ECTI Transactions on Computer and Information Technology
Journal homepage: https://ph01.tci-thaijo.org/index.php/ecticit/
Published by the ECTI Association, Thailand, ISSN: 2286-9131

Sentiment Analysis on Large-Scale Covid-19 Tweets using Hybrid Convolutional LSTM Based on Naïve Bayes Sentiment Modeling

Arunava Kumar Chakraborty[1], Dipankar Das[2] and Anup Kumar Kolya[3]

ABSTRACT

Millions of lives were affected rapidly throughout the world when the Covid-19 outbreak spread by leaps and bounds. During this catastrophic period, people used to express their condolence as well as emotions through different social networks. In order to analyze the public comments on Twitter, an experimental approach is developed based on popular words regarding this pandemic. In this paper, various NLP-based research works are discussed on sentiment analysis, trend prediction, topic modeling, learning mechanisms, etc. Furthermore, the hybrid deep learning models are developed based on the Naïve Bayes sentiment model to predict the sentiment from the collected huge number of Coronavirus-related tweets. After performing the n-gram analysis, the Covid-19 specific words are extracted based on their popularity. The public sentiment trend has been analyzed using the extracted topics related to Covid-19 and the tweets are classified according to their sentiment scores. The distinguished sentiment ratings are assigned to the collected tweets based on their sentiment class. Then Convo-Sequential and Convo-Bidirectional long-short term networks are trained using fine-grained sentiment-rated tweets to categorize Covid-19 tweets into five different sentiment classes. Finally, our proposed Convo-Sequential and Convo-Bidirectional LSTM models achieved 84.52% and 85.03% of validation accuracy respectively for the first phase dataset whereas using the second phase dataset the models obtained the validation accuracy of 86.58% and 87.22% respectively.

Article information:

Keywords: Covid-19, Naïve Bayes Classifier, Hybrid Convolutional LSTM, Fine-grained Sentiment Analysis

Article history:
Received: May 6, 2023
Revised: July 1, 2023
Accepted: August 3, 2023
Published: August 12, 2023 (Online)

DOI: 10.37936/ecti-cit.2023173.252549

1. INTRODUCTION

The Covid-19 pandemic was initially detected in Wuhan, Hubei Province, China on 31st December 2019, and was continually expanding around the world. On March 11, 2020, the World Health Organization (WHO) officially declared the Covid-19 outbreak as a pandemic, after witnessing the trend of its spread [1]. This virus spread from China to Russia, the United States, the United Kingdom, Brazil, Spain, Italy, and many other countries by infecting and killing thousands of people. The pandemic has claimed the lives of millions of people, prompting many countries to take a hard lockdown to prevent the spread of this virus. According to doctors, even after the discovery of the covid-19 vaccine, it is important to maintain social distancing otherwise, the spread of the virus will never be stopped [2]. Various social media platforms have been instrumental in regularly disseminating information about the epidemic around the world. Many people have shared their views about the disease on Twitter during the lockdown, so we've been inspired to analyze a huge amount of Twitter data to measure people's awareness about the epidemic.

Many people in different countries have used their first language to post their opinions on social media so we initially faced many challenges during the process of streaming English tweets around the world

[1,3] The authors are with Department of Computer Science & Engineering, RCC Institute of Information Technology, Beleghata, Kolkata - 700015, India, E-mail: arunava95.akc@gmail.com and anupkumar.kolya@rcciit.org.in
[2] The author is with Department of Computer Science & Engineering, Jadavpur University, Kolkata 700032, India, E-mail: dipankar.dipnil2005@gmail.com

[3]. However, exclusive English tweets on Covid-19 have been collected in two phases over seven months for August - October 2020 and April - June 2021 respectively. The first and second-phase datasets have been developed with almost 320k and 489k Covid-19 tweets. Then we used the n-gram model to analyze the trend of tweets so that the most relevant grams can be extracted for further computation. The most popular Covid-19 related words have been identified from the word corpus using the Bag-of-Words model. The sentiment polarities of the pre-processed tweets have been evaluated to classify them into positive, negative, and neutral classes. For the fine-grained classification of positive and negative tweets, we used the Naïve Bayes classifier for assigning the refined sentiment scores to the classified tweets based on the probability of being in the binary classes. Instead of using conventional Machine Learning algorithms, Hybrid Deep Learning models have been introduced since the deep hybrid neural networks are beneficial in text generation, word representation estimation, vector representation, feature presentation, sentence modeling, and sentence classification. Finally, both Hybrid Convolutional LSTM networks have been trained using fine-grained tweets to classify the tweets into five different sentiment classes.

1.1 Motivation & Objective

The inspiration behind this work is to study the recent workflow on Covid-19 from a dedicated natural language research perspective and to analyze the behavioral aspects and nature of people around the world about this epidemic. Twitter is one of the most famous microblogging social media platforms which can be considered as a source of valuable information. In 2021, out of 330 million active Twitter users from around the world, India has 17.5 million active Twitter users, making India currently the third-highest user country in the world[1]. Twitter users tweet their expressions in 280 characters. As many people have shared their feelings about the epidemic on Twitter during the lockdown episode due to the epidemic situation, we have been inspired to analyze the public's emotional state from time to time.

This pandemic not only claimed the lives of millions but also plunged the entire world economy into recession. During this difficult time, many people lost their jobs and because of this, the unemployment rate is increasing day by day all over the world. The various NLP-based research works have been discussed along with their respective proposals, methodologies, and contributions from recent years. Since both datasets have been developed by collecting raw tweets without any sentiment label, we evaluate the polarity ratings of the tweets using the NLTK-based sentiment analyzer and classify them into positive, negative, and neutral classes. The polar tweets are analyzed using the Naïve Bayes sentiment model for fine-grained classification. Furthermore, the neutral tweets and the fine-grained tweets are combined for final analysis. The major objective of this research work is to analyze the public sentiment for finding the global sentiment trend about this pandemic and finally train both Convo-Sequential and Convo-Bidirectional LSTM models based on the fine-grained tweets to predict the sentiment class of the unseen tweets by analyzing correlations between tweet words.

[1] www.statista.com

1.2 Research Contribution

The novel contributions of this paper are:

- Lexical N-gram Model for extraction of most popular unigrams, bigrams, and trigrams to find the trend of tweet-specific words.
- Covid-19 Specific Word Identification based on the Bag-of-Words Model to find the impact of those words over the tweets.
- Sentiment Analysis of the unlabeled tweets to evaluate the polarity ratings of each of the tweets.
- Naïve Bayes sentiment model for evaluation of refined sentiment ratings for positive and negative tweets based on the previously extracted features.
- Fine-grained Classification of tweets based on their sentiment polarity ratings using the Naïve Bayes Model for finding the global sentiment trend.
- Sentiment classification using Hybrid Convolutional LSTM based on the fine-grained tweets and their refined sentiment ratings for sentiment prediction.

The organization of the paper is as follows: Section 2 represents the literature review on previously published research works. The system architecture of our proposed methodology for this research work has been discussed in Section 3. In Section 4, the structure of the Covid-19 dataset and proposed preprocessing technique have been demonstrated. The trend of tweet words using the n-gram model has been identified as Feature I in Section 5. In Section 6, the Covid-19 specific words have been extracted as Feature II from the Bag-of-Words model based on the word lexicon. Section 7 consists of the evolutionary classification of the sentiment-rated tweets considering their sentiment polarities. We demonstrated the Naïve Bayes Model to evaluate the refined sentiment ratings of the positive and negative tweets in Section 8. The global sentiment trend on Covid-19 throughout the world has been presented in Section 9. In Section 10 we trained the Hybrid Convolutional LSTM models based on the fine-grained tweets along with their sentiment ratings. Section 11 represents the comparative analysis of the performance of the proposed mode with other similar research works. Finally, Section 12 concludes with the future prospects of the research work.

