

Chapter 1

Introduction

The novel Coronavirus, SARS COV-2, was recognized for first time in early December 2019, and according to World Health Organization (WHO), until September 30th, 2022, more than 600 million individuals have been infected with the virus, resulting in around 6.5 million deaths worldwide¹. The daily reports of new infections and deaths have been fluctuating, and the numbers differ from country to country. COVID-19 virus is spreading rapidly and causing widespread respiratory problem. Due to this people suffer from extreme breathing problem and body oxygen level decrease severally. This can increase the likelihood of organ damage and even death [81, 38, 192, 114]. The WHO has designated it as one of the world's largest and deadliest pandemics [1, 206]. As the pandemic continues, the number of cases increases and patients have experienced severe breathing difficulties, resulting in a significant amount of deaths; so, there have been various causes to be worried regarding this viral outcomes [73]. Each country has followed distinct procedures and done various precautionary actions to avoid the transmission of COVID-19.

The virus has different effects on different people. Most people start out with a cold and fever, but then they get serious breathing problems, feel tired and sore, and lose their sense of taste and smell. The virus has mutated into a large number of variants, and each one grows more severe as a new variety emerges.

The virus spreads when an individual comes in contact or does not maintain physical distance with an infected person. When a patient yawns, sneezes, coughs, speaks, sings or breathes, it transmits from the mouth or nose. Several guidelines have

¹<https://covid19.who.int/>

been advised:

- Wear your mask to ensure that it covers both your nose and mouth whenever you step out of the house.
- Properly wash your hands.
- Sanitize yourself.
- Avoid eating or drinking anything cold.
- Eat nutritious foods to strengthen your immune system.
- Maintain a physical distance while interacting with a group of people.
- Avoid all forms of direct physical contact.

Taking care of oneself involves caring for others as well. It would be a lot simpler to stop the spread of the virus if everyone aware of the consequences this sickness might cause. Researchers and medical professionals have devoted significant effort to create COVID-19 vaccines. While the vaccination is going on several people also experienced a few side effects, such as fever, pain, and weakness. It is important to note down that getting vaccinated is not permission to walk around without masks and engage in personal contact with strangers. Periodically, new strains of the virus have emerged, and the severity of the sickness has increased with each new strain. Only through collaborative efforts we will be able to stop the spread of the virus.

1.1 Observation and Motivation of the works

As the COVID-19 pandemic continues to unfold, there is an escalation in cases and an increasing number of patients facing severe respiratory issues, leading to a significant loss of life. This creates a pressing need to address and find solutions for the problems associated with COVID-19.

In this scenario, pandemic has become a worldwide priority for research on the subject. However, in order to make innovative experimental data or theoretical ideas available to a wide group of users, it is important to sort through individual publications for relevant information. Researchers in several domains, including

physicians, epidemiologists, and policymakers, must have access to the most recent articles in their respective domains. With thousands of new articles and research papers being published every week on the virus, it can be overwhelming and laborious to manually sift through and identify relevant information. Automated classification systems can help expedite the research process by quickly and accurately categorizing articles based on their content and relevance, making it easier for appropriate authorities and persons to efficiently navigate the large volume of COVID-19 related literatures to identify important findings and draw meaningful conclusions.

Apart from health crisis, COVID-19 and lockdown have severe consequences, including shut down of school-college-university, office, companies and industries, cancelation of sports and social events, resulting loss of employment and decrease in gross domestic product (GDP). People are forced to remain within the walls with minimal engagement incurring a deep affect on their physical and mental health. Being quarantined inside their household and without physical interaction with outside, people become more addicted towards smart phones, social medias and online blogs/forums in this period [25, 82]. They convey their sentiments and feelings through these digital platforms. Hence, widely used social medias (Facebook and Twitter) have evolved into valuable repositories of information about public sentiment and perception towards lockdown and government policies. Consequently sentiment or opinion mining from these web (social media) corpora has become important to help the governments, policy makers and other concern organizations to figure out useful information and the public opinions for further policy making and opting better precautionary measure to combat the situation.

