region estimated that health misinformation led to over 5,800 hospitalizations because of negative responses to so-called miracle cures. These cases put a strain on already stretched hospitals and diverted resources to real cases of emergency.

The main problem of the misinformation crisis is the increasing lack of trust. Under the condition of weak institutional trust, rumors fill the void. An example is a 2025 youth poll commissioned by UNICEF in Bangladesh that found that two out of three youth reporting had 'fake news' as their number one cause of stress on social media platforms. This is not just a psychological stress but a stress that affects the behavior. Research has established that repeated exposure to unproven health claims has a serious negative impact on the chances of parents vaccinating their children. In societies that have lacked trust in the societal institutions, cultural fears, and emotional vulnerabilities, rumors propagated digitally gain advantage on such weaknesses, and uncertainty is met with suspicion.

In order to address this challenge, scholars and policy makers are increasingly making use of Natural Language Processing (NLP), a sub-field of artificial intelligence that helps computers to comprehend and interpret human language. This is a peculiar task in the context of Bangladesh. Bangla is a linguistically rich language which is morphologically shaded and often blended with English in day to day internet communication. Slang, emojis, phonetic spellings, and code-mixed phrases, which can not be filtered through simple keywords, are found in posts. The traditional moderation instruments that have been trained on English-language templates fail to identify minor tendencies of misinformation that are concealed in the Bangali discourse.

Recent studies are aimed at constructing Bangali-specific NLP pipelines that can identify malicious health discourses. Bangla-BERT has become one of the best-performing transformer-based language models with an accuracy rate of about 92.63 percent in the process of identifying health-related misinformation-significantly better than more traditional machine learning models like Random Forest models with a maximum accuracy of about 76 percent in identifying misinformation. Transformer architectures, unlike the keyword-based filters, examine contextual word-word relations, and are capable of grasping a semantic nuance and intent. An example of this can be the worried parent who raises the question, Is this vaccine safe to my child? is different to a coordinated post that says, this vaccine is an experiment that is kept secret-spread word before they take this down! Advanced systems of NLP are able to recognize the difference between inquiry and incitement through a syntax, sentiment, and contextual embedding.

In addition to the interpretation of texts, NLP systems are also able to detect the structural patterns within the domains of misinformation networks. In Bangladesh, rumors are spread in the form of chain messages

with the most common being spread using the encrypted platform of WhatsApp. Some of these messages usually include linguistic fingerprints: calls to an urgent action, emotional capitalization (URGENT!), allusions to unknown insiders, or references to religious authority. With such recurring rhetorical markers, researchers can identify potentially harmful discourse early before going viral by training models to identify such signs of rhetorical patterns. Pattern recognition is not limited to language, but also dissemination behavior, and organized inauthentic activity, repetitive phrasing in many different accounts, and unusual sharing rates may indicate amplification campaigns run by bots.

Notably, the aspect of implementing NLP in the field of public health governance is not censorship but resilience. A proactive approach considers the AI systems as an early warning radar. Suppose a real-time Myth-Buster Dashboard which sends a warning notification to the health officials when a rumor about Vitamin A capsules starts trending in Mymensingh. This type of system might facilitate specific countermeasures: radio announcements in the local language, SMSs clarifying something, briefings of community health workers, quick-response posts on social media targeting the particular claim. Authorities can pre-bunk falsehoods instead of debunking them once they have been said by acting early, before the misinformation penetrates the realm of public discourse.

The best way of restoring social trust is through transparent communication. On occasions when AI systems identify content as misleading, there should be explanations. The system could show that a particular claim is inconsistent with proven data by recognized medical authorities or based on already fact-checked stories instead of simply declaring a post to be false. This openness will decrease the views of arbitrary censorship and bring technology out as a shield instead of a sword. Over time, such practices will help regain faith in mass health communication.

It is proposed that behavioral nudges can help avert the spread of misinformation to a large extent. In a study conducted in 2026, users who received AI-generated alerts that they were wrong progressed by 40 percent to post suspicious health content. These notifications never blocked data, they just created friction - making the user ask themselves whether to send or not. Effectively, NLP is the immune system of the digital ecosystem, which detects the pathogens in the information stream. However, immunity goes beyond detection. The long term solution is the digital literacy. The users need to acquire the critical thinking skills they will require to sift through claims by themselves, especially when it comes to emotionally charged situations.

In the future, the vision of a smart Bangladesh should include an efficient computational linguistics framework. Local datasets are the most important to be invested in. Online youth communication is dominated by code-mixed language -Bangla mixed with English. Lacking representative datasets that represent these hybrid forms, NLP models are under threat of being blind. Dialect and socioeconomic diversity can be better trained by continuous retraining of models that use regionally diverse data.

There are also ethical considerations that are urgent. The misinformation detection systems should be audited to detect bias so that it will not excessively detect the voices of minorities or valid scientific disagreement. It is necessary to have transparency in the design of models, independent control, and clear requirements on the classification of content. Extreme moderation may destroy the same trust which these systems are meant to influence. Living with freedom or expression and protection against harm necessitates delicate frameworks of governance that are guided by legal knowledge, technological knowledge and sociocultural knowledge.

The effectiveness of this digital defense will be determined by the cross-sector cooperation. Anonymized data and best practices need to be shared by technology companies, telecommunications providers, public health agencies, as well as fact-checking organizations. Rumor Scanner is already a platform that helps to debunk viral claims in Bangladesh. Their knowledge can be introduced into automated NLP systems and improve the speed of detection cycles. Schools, in their turn, will be able to optimize algorithms and measure performance rates and make sure that models will adjust to the changing misinformation strategies.

Finally, social media misinformation is no longer a minor technological issue. It is a determinant of health outcomes. A viral lie can sabotage the work of vaccination programs, scare away those who want to visit the hospital, or create dangerous self-medication in hours in an interconnected society. Capability to make out believable data and viral trickery may be the most vital health device in a Bangladeshi family. NLP will provide the institutions with digital ears that can differentiate between signal and noise. Its success, however, lies in human custodianship ethics, openness in its government, and long-term involvement of the people.

With Bangladesh in the various twists and turns of digital transformation, the fact that NLP has become a part of the national approach to its public health is a need and a chance. Early detection of misinformation, specific communication, and development of digital literacy will help the country strengthen its protection against the infodemic. This is not intended to make people stop talking, but to lift the truth. By so doing, Bangladesh is asserting that in the digital world, information must also be kept safe in order to secure the health.