2. LITERATURE REVIEW

Recently, a wide variety of work has been found in the research domain of Natural Language Processing. In this limited literature review, we have presented the proposed work, methodology, and analysis for some of those state-of-the-art research works. The entire literature review is mainly divided by considering three different aspects e.g., sentiment analysis from social media, the healthcare industry, and the Covid-19 pandemic.

2.1 Sentiment Analysis from Web and Social Media

Different types of text are available in traditional media, and among them, microblogging messages are noisy, brief, and socially embedded. In order to manage networked texts in microblogging, several researchers developed a unique sociological technique (SANT) [4]. They used social theories to extract sentiment relations between tweets and created an optimization approach for SANT to obtain consistent performance for various volumes of training data. The probability model was used to build a semi-automated sentiment analysis based on the online social network [5]. The authors developed a model that scans a train set of text messages to prepare a sentiment lexicon that includes the list of words that appeared in the messages along with the probability that a text message having those words was carrying a positive sense. The positivity score of text messages of the test set was then calculated and each message was classified as positive or negative based on the threshold value established using a train set. Few researchers implemented a model that can use Twitter data to demonstrate a potentially integrated methodology based on sentiment analysis and social network analysis [7]. They collected three types of data from #SamSmith (Grammy Awards in 2015) and #Ukraine (crisis of 2014) Twitter channels. They used a classifier based on the Multinomial Naïve Bayes algorithm to identify the tweets from different sentiment classes.

In order to perform Topic-based and Message-level sentiment analysis, some researchers described a deep LSTM architecture [10]. On top of the word embeddings pre-trained on a large collection of Twitter conversations, the authors deployed LSTM networks enhanced with two types of attention mechanisms. According to some researchers, the bidirectional LSTM with two-dimensional max-pooling can improve text categorization [8]. They proposed two combinational models: BLSTM-2D Pooling and BLSTM-2DCNN and used a 2D filter size and 2D max-pooling size as (5,5) for training purposes. Finally, they achieved the 52.6% of the highest prediction accuracy. A neural network-based Emotion Recognition framework was developed by a group of researchers using the LSTM hyperparameter optimization mechanism [12]. They discovered that adjusting LSTM hyperparameters improved the recognition rate of four-quadrant dimensional emotions by 14 percent. Finally, the model based on the optimized LSTM classifier obtained 77.68% accuracy by applying the Differential Evolution algorithm. Another research work represents the use of a multimodal feature learning approach using neural network-based Skip-gram and Denoising Autoencoders [9]. The models analyzed the sentiment of micro-blogging content like tweets containing short text and images. The proposed semi-supervised model CBOW-LR used the sentiment polarity classifier to learn the vector representation and achieved better accuracy over CBOW representation on a similar volume of tweets. In comparison to previous state-of-the-art approaches, another hybrid model (CBOW-DA-LR) works in an unsupervised and semi-supervised manner and achieved higher classification accuracy.

A novel learning strategy was introduced to overcome language variation, considering the tendency of socially proximate individuals in 2017 [11]. The proposed approach focused on different local regions of the social network to capture subtle shifts in meaning across the network. The model was formulated based on the social attention mechanism to predict the weighted combination of the outputs of the basic models considering the unique weighting of authors in the social network. The model outperforms the standard CNN model in analyzing the sentiment ratings of Twitter and review data. Although some researchers discovered that combining deep CNN with word embedding is preferable instead of using the TF-IDF vectorizer [15, 16]. A generative model HUB was built to incorporate two companion learning tasks of opinionated content modeling and social network structure modeling to examine the problem of user behavior modeling using many sources of user-generated data [13]. On two large collections of review datasets from Amazon and Yelp, the predictive user behavior models outperformed the comparable baseline algorithms in sentiment classification and item/friend recommendations. Topic modeling was used to automatically extract the many topics of climate change conversation, and sentiment analysis was utilized to categorize tweets as positive, neutral, or negative [14]. The pooled Latent Dirichlet Allocation (LDA) model was trained by looking at word co-occurrences in collected tweets. Finally, based on the average UMass coherence score, the pooled LDA model outperforms both the unpooled LDA and Biterm Topic Model.

2.2 Sentiment Analysis in Healthcare

The public's anxiety over a global pandemic can be measured using Sentiment Analysis for textual categorization. Some researchers developed the Epidemic Sentiment Monitoring System (ESMOS), which can be used to track disease progression and peaks [17]. Multinomial Naïve Bayes produced better prediction accuracy in categorizing negative versus neutral tweets. Using natural language processing, a group of academics from Johns Hopkins University investigated mental health occurrences from widely available Twitter data [18]. They examined the characteristics of post-traumatic stress disorder (PTSD), depression, bipolar disorder, and Seasonal Affective Disorder (SAD) and provide some insights regarding the classification of quantitative linguistic data. In order to analyze the sentiment polarities (positive, negative, or neutral) of textual data, there are some commercial and non-commercial applications. Semantria and TheySay are two prominent commercial tools, whereas WEKA and Google Prediction API are two popular non-commercial tools. Another study highlights the difficulty of analyzing sentiment using several tools in the healthcare domain [19]. The authors presented a performance comparison of the various tools across huge datasets and discovered that WEKA performed best on their data. Another study demonstrates a sentiment-tracking algorithm for detecting symptoms of depression from tweets [20]. The sentiment classifier categorized the depressive tweets within three sub-classes i.e., depressed mood, disturbed sleep, and loss of energy. According to the researchers, the machine learning algorithms enhanced the precision of simple keywords for accurately recognizing depressed symptoms. A group of researchers proposed the iDoctor healthcare recommendation system based on hybrid matrix factorization methods [21]. They employed the LDA model for topic modeling to capture the user's latent priority, and doctor features may be recovered from user review comments on doctors. Finally, both features were used in matrix factorization to provide more accurate counseling. A group of researchers developed a system for analyzing the public's thoughts regarding health technology and identifying their requirements by using social media to uncover trends in health information [23]. SentiWordNet was used to classify the sentiment of the collected tweets based on the most frequently used words in the corpus. Another comparative study based on sentiment analysis in healthcare represents that SVM with Naïve Bayes classifier performs better than different supervised algorithms [24]. Some researchers proposed a psychometric analyzer to analyze the sentiment and emotion of patient health based on previous medical history and recorded speech [22]. The researchers used the Support Vector Machine (SVM) and Support Vector Regression (SVR) for sentiment analysis, as well as the Decision Tree to detect emotions.

2.3 Sentiment Analysis on Covid-19

Machine learning algorithms also performed better to predict the sentiment ratings from Covid-19 tweets. Some academics have used adaptive K-means clustering on Covid-19 tweets [26]. The n-gram model was used to examine the tweet patterns. As a result, they noticed a variation in the frequency of n-grams in the dataset. According to the comparative performance analysis of four different ML-based classifiers, it has been found that the Naïve Bayes classifier outperforms the Logistic Regression classifier and achieved 91% classification accuracy for short Tweets [29].

In order to estimate the number of Covid-positive cases in different states of India, a Covid-19 trend model was proposed by some researchers [25]. The authors primarily focused on LSTM-based prediction models and used historical data to test several LSTM variations such as stacked, convolutional, and bi-directional LSTM. Finally, they discovered that for short-term prediction, the Bi-LSTM outperforms other LSTM models. An interactive web application was developed for real-time tweet tracking on COVID-19 [28]. The bigrams were extracted from the tweets and the LDA has been used to extract the popular topics on Covid-19. Finally, the Sentiment Intensity Analyzer was used from the NLTK library for finding the sentiment polarities of each tweet. A deep LSTM network was developed to analyze the sentiment from Coronavirus-related tweets based on the specific topics extracted by the LDA model [30]. The model classified the tweets into positive, negative, and neutral classes and achieved 81.15% of classification accuracy. In another study, it has been found that the CNN-based LSTM model with *GloVe* embeddings can perform better sentiment prediction on large-scale Covid-19 tweets [27]. Some academics proposed an LSTM network on John Hopkins University data to analyze the Covid-19 trend [31]. The model achieved 93.40% for short-term and 93.40% for long-term tweets. A few researchers suggested a hybrid artificial intelligence (AI) methodology for predicting the global Covid-19 trend [32]. RoBERTa is a type of BERT model for textual feature selection. To analyze the Coronavirus trend, they created an improved susceptible–infected (ISI) model embedded with the LSTM network to determine the distinct infection rates.