However, such content is not necessarily reliable always; it frequently contains gossip, disinformation, rumours and fake news. This kind of information is all over the internet in an effort to influence public opinion [128, 129, 130], scam link clicks of money [24], and influence significant events [190]. The rumors are intentionally created by people and spread through various web platforms. It hampers pandemic prevention, control, social governance, etc. Due to the public's interest, all social networking newscasts must adhere to a verifiability standard.

From these observations we have motivated and decided to work on the following areas:

- Automatic classification of COVID-19 related articles.
- Classifying COVID-19 related Rumors and Fake News.
- Analyze public sentiments on different issues during lockdown and viral outbreaks.

In the next section we have identified and highlighted the specific problems which are addressed in this thesis.

1.2 Problem Statement

From the observation mentioned above a set of problems related to COVID-19 and lockdown, are considered to be addressed in this thesis.

The initial problem has been the classification of COVID-19 associated research articles. Article classification is the process of categorizing or labelling articles into predefined classes based on their content, topic and other relevant attributes. It involves using NLP approaches to examine the text of an article and assign it to a specific category. It enables efficient searching, personalized content recommendations, knowledge discovery, decision making support, and content filtering. Overall, article classification improves information organization, enhances user experience, aids in research and analysis, and promotes effective information management. In this pandemic, a massive number of articles that related to COVID-19 have been published. Medical practitioners help themselves stay updated with the latest research and findings related to the virus by going through these articles. As well, it permits researchers to navigate through an enormous amount of scientific literature and identify relevant studies for additional research. They also identify research gaps and emerging trends, enabling them to contribute to the collective knowledge of COVID-19 and potentially make significant advancements in the field. Researchers and medical practitioners still face huge knowledge problems while fighting against this pandemic, as they have to go through all of these research papers manually. Hence, one of the most significant responsibilities is to automate COVID-19 article classification to improve the research work and search procedure. This has been considered as one of the research problem in this thesis and discussed subsequently (Chapter 3).

Classification of Rumours on COVID-19 related issues is taken as the second problem in the thesis. Rumor classification signifies detecting and categorizing rumors or unverified information circulating in various forms, such as social media, news articles, or online platforms. It involves analyzing the content and context of the information to determine its veracity or reliability. Rumor classification aims to distinguish between true information, false information, and information that cannot be validated at the time of assessment, helping to reduce the dissemination of misinformation and providing clarity to the public during undefined situations, such as COVID-19 pandemic. As a result of lockdown during the viral outbreaks people are quarantined within the walls with minimum engagement and physical interaction with outside world. In this period, they are more addicted towards smart phones and digital Medias (specially social media and online blogs/forums etc). But the online contents are not reliable always; it frequently contains gossip, misinformation, rumours and fake news. This kind of information is all over the internet in an effort to influence public opinion [128, 129, 130], scam link clicks of money [24], and influence significant events [190]. Lack of knowledge and proper information may create public anxiety and panic which could be devastating in pandemic. Consequently detecting rumours has become obligatory for the economy and public safety. But analyzing such huge amount of information, specially the informal content of online media is a mammoth task. Automated intelligence techniques are required. It has been considered as the second problem and discussed in Chapter 4.

Next towards the end of the thesis we have taken two sentiment classification problems on twitter dataset. Opinion mining or sentiment analysis is technique in NLP that aims to identify sentiment or opinion directed at specific topic or event or a given target entity.

Recently, a growing interest has been observed among researchers for sentiment analysis in both education and industry. The majority of sentiment analysis studies have primarily concentrated on determining the entire sentiment of a text. However, there is often a desire to understand other's perspectives on certain aspects of a term or subject. For example, Consider a customer review of a smartphone:

Review: "The camera quality of this smartphone is excellent, but the battery life is disappointing."

In ABSA, we would identify the aspects or features mentioned in the review and analyze the sentiments associated with each aspect. In this example, the aspects are “camera quality” and “battery life.”