3. PROPOSED METHODOLOGY

Large-scale tweet collection is the foundation of this research as well as the first phase of our proposed method. Initially, we collected 320k tweets related to Covid-19 during the phase of August - October 2020. The second dataset consists of 489k Covid-19 tweets which have been collected during the phase

April - June 2021. Then using the NLTK[2] (Natural Language Toolkit) package, the collected tweets are cleaned and pre-processed from both corpora by eliminating noisy data from the tweets. Then the tweets are tokenized to extract the smaller tokens from the tweets and lemmatization is used to convert the tokens into their original word lemma. After cleaning and pre-processing the most frequent unigrams, bigrams, and trigrams are extracted from the dataset. Then we identify the Covid-19 exclusive words and their frequency of recurrence in the corpus to understand the impact of those words on the tweets.

On the other hand, the NLTK-based sentiment analyzer has been used to evaluate the sentiment polarity of each cleaned and pre-processed tweet. Then using evolutionary classification, the tweets are classified into *positive, negative,* and *neutral* classes based on their *compound* sentiment ratings. Then Naïve Bayes Sentiment model is developed based on the popular grams to perform fine-grained sentiment classification on positive & negative tweets and assign the Refined Sentiment rating to fine-grained tweets. The model calculates the probability scores of each tweet for being in respective classes. We further classify the tweets into four sentiment classes i.e., Most Positive (1.0), Positive (0.5), Negative (-0.5), and Most Negative (-1.0) based on the probability scores. Then we merged the previously categorized neutral tweets with fine-grained data by assigning the class as Neutral (0.0). Finally, we calculate the daily Global Sentiment Trend for Covid-19 using refined sentiment ratings.

Considering the fine-grained tweets from the corpora we label them into five integer values (from 0 to 4) and transform them into word vectors for further processing. The datasets are divided into 80:20 ratios for training and validation of the hybrid models. Finally, both Convo-Sequential and Convo-Bidirectional deep LSTM networks are trained using the vectorized tweets and their corresponding labels to categorize the tweets into sentiment classes. After completion of training the proposed models achieve certain accuracy considering predicted and evaluated data from both datasets. We also train these models on different widely available public corpora to identify the noticeable performance efficiency of our system by performing a comparative study with other previous benchmark experiments. In Figure 1, we present the Proposed System Architecture of this research work.

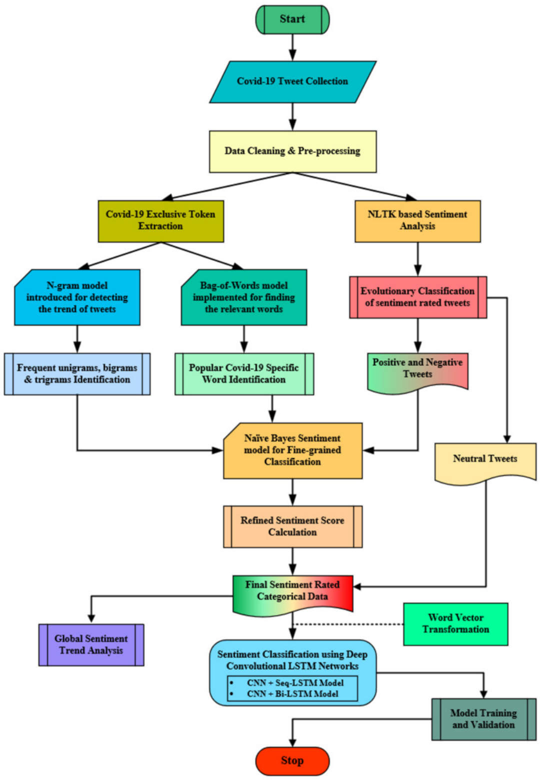

Fig.1: *Proposed System Architecture of Covid-19 Sentiment Analysis using Hybrid Convolutional LSTM Based on Naïve Bayes Sentiment Modeling.*

4. PREPARING COVID-19 DATASET

During the past two years, almost every country has been affected by the coronavirus throughout the world. On January 30, 2020, WHO declared the outbreak a Public Health Emergency of International Concern, and on March 11, 2020, it was declared a pandemic. According to government reports, the United States, Italy, Spain, and France had the most confirmed cases. Even in countries where the new coronavirus hasn't spread as widely, they're still under a lot of pressure. To stop the infection rate at a local level, many governments have announced restrictive measures such as lockdown, home isolation, and stay-at-home orders. However, some people were surprised to find that if their government had taken appropriate measures to curb this pandemic at an early stage, they might not be facing the current situation today. It's been two years since the Covid-19 epidemic broke out, claiming millions of lives and changing the way we communicate and explore the world with each of us.

4.1 Collection of Tweets

The live streaming of tweets from Twitter was started after WHO declared Covid-19 as a pandemic. Since this pandemic has affected the entire planet, we have collected Covid-19-related English tweets at a frequency of 10k per day in two different periods, beginning in August - October 2020 and ending in April - June 2021. We prepared the first phase dataset of about 320k tweets collected from 20th August to 20th October 2020 when the pandemic was spreading with its fatal intensity. Since the beginning of 2021, with the approval of WHO, various companies

[2]https://www.nltk.org/

like Pfizer, and BioNTech have increased production of the covid vaccine in response to worldwide growing demand. After six months we again start collecting tweets from Twitter as at that time first dose vaccination was started in different countries. Almost 489k tweets were collected in the period from 26th April to 27th June 2021 for the second phase dataset.

Unsurprisingly, these tweets were significantly noisy with grammatical errors, misspelled words, extra white spaces, numbers, URLs, misspelled punctuations marks, stop words, misused emojis and emoticons, similar tweets, tweets without any information, and so on. The datasets contain important information about most of the tweets as attributes e.g., -

- id [Number]
- created_at [DateTime]
- source [Text]
- original_text [Text]
- favorite_count [Number]
- retweet_count [Number]
- original_author [Text]
- hashtags [Text]
- user_mentions [Text]
- place [Text]

Twitter users can typically live stream only 1-2 percent of total tweets on a given keyword using the Twitter API. However, we were able to collect tweets at a rate of about 10k each day. Twitter used to allow a maximum of 140 characters per tweet, including emoticons, until October 2017. However, on 7th November 2017, Twitter expanded the character limit per tweet to 280 characters. In this research, we considered the raw tweets (original_text) and the creation date (created_at) attributes for further analysis.

Finally, we collected 3,20,316 tweets for the first phase dataset and 4,89,269 tweets for the second phase dataset using 10 hash-tagged keywords i.e., ***#covid-19, #coronavirus, #pandemic, #covid, #lockdown, #homequarantine, #quarantinecenter, #socialdistancing, #stayhome***, and ***#staysafe***. We selected these keywords since these were the most used terms by different news channels, verified social media handles, and government authorities during that period.

4.2 Data Preprocessing

In NLP, text preprocessing is an essential step for cleaning and preparing text data for building a Machine Learning model. Data pre-processing is primarily used to clean raw data by performing a set of operations to produce a better outcome for subsequent analyses. The Twitter RESTful API tweepy[3] was used to access data about both Twitter users and what they are tweeting about this pandemic. As the raw tweets were collected for developing the datasets, they contain lots of duplicate tweets having the same status ID. However, all the duplicate tweets have been removed from the datasets and 1,20,509 and 1,47,475 tweets have been found from the first phase and second phase Covid-19 datasets. After removing all the duplicate tweets, the final size of the datasets was 35MB and 43MB respectively.

The tweets were pre-processed by creating a user-defined pre-processing function using NLTK, a Python toolkit for Natural Language Processing. It converts all tweets to lowercase in the first phase. The tweets are then stripped of all excess white spaces, numerals, special characters, ASCII characters, URLs, punctuation, and stop words. Then the hash-tagged keywords are removed from the tweets since the higher frequency of the hash-tagged keywords may affect the feature extraction process. Tokenization is the method for splitting a large chunk of text into smaller parts called tokens. The tweets have been tokenized using the function so that the sentence can be divided into smaller parts of the word. Lemmatization is a process that reduces words to their base or dictionary form, known as a lemma. Using lemmatization, the pre-processing function has normalized the words by reducing them to a common base. In Figure 2, the daily distribution of tweets has been presented from both datasets.

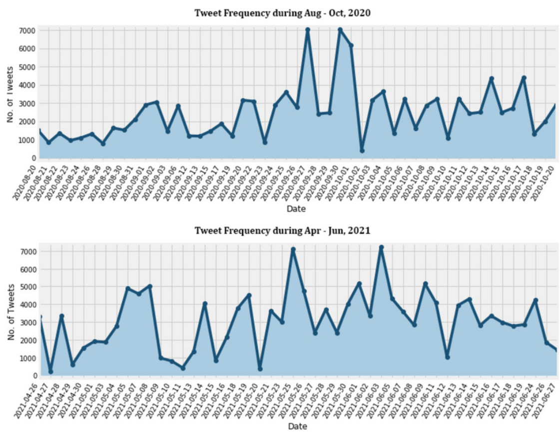

Fig.2: *Tweets distribution of the Covid-19 tweets corpora.*

5. FEATURE I: WORD TREND IDENTIFICATION USING N-GRAM

In Natural Language Processing, lexical n-gram models are commonly used for statistical analysis and syntax feature mapping. We found the popularity of words or a collection of adjacent words by developing an n-gram model to analyze the generated corpus of tokenized words. Here the probability chain rule has been used to determine the likelihood of the occurrence of a series:

$$\left.\begin{aligned} P(x_1, x_2, x_3, \ldots x_n) &= P(x_1)P(x_2|x_1)P(x_3|x_1, x_2) \\ &\ldots P(x_n|x_1, x_2, x_3, \ldots x_{(n-1)}) \\ &= \prod_{i=1}^{n} P(x_i|x_1^{(i-1)}) \end{aligned}\right\} \quad (1)$$

[3]https://www.tweepy.org/

For example, a sentence can be considered as "Now Covid-19 wave is increasing". Now according to the probability chain rule, *P("Now Covid19 wave is increasing") = P("Now") x P("Covid19" | "Now") x P("wave" | "Now Covid19") x P("is" | "Now Covid19 wave") x P("increasing" | "Now Covid19 wave is")*.

The probabilities of words in each sentence after applying the probability chain rule:

$$\left.\begin{aligned} P(W_1^n) &= \prod\nolimits_j P(W_j|W_1, W_2, W_3, \ldots W_{(j-1)}) \\ &= \prod\nolimits_{j=1}^{n} P(W_j|W_1^{(j-1)}) \end{aligned}\right\} \quad (2)$$

A Markov assumption states that the probability of a word is entirely dependent on the probability of the previous word. According to this probabilistic model, the probability of some future unit can be predicted without looking too far into the past. The bigram model estimates the probability of a word by using only the conditional probability $P(W_i|W_{i-1})$ of one preceding word on the given condition of all the previous words $P(W_i|W_1^{i-1})$ [6].

$$P(W_1, W_2) = \prod\nolimits_{i=2} P(W_2|W_1) \quad (3)$$

The expression for the bigram probability is –

$$P(W_k|W_{(k-1)}) = \frac{count\ (W_{(k-1)}, W_k)}{count(W_{(k-1)})} \quad (4)$$

The most popular unigrams, bigrams, and trigrams have been extracted from the corpora using the n-gram model. The n-gram analysis provides a summary of the most popular topics to determine the reason behind a sentiment for a better understanding of human emotions on Covid-related tweets. From this analysis, it is evident that there is a significant difference between the popularity of grams identified during the two different time phases as some new grams have been detected in the period Apr – Jun 2021.

According to this n-gram model, the popularity of trigrams is lower than that of bigrams, and the popularity of unigrams is the highest. However, Table 1 represents some of the most frequent unigrams, bigrams, and trigrams from both datasets along with their popularity count.

***Table 1:** Most Popular N-grams.*

	Covid-19 Dataset I (Aug - Oct 2020)	**Covid-19 Dataset II (Apr - Jun 2021)**
n = 1	('case', 10654), ('new', 9796), ('trump', 7603), ('people', 6957), ('death', 5836)	('vaccine', 19776), ('new', 12009), ('case', 11607), ('people', 8414), ('death', 6365)
n = 2	(('new', 'case'), 2690), (('tested', 'positive'), 1997), (('test', 'positive'), 1147), (('white', 'house'), 993), (('president', 'trump'), 801)	(('new', 'case'), 3478), (('fully', 'vaccinated'), 1013), (('case', 'death'), 912), (('public', 'health'), 848), (('vaccine', 'appointment'), 828)
n=3	(('report', 'new', 'case'), 366), (('new', 'case', 'new'), 184), (('new', 'case', 'death'), 170), (('case', 'new', 'death'), 169), (('president', 'donald', 'trump'), 158)	(('appointment', 'available', 'walgreens'), 658), (('vaccine', 'appointment', 'available'), 628), (('new', 'available', 'appointment'), 578), (('available', 'appointment', 'detected'), 578), (('appointment', 'detected', 'provider'), 578)

6. FEATURE II: COVID-19 SPECIFIC WORD DETECTION

After the completion of the pre-processing phase, the Bag-of-Words (BOW) model has been proposed based on frequently occurring words from the Covid lexicon, and we have extracted the most common Covid-19 specific words. The word cloud is mainly used for the novel visual representation of text data.

Several words have been identified in the prepared word lexicon at various times and in different positions within the tweets. Identifying the most popular words from both corpora helps to analyze the impact of those words on the tweet. The lists of almost 1.06 and 1.32 million words have been identified from both generated corpora. For the most popular Covid-19 exclusive words, the frequency of each token has been evaluated from the obtained lists. The popularity score of each word has been calculated from both datasets.

Following the identification of word frequency, the likelihood of occurrence for each word has been calculated based on a total of 10,66,548 and 13,22,513 words from both generated corpora. Table 2 represents the probability scores of some most popular words from both first and second phase datasets.

$$P(W_i) = \frac{count(W_i)}{\sum_{i=0}^{n} count(W_{i=0}^{n})} \quad (5)$$

***Table 2:** Popularity & Probability of Most Frequent Words.*

	Covid-19 Dataset I (Aug - Oct 2020)		**Covid-19 Dataset II (Apr - Jun 2021)**	
	Words	**Probability**	**Words**	**Probability**
1	case	0.009989	vaccine	0.014953
2	new	0.009185	new	0.009080
3	trump	0.007129	Case	0.008776
4	people	0.006523	People	0.006362
5	death	0.005472	death	0.004813

During the analysis, a total of 39,834 and 55,816 unique words have been extracted from both Covid-19 corpora. It is also evident that in 2021 the "vaccine" keyword achieved much higher popularity than other important keywords. It signifies the global impact on the availability, production, usage, and other aspects of the Covid-19 vaccine.