Aspect: **Camera Quality**

Sentiment: Positive

Aspect: **Battery Life**

Sentiment: Negative

ABSA enables us to grasp the sentiments conveyed regarding various aspects of a product or service, providing a more granular analysis than general sentiment analysis. In response to this concern, the concept of ABSA was introduced in 2010 [186]. Despite the fact that some of the ABSA techniques are available, most of these are experimental prototypes [113]; hence ABSA is still an open area to explore for NLP and data science researchers. An ABSA system processes a collection of texts (such as tweets) that detail specific information (such as, physical shut down of school-college-university and online teaching-learning with remote participant) and estimate their sentiments. Because of lock down school-college-university, office, companies and industries are remained closed, all social and sports events are also cancelled or rescheduled, resulting loss of employment and decrease in gross domestic product (GDP). People are forced detach from outside world to avoid infection. This makes them lonely and physically isolated from others. Consequently they are more habituated towards smart phones, social media. They used to share their daily activities, sympathy and emotions through these online media. As a result, widely used social media platforms like Facebook and Twitter have transformed into massive sources of information about public opinion and perception towards lockdown and government policies. Consequently mining these web (social media) corpora has become important to help the governments, policy makers and researchers to figure out useful information and the public opinions for further policy making and opting better precautionary measure to combat the pandemic. To address these problems we have considered ABSA to identify sentiment of Indians towards the lockdown inventiveness taken by the Government of India to control the transmission of Coronavirus as the third

problem of the thesis. This has been discussed in details in Chapter 5.

Like the other activities the teaching-learning process also brutally affected COVID-19 and lock-down. All the academic institutions are focusing on elevating the educational activities via online during this period. During the Virus outbursts the teaching-learning-evaluation process involving traditional classroom framework has drastically shifted to online mode with remote participants rather than physical classroom learners. But existing setup for such remote learning only restricted to audio-visual contents and ICT based classroom etc. Mainly the Lab based subjects and kids' education face a major setback. Absence of communication with fellow participants, lack of interaction with instructors and passive participation in these modes of learning incur minimum involvement of students in the class. As a consequence, concentrations of the learners often break without any interactions in the online classes. But, although its several drawbacks online education has come out as better than nothing, as the physical class is not feasible during the period. People from all over the world have flocked on social media to convey their opinions and feelings about online teaching-learning. Consequently it has become important to help the governments/ policy makers/ educational institutions figuring out useful information and the opinions of students and guardians about online education during the time. To address this, an ABSA system to predict the opinion of common public underlying various aspects of online teaching-learning in this scenario has been considered as the final problem of this thesis. It has been detailed in Chapter 6.

In the next section we have highlighted the major Contributions of the thesis.

1.3 Contributions of the Dissertation

In the previous section we have identified the problems that are addressed in this thesis. Here we have briefly highlighted the major Contributions of dissertation as follows:

Attention-based Bidirectional LSTM with Embedding technique for classification of COVID-19 articles:

A DL based multi-class classification system named ABiLSTM has been developed to accurately classify and categorize COVID-19 related research articles, which can aid in addressing the challenges posed by the novel Coronavirus. This system incorporates

an embedding layer, BiLSTM, and attention mechanism, utilizing various identified candidate attributes to improve the efficiency of model. It plays a crucial role in guiding research efforts towards the right direction, expediting the development of treatments and diagnostics for COVID-19, and facilitating the implementation of necessary preventive measures. This DL model is applied to solve the first issue mentioned in my thesis and thoroughly elucidated in Chapter 3.

Attention-based Bidirectional LSTM with BERT Embedding for classification of Imbalanced COVID-19 Rumors:

A density-based synthetic minority oversampling technique (DSMOTE) that combines the benefits of optimised DBSCAN and SMOTE in order to balance the COVID-19 rumors dataset, has been suggested in Chapter 4. Then, attention based BiLSTM with BERT model (BERT-ABiLSTM) has been developed for COVID-19 rumor classification. The BERT embedding layer, BiLSTM, and Attention Mechanism are combined together to improve the classifier's performance using different balanced candidate methods. This system efficiently identifies and categorizes rumors into relevant classes, serving the purpose of preventing unnecessary panic, maintaining socio-psychological and economic stability, and assisting the government in effectively managing the COVID-19 crisis. This ensemble deep learning model presented in this thesis addresses the second problem described and elaborated in Chapter 4.