7. NLTK-BASED SENTIMENT ANALYSIS & CLASSIFICATION

Nowadays deep learning-based sentiment analysis become more popular for extracting information and text prediction to analyze public behaviors about any event. Sentiment analysis is mainly used to measure public opinion trends. It is a specific type of data mining through natural language processing (NLP), computational linguistics, and text analysis. The subjective information can be analyzed and extracted from social media to categorize the text into multiple sentiment classes.

NLTK sentiment analyzer has become a popular choice in gold standard research works for evaluating sentiment polarity, as it provides reliable and accurate results for analyzing the sentiment of text data. The Sentiment Analyzer function of the NLTK library is used to evaluate the sentiment polarities of the preprocessed tweets using a set of predefined rules and heuristics for further analysis. Since both datasets have been developed by collecting raw tweets without having any sentiment label, sentiment analysis has been required to understand the overall distribution of tweets from different sentiment classes. The sentiment polarity of each tweet has been evaluated using the NLTK-based Sentiment Analyzer to get the polarity scores for positive, negative, and neutral classes and evaluate the compound sentiment scores of each tweet.

The compound scores are computed by summing the valence scores of each word in the lexicon, adjusted according to the rules. Then the valence scores normalized to be between -1 (most extreme negative) and +1 (most extreme positive) for producing a single unidimensional sentiment measure considering a given sentence. The more the compound score tends to +1, the more the sentence will be positively sensitive and the more it moves towards -1, the more the sentence will be negatively sensitive. The tweets have been classified using the sentiment classification algorithm based on the compound polarity scores into three different classes i.e., *positive* (polarity > 0), *negative* (polarity < 0), and *neutral* (polarity = 0). The positive and negative tweets will be further used for fine-grained sentiment analysis.

8. NAÏVE BAYES SENTIMENT MODEL DEVELOPMENT

The Naïve Bayes algorithm is a standard and effective classification algorithm in machine learning. To perform this classification, the Bayes theorem has been applied with a strong independence assumption between the features. According to several NLP-based research works the Naïve Bayes classifier yields better outcomes when it is used for textual data analysis. The probabilistic classifier considers the high dimensional features by assuming that the occurrence of certain words is independent of the occurrence of other words. The naïve Bayes model will prepare the probability table based on the respective feature set. Finally, the Bayes theorem will be used to calculate the posterior probability.

The Naïve Bayes sentiment model has been proposed based on a Multinomial Naïve Bayes classifier for refining the polarity ratings for all *positive* and *negative* tweets. A naïve Bayes classifier is a generative classifier that performs better than other baseline algorithms on textual fragments for classifying them based on their polarity ratings [3]. In the phase of feature extraction, we extracted the frequent unigrams, bigrams, and trigrams from the N-gram model as well as the Covid-19 specific words from the Bag-of-Words model and our proposed model takes those extracted features as its inputs for further computation. The model computed the probability of each Positive and Negative tweet occurring in the respective classes. The computer probability ratings are between 0.0 to 0.5 for negative tweets and 0.5 to 1.0 for positive tweets. The binary tweets are categorized between the Most Positive (prob >= 0.75), Positive (prob < 0.75), Negative (prob > 0.25), and Most Negative (prob <= 0.25) classes. In this phase, the previously classified *neutral* tweets are merged with the Naïve Bayes sentiment-rated data. By using the classification algorithm presented in Table 3, the *positive, negative*, and *neutral* tweets are categorized into Most Positive (1.0); Positive (0.5), Neutral (0.0), Negative (-0.5), and Most Negative (-1.0) sentiment classes along with the refined sentiment ratings based on their probability score assigned by Naïve Bayes model.

Table 3: *The Fine-grained Sentiment Classification Algorithm.*

```
Algorithm Fine-Grained Sentiment Classification (class, nb_proba, refined_sentiment):
for each i in range (0, len(tweet.index)):
    if tweet_i[class] == 'positive' and tweet_i[nb_proba] >= 0.75:
        tweet_i[refined_sentiment] ← 1.0    # assigned 1.0 for Most Positive Tweets
    elif tweet_i[class] == 'positive' and tweet_i[nb_proba] < 0.75:
        tweet_i[refined_sentiment] ← 0.5    # assigned 0.5 for Positive Tweets
    elif tweet_i[class] == 'neutral':
        tweet_i[refined_sentiment] ← 0.0    # assigned 0.0 for Neutral Tweets
    elif tweet_i[class] == 'negative' and tweet_i[nb_proba] > 0.25:
        tweet_i[refined_sentiment] ← -0.5   # assigned -0.5 for Negative Tweets
    else:
        tweet_i[refined_sentiment] ← -1.0   # assigned -1.0 for Most Negative Tweets
end
```

9. GLOBAL SENTIMENT TREND ANALYSIS

9.1 Overall Average Sentiment Trend

During this analysis, the fine-grained sentiment-rated tweets along with the refined sentiment ratings have been used to identify the average sentiment trend. Some tweets from each sentiment class along with their Naïve Bayes probability scores have been presented in Table 4.

Table 4: *Covid-19 Tweets from Different Sentiment Classes.*

Date	Original Tweet	Naïve Bayes Probability Score	Refined Sentiment Rating	Sentiment Class
27-09-2020	***Neo Irene Akridge*** **@WriterOnRetreat** GOP the party of lies, crimes, hypocrisy, caging children, killing 200,000 Americans with COVID-19, forcing hysterectomies on immigrant women, tearing families apart, allowing bounties on our troops... Screwing over Merrick Garland.	0.10	-1	Most Negative
29-09-2020	**Aziz Shroff @azizshroff1** What is the need for a new Parliament building when funds are required for healthcare to tackle the COVID-19 pandemic? Absolute waste of tax payers money just to satisfy one man's ego.	0.31	-0.5	Negative
14-05-2021	**zindadilkashmir @jindadilkashmir** Health is not a condition of matter, but of Mind. An awareness campaign on #Covid19 was organised by #IndianArmy for the locals of #Bunedevsar #Kashmir #KashmirFightsCovid #COVIDSecondWave #coronavirus #COVID19Vaccine #MaskUpIndia	-	0.0	Neutral
04-05-2021	**Merrilee Fullerton, MPP @DrFullertonMPP** We have enhanced testing requirements during the second wave, recognizing the importance of identifying a case of the virus before it can spread from the community into a long-term care home.	0.74	0.5	Positive
30-04-2021	**DRM Salem @SalemDRM** Due to COVID-19 issues all passengers are being permitted through main entry only to enable monitoring and ticketing activities. However the matter has been forwarded to the concerned higher authority to take further action to prevent spreading of COVID-19 pandemic.	0.99	1.0	Most Positive

The date-wise average sentiment ratings have been calculated from the refined sentiment ratings to find the average sentiment trend globally. It has been found that there is a moderate contrast between the data from different time phases. During the analysis period, a certain drop in the average sentiment is noticeable due to the massive impact of this pandemic. Various emotions can be identified from the tweets presented in Table 4 which may cause changes in the worldwide sentiment trend.

9.2 Sentiment Trend Shift Detection

At the time of executing the sentiment trend analysis, it has been found that daily sentiment polarities changed continuously for positive and negative categories. So, the date-wise average positive and negative polarities have been calculated from both datasets to detect the polarity shift during the time phases Aug – Oct 2020 and Apr – Jun 2021. The Identification of Sentiment Trend shift helps to understand the changes in the daily positive (0.0 to 1.0) and negative (-1.0 to 0.0) polarity trend. The average sentiment trend shift has been visualized using a 3D scatter plot to present the daily sentiment trend shift for both positive and negative aspects. In Figure 3, we present the graphical representation of the worldwide average sentiment trend shift.

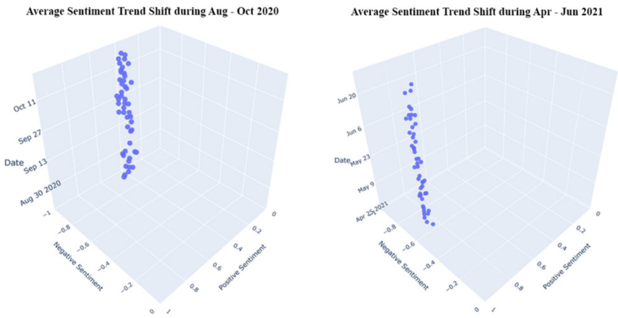

Fig.3: *Graphical representation of the Average Sentiment Trend Shift throughout the world.*

It is evident from the figure that the average sentiment trend from the period Apr – Jun 2021 is shifted towards more positive and less negative compared to the Aug – Oct 2020 phase due to the terrible situation during the initial wave of the Covid-19 pandemic. This change in sentiment reflects the impact of the pandemic situation and the associated challenges faced by individuals and communities during that time.

10. SENTIMENT MODELLING USING HYBRID CONVOLUTIONAL LSTM

10.1 Architecture of Hybrid Convolutional LSTM Model

According to numerous researchers, the LSTM (Long-Short Term Model) network outperforms similar neural network models in conventional sentiment analysis of textual data as it produces the most effective predictions. As a special RNN structure, this model can forecast future predictions based on the various extracted features from the dataset. The data moves through the cell states in LSTM so that this network can accurately recollect or overlook things depending on the hyperparameters associated with it. For the historical datasets, the information gathered over progressive period frames and generally LSTM is proposed to be an efficient methodology to produce forecasts with these datasets. In LSTM, the memory cell c_t acts as an accumulator of the state information. The self-parameterized controlling gates can modify the LSTM cell. Every time for new input, if the input gate i_t is on then its information will be collected to the cell. The past cell information will be forgotten if they forget gate f_t is activated. The output gate o_t controls the flow of the information in the rest of the network by propagating the present cell output c_t to the final state h_t. The three gates and memory cells control the information flow throughout the network to trap the gradient in the cell and prevent them from disappearing too rapidly. The mathematical representation of the status of the gates and cell are presented using the following equations where W denotes the weight matrices, σ denotes the logistic sigmoid activation function, and '$\bigodot$' denotes the Hadamard Product.

$$\left.\begin{aligned} i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci} \bigodot c_{t-1} + b_i) \\ f_t &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf} \bigodot c_{t-1} + b_f) \\ c_t &= f_t \bigodot c_{t-1} + i_t \bigodot tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \\ o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co} \bigodot c_t + b_o) \\ h_t &= o_t \bigodot tanh(c_t) \end{aligned}\right\} \quad (6)$$

Although for handling local dependencies between neighbor words, the LSTM has proven powerful at the time of sentiment classification over large data, it had to deal with excessive redundancy. The proposed hybrid Convolutional LSTM network takes word embeddings as its input and feeds them into *Convolutional* layers to extract local features. After that, the output of the Convolutional model has been transmitted to the LSTM network as input to learn long-term dependencies between the sequence of words for fine-grained sentiment classification. We exploit both Convo-Sequential and Convo-Bidirectional LSTM models for sentiment evaluation of the fine-grained Covid-19 tweets. Figure 4 represents the detailed architecture of the proposed hybrid model.

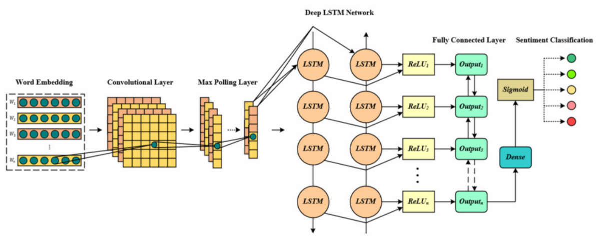

Fig.4: *System Architecture of proposed Hybrid Convolutional LSTM model.*

10.2 Fine-grained Sentiment Classification

In order to perform sentiment classification, we labeled the fine-grained tweets with integer values (from 0 to 4) based on the associated refined sentiment scores since the classifiers cannot fit the non-integer target class values. The cleaned and pre-processed tweets have been labeled as Most Negative (0), Negative (1), Neutral (2), Positive (3), and Most Positive (4) to prepare a new dataset. The dataset has been divided into two sets i.e., X and y sets based on the independent and dependent attributes. Furthermore, we split the X and y sets by an 80:20 ratio to prepare the training sets (X_train, y_train) with 80% of the total data and the test sets (X_test, y_test) with 20% of the total data respectively. This model generated a substantial number of Covid-19 specific words from the new dataset. We have converted these words into word vectors using word2vec by setting the vector dimension as 200 for each collected word within a tweet. New X_train and X_test sets have been prepared based on the evaluated word vectors for further computation. The word vectors and sentiment scores from the new training set (X_train) are supplied as the first layer of inputs into the model.

In this experiment, the *TensorFlow*[4] framework and *Keras*[5] library have been used to add *Sequential* and *Bidirectional* LSTM models along with *Embedding, Convolutional, Max Pooling,* and *Dense* layers. The models have been trained for 6 epochs with two types of outline activation functions with parameters, optimizer, loss, and accuracy. The *Embedding* layer has been used to initialize the words to assign random weights and it learns the embedding to embed all words in the training dataset. In this layer, input length is the maximum length of a tweet that has been used to represent a particular weighted word from a sentence. The *Convolutional layer* with ReLU (Rectified Linear Unit) activation function has been used to take word embeddings as its input from the *Embedding* layer. At the end of this network layer, the *Max Pooling* layer has been used to interpret all the n-gram features to generate sentence representations. The *Dense* or fully connected layer has been used to classify the extracted features of the *Convolutional* layers. Using a dense layer, every current input (neuron) in the network layer is connected to every input in the proceeding layer of the network. We have used the ReLU activation function with 128, 64, and 32 units for the initial set of Dense layers, and the Sigmoid activation function with 5 units for the outermost Dense layer. The final *Dense* layer classifies the tweets within five sentiment classes by predicting the sentiment class of each tweet. During the training, we used 32 batches, 2 verbose, and a learning rate of 0.001 for both models. The dropout technique has been used to prevent the models from overfitting by dropping the irrelevant information from the network which does not contribute to further processing to enhance the performance of the models. Table 5 represents the training and validation accuracy vs. loss using Convo-Sequential LSTM on both datasets respectively.

After completion of the training of the Convo-Sequential LSTM model on Covid-19 Dataset I and Covid-19 Dataset II, finally, the models achieved 90.12% and 90.30% of overall training accuracy whereas 84.52% and 86.58% of validation accuracy on the testing data respectively. Table 6 represents the training and validation accuracy vs. loss using Convo-Bidirectional LSTM respectively.

[4]https://www.tensorflow.org
[5]https://keras.io/

Table 5: *Training and Validation Accuracy vs. Loss using CNN+Seq-LSTM Model.*

	Epochs	Train Loss	Train Accuracy	Val. Loss	Val. Accuracy
Covid-19 Dataset I (Aug – Oct 2020)	*Initially*	23.65%	72.98%	16.57%	82.49%
	2nd	17.25%	80.47%	14.77%	84.95%
	3rd	13.05%	85.21%	15.52%	85.31%
	4th	10.65%	87.72%	18.67%	84.68%
	5th	9.10%	89.24%	21.22%	84.67%
	6th	8.20%	**90.12%**	23.50%	**84.52%**
Covid-19 Dataset II (Apr – Jun 2021)	*Initially*	19.93%	77.93%	14.09%	85.22%
	2nd	15.61%	82.14%	13.21%	86.49%
	3rd	12.22%	86.07%	13.72%	86.92%
	4th	10.10%	88.18%	14.77%	87.12%
	5th	8.77%	89.53%	17.50%	86.99%
	6th	8.01%	**90.30%**	19.49%	**86.58%**

Table 6: *Training and Validation Accuracy vs. Loss using CNN+Bi-LSTM Model.*

	Epochs	Train Loss	Train Accuracy	Val Loss	Val Accuracy
Covid-19 Dataset I (Aug – Oct 2020)	*Initially*	23.35%	73.22%	16.30%	82.88%
	2nd	16.87%	80.95%	15.42%	84.15%
	3rd	12.74%	85.59%	16.56%	84.97%
	4th	10.30%	88.07%	17.68%	84.84%
	5th	8.85%	89.47%	18.66%	84.90%
	6th	7.95%	**90.33%**	22.04%	**85.03%**
Covid-19 Dataset II (Apr – Jun 2021)	*Initially*	19.69%	78.27%	14.11%	85.21%
	2nd	15.36%	82.33%	13.36%	86.47%
	3rd	11.88%	86.29%	13.46%	87.02%
	4th	9.74%	88.54%	15.61%	87.02%
	5th	8.59%	89.72%	16.09%	86.35%
	6th	7.77%	**90.48%**	18.08%	**87.22%**

After completion of the training of the Convo-Bidirectional LSTM model on Covid-19 Dataset I and Covid-19 Dataset II, finally, the models achieved 90.33% and 90.48% of overall training accuracy whereas 85.03% and 87.22% of validation accuracy on the testing data respectively. It has been found that the Convo-Bidirectional LSTM network performs better than the Convo-Sequential LSTM network on both datasets because they can learn each token from the sequence based on both the past and the future context of the token.

There is a substantial loss difference between the training and testing epochs, as can be seen in both figures. This represents small overfitting of the data, which can be explained by the fact that sentiment associated with the tweet streaming keywords varied periodically during the tweet collection phase.

10.3 Classification Report & Confusion Matrix

In this experiment, fine-grained tweets have been considered from both validation datasets to evaluate the classification accuracy during sentiment prediction of the hybrid convolutional LSTM models. Table 7 represents the classification report to present the performance regarding the classification of fine-grained tweets along with the different classes from both first and second phase datasets.

In Figure 5, the heatmaps for the confusion matrix are presented to identify the differences between predicted and actual tweets along with the different classes from both first and second phase datasets.

Table 7: *Classification Report.*

		CNN + Seq-LSTM			CNN + Bi-LSTM			Support
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	
Covid-19 Dataset I (Aug – Oct 2020)	*M. Neg. (0)*	0.79	0.84	0.81	0.85	0.74	0.79	2603
	Neg. (1)	0.84	0.78	0.81	0.80	0.86	0.83	5147
	Neu. (2)	0.98	0.95	0.96	0.98	0.95	0.97	6739
	Pos. (3)	0.64	0.71	0.68	0.67	0.70	0.69	3568
	M. Pos. (4)	0.85	0.87	0.86	0.87	0.87	0.87	5902
	Avg. / Total	**0.85**	**0.85**	**0.85**	**0.85**	**0.85**	**0.85**	**23959**
Covid-19 Dataset II (Apr – Jun 2021)	*M. Neg. (0)*	0.84	0.84	0.84	0.84	0.85	0.84	3145
	Neg. (1)	0.83	0.76	0.80	0.82	0.80	0.81	5111
	Neu. (2)	0.98	0.96	0.97	0.99	0.96	0.97	9462
	Pos. (3)	0.55	0.65	0.60	0.60	0.56	0.58	2944
	M. Pos. (4)	0.89	0.91	0.90	0.88	0.93	0.91	8613
	Avg. / Total	**0.87**	**0.87**	**0.87**	**0.87**	**0.87**	**0.87**	**29275**

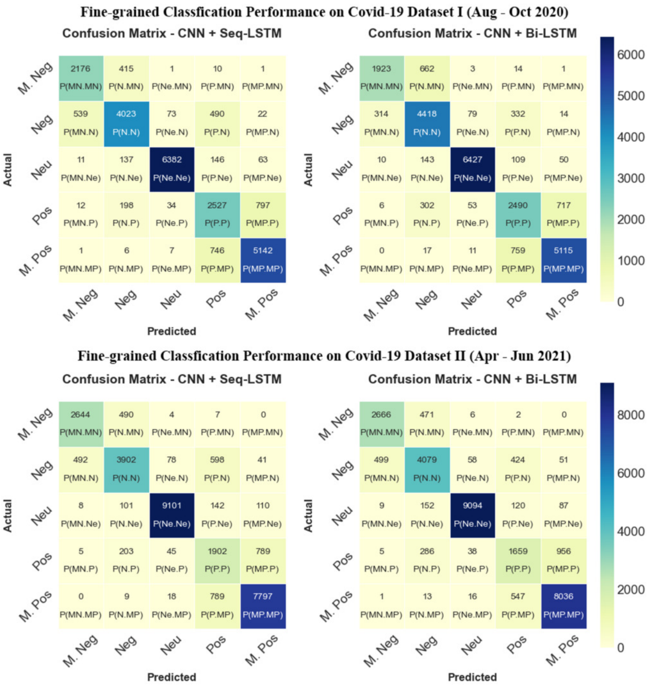

Fig.5: *Confusion Matrix.*

11. COMPARATIVE ANALYSIS

The comparative analysis is presented here to analyze the performance of the proposed deep Convolutional-LSTM models with different previous benchmark experiments on large-scale Covid-19 data as well as several public corpora. Although, in our previous experiments we have already trained these models on two different datasets collected during the phases Aug – Oct 2020 and Apr – Jun 2021, and the models performed consistently well for sentiment prediction of tweets. Through this study, the efficiency and the consistency of these models have been compared to the different State-of-Art approaches we have found during the phase of the literature survey.

11.1 Benchmark Comparison of NLP-based Experiments on Covid-19 Data

Here we have demonstrated performance comparisons with some recent experiments on Covid-19 using different machine learning and deep neural network models along with any baseline algorithm(s). The main objective of this analysis is to identify the classification accuracy for sentiment or emotion prediction achieved by other models considering similar large-scale textual data. Table 8 represents some comparisons of prediction accuracy achieved by various models we have identified throughout the survey.

Table 8: *Performance Comparison between Different ML & DNN Models on Covid-19 Tweets Data.*

Machine Learning or Deep Neural Network Classifiers	Sentiment Classification Accuracy
CNN w. *GloVe* Embeddings (3 classes) [27]	90.67%
CNN + Bi-LSTM w. Naïve Bayes Sentiment Model (5 classes)	**87.22%**
CNN + Seq-LSTM w. Naïve Bayes Sentiment Model (5 classes)	**86.58%**
XLNet on English Tweets (6 classes) [36]	84.70%
Stacking Classifier w. SVC + LR (3 classes) [37]	83.50%
MLP w. pre-trained Word2Vec embedding (3 classes) [35]	83.00%
LDA + Deep LSTM (3 classes) [30]	81.15%
Logistic Regression w. trigrams + Tf-Idf (3 classes) [33]	81.00%
SVM w. Gaussian Membership based Fuzzy logic (3 classes) [33]	79.00%
Deep RNN Model (4 classes) [34]	76.71%

During the analysis, it has been found that our proposed deep CNN + Bi-LSTM and CNN + Seq-LSTM classifiers perform better than different State-of-Art machine learning and deep neural models. But both models achieved slightly lower accuracy on validation data compared to a similar model. However, the performance of these ML or DNN models may vary since different datasets along with several parameters and target classes have been used for the selected experiments in this comparative analysis. The CNN model with *GloVe* Embeddings [27] achieved the highest accuracy than the proposed hybrid classifiers considering 3-class classification.

11.2 Comparative Analysis of State-of-Art Experiments on IMDB Dataset

Stanford's large movie review dataset, also known as IMDB is an open-source public corpus that is widely used for sentiment classification. Here the performance comparisons have been demonstrated with some previous experiments on IMDB review sentiment analysis. Table 9 represents some comparisons between various benchmark experiments.

The main intention behind this analysis is to identify the classification accuracy achieved by other deep learning models. From the analysis, it has been found that the proposed hybrid Convolutional LSTM neural classifiers based on the Naïve Bayes Sentiment model outperform different State-of-Art deep neural models.

11.3 Comparisons with Benchmark Experiments on Other Public Corpora

The performance of our models has also been compared with previous State-of-Art experiments on different open-source public corpora. In order to perform this comparative analysis, two datasets have been considered i.e., Amazon customer review and Stanford Sentiment Treebank (SST). In Table 10, we present the comparisons between the classification performance between previous benchmark experiments and our proposed models.

Table 9: *Performance Comparison between Different Deep Neural Network Models on IMDB Data.*

Deep Neural Network Classifiers	Sentiment Classification Accuracy
CNN + Bi-LSTM w. Naïve Bayes Sentiment Model	**90.44%**
CNN + Seq-LSTM w. Naïve Bayes Sentiment Model	**90.26%**
Ensemble LSTM + CNN [43]	90.00%
CNN + LSTM w. Combined Kernels [39]	89.50%
CNN [43]	89.30%
CNN - LSTM w. Word2Vec embedding [41]	89.20%
LSTM [43]	89.00%
CNN + LSTM w. Vanilla or Multiword Pre-processing [42]	88.90%
CNN w. Multiword Pre-processing [42]	88.10%
CNN [41]	87.70%
MLP [41]	86.74%
Vanilla Neural Network [40]	86.67%
LSTM [41]	86.64%
LSTM w. Tuning and Dropout [44]	86.50%
Recursive RNN [38]	83.88%

Table 10: *Performance Comparison with Benchmark Experiments on Public Corpora.*

Public open-source corpora	Previous Benchmark results	Our results	
		CNN + Seq-LSTM	CNN + Bi-LSTM
Amazon customer review dataset	90.00% [45]	**99.91%**	**99.92%**
Stanford Sentiment Treebank (SST) dataset	86.99% [46]	**90.07%**	**90.25%**

From this comparison, it can be concluded that the proposed models achieved consistent validation accuracy over these datasets for sentiment classification.

12. CONCLUSION & FUTURE SCOPE

In this paper, the various recent NLP research works were discussed on different large-scale event analyses. We proposed a novel experimental approach, mainly focused on Fine-grained Sentiment Classification on Covid-19 tweets using the deep hybrid classifiers. The popularity of a group of words was analyzed using the n-gram model and the most popular Covid-19 specific words were extracted as two main features of the datasets. The tweets were initially categorized based on the sentiment polarities calculated by an NLTK-based sentiment analyzer. However, later the Naïve Bayes sentiment model was developed to assign the fine-grained sentiment scores to the polar tweets based on their probability of occurring in the binary classes, and the tweets are categorized into four sentiment classes. The previously categorized neutral tweets are combined with these fine-grained tweets to assign the refined sentiment scores to all tweets using the fine-grained sentiment classification algorithm. During the analysis, we considered all the Most Positive, Positive, Neutral, Negative, and Most Negative tweets to train the deep learning-based hybrid LSTM network. The dataset was divided into 80:20 ratio i.e., 80% for training and 20% for testing purposes. With the first and second phase Covid-19 datasets, the proposed Convo-Sequential LSTM and Convo-Bidirectional LSTM classifiers were trained for 6 epochs considering certain parameters. Finally, the models achieved 84.52% and 85.03% of validation accuracy respectively for the first phase dataset whereas using the second phase dataset these models obtained validation accuracy of 86.58% and 87.22% respectively.

The second wave of coronavirus infection is considered more dangerous than the first because it has already spread around the world. Many countries have taken different strict measures to stop the spread of the new strain of coronavirus. Analyzing this situation, we hope that we will see a lot of such research work shortly as there has been tireless research based on Covid-19 over the last year and a half. We will also continue our relentless efforts to improve this research work considering specific events related to the Covid-19 pandemic. For future work, we want to extend this work by extracting the tweets from the datasets for which the Convo-Bidirectional LSTM model performs better than the Convo-Sequential LSTM model. We will analyze the impact of different factors on sentiment analysis, such as the demographic information, geographical locations of authors, and other associated parameters of the tweets to retrieve valuable insights from the collected data. Finally, considering the novelty of our work we are eager to publish our Covid-19 data on an open-source data repository platform e.g., GitHub[6] so that other researchers can experiment with our findings and compare their results with our outcomes.

[6] https://github.com/ArunavaKumar

References

[1] S. Tuli, S. Tuli, R. Tuli ans S. S. Gill, "Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing," *Internet of Things*, vol. 11, 2020.

[2] A. D. Dubey, "Twitter Sentiment Analysis during COVID-19 Outbreak," 2021. [Online]. Available: SSRN,`https://ssrn.com/abstract=3572023`

[3] S. Das and A. K. Kolya, "Sense GST: Text mining & sentiment analysis of GST tweets by Naive Bayes algorithm," *2017 Third International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)*, Kolkata, India, pp. 239-244, 2017.

[4] X. Hu, L. Tang, J. Tang and H. Liu, "Exploiting social relations for sentiment analysis in microblogging," in *Proceedings of the sixth ACM international conference on Web search and data mining (WSDM '13)*, pp. 537-546, 2013.

[5] H. Lee, Y. Han, K. Kim and K. Kim, "Sentiment analysis on online social network using probability Model," in *Proceedings of the Sixth International Conference on Advances in Future Internet*, pp. 14-19, 2014.

[6] D. Jurafsky, *Speech and Language Processing*, India:Pearson Education, 2000.

[7] P. Fornacciari, M. Mordonini and M. Tomaiuolo, "Social network and sentiment analysis on Twitter: Towards a combined approach," in *KDWeb*, pp. 53-64, 2015.

[8] P. Zhou, Z. Qi, S. Zheng, J. Xu, H. Bao and B. Xu, "Text Classification Improved by Integrating Bidirectional LSTM with Two-dimensional Max Pooling," 2016. [Online]. Available: arXiv preprint arXiv:1611.06639,`https://doi.org/10.48550/arXiv.1611.06639`

[9] C. Baecchi, T. Uricchio, M. Bertini, and A. Del Bimbo, "A multimodal feature learning approach for sentiment analysis of social network multimedia," *Multimedia Tools and Applications*, vol. 75, no. 5, pp.2507-2525, 2016.

[10] C. Baziotis, N. Pelekis and C. Doulkeridis, "DataStories at SemEval-2017 Task 4: Deep LSTM with Attention for Message-level and Topic-based Sentiment Analysis," *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pp. 747–754, 2017.

[11] Y. Yang, and J. Eisenstein, "Overcoming language variation in sentiment analysis with social attention," *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 295-307, 2017.

[12] B. Nakisa, M. N. Rastgoo, A. Rakotonirainy, F. Maire and V. Chandran, "Long Short Term Memory Hyperparameter Optimization for a Neural Network Based Emotion Recognition Framework," in *IEEE Access*, vol. 6, pp. 49325-49338, 2018.

[13] L. Gong and H. Wang, "When sentiment analysis meets social network: A holistic user behavior modeling in opinionated data," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 1455-1464, 2018.

[14] B. Dahal, S. A. P. Kumar and Z. Li, "Spatiotemporal Topic Modeling and Sentiment Analysis of Global Climate Change Tweets," *Social Network Analysis and Mining*, vol. 9, no. 1, pp.1-20, 2019.

[15] N. C. Dang, M. N. Moreno-García and F. De la Prieta, "Sentiment analysis based on deep learning: A comparative study," *Electronics*, vol. 9, no. 3, pp. 483, 2020.

[16] P. Chauhan, N. Sharma and G. Sikka, "The emergence