Aspect Based Sentiment Analysis using Multi Criteria Decision Making and Deep Learning under COVID-19 Pandemic in India:

An ABSA system has been suggested to predict the sentiment of the Indian public regarding the lockdown measures implemented by the Government of India to combat the disperse of the virus. In the suggested methodology, tweets are grouped into clusters using k-means algorithm. Aspect terms are then recognised using a Multi Criteria Decision Making technique (BWM), which takes into account various features such as PMI, TF-IDF, Dice Coefficient, and Jaccard Coefficient from each cluster. Subsequently, sentiment regarding each aspect is extracted from individual tweets. Finally, the sentiment polarities are aggregated using the Dempster-Shafer Aggregation Method. This methodology allows for comprehensive sentiment analysis based on

aspect terms extracted from tweets. Next, to classify the Twitter dataset, Bi-GRU model has been developed. This study addresses the third problem outlined in the thesis and thoroughly discussed in Chapter 5.

Sentiment Analysis on Online Education Using Multi-Criteria Decision Making and Deep Learning during COVID-19:

An aspect-based approach is proposed for analyzing public sentiment related to teaching-learning process throughout the COVID-19 transmission using a Twitter dataset. The suggested method uses the k-Mediod algorithm to cluster tweets and employs the Multi Criteria Decision Making approach (AHP) to select aspect terms based on attributes such as PMI, TF-IDF, Dice Coefficient, Jaccard Coefficient and TextRank from each cluster. Public sentiment is then extracted from individual tweets based on identified aspect terms. The results indicate that there is a general disagreement among people regarding most aspects of online education during the pandemic. Furthermore, a TRANS-Att-BiGRU model is introduced for sentiment classification, incorporating an embedding layer, BiGRU, and attention mechanism to enhance the classification performance. This deep learning model presented in this study addresses the end problem outlined in the thesis and thoroughly discussed in Chapter 6.

1.4 Organization of the Thesis

The thesis is structured into seven chapters. The summary of each chapter is as follows:

Chapter 2 presents related literature in the field of text classification and sentiment analysis, different techniques, and their applications.

Chapter 3 presents an attention-based bidirectional LSTM with embedding technique for classification of COVID-19 articles.

Chapter 4 describes a method for classifying imbalanced COVID-19 rumors. It consists of a deep learning model named attention-based bidirectional LSTM with BERT embedding to identify the rumors.

Chapter 5 includes a DL method to design an ABSA system to assess the sentiments of Indians regarding the lockdown measures implemented by the Government of India

to curb the spread of Coronavirus.

Chapter 6 presents a DL method for analysing the sentiment of student/teacher/parent about the online learning during COVID-19 pandemic situation.

Chapter 7 provides a conclusion to the thesis by emphasising the primary contributions and significance of the research. Future scopes of the present works are mentioned towards the end of this chapter.

1.5 Summary

The COVID-19 has prompted a surge in AI technology, specifically in the areas of article classification and analysis of sentiment. This thesis focuses on the origin, motivation, and objectives of the work, highlighting the problem statement and acknowledging the various challenges involved. The contributions of the thesis encompass multiple methods for automatic article classification and sentiment analysis, with the aim of leveraging technology to benefit society. Key research contributions include the classification of COVID-19-related articles and rumors, aspect-based sentiment analysis of Indians during lockdown and unlock phases, and analyzing public sentiment on online learning, benefiting medical practitioners, policymakers, and society as a whole. The chapter concludes by outlining the organization of the entire thesis. The research findings from this study will facilitate the creation of automated classification and sentiment analysis modules in the near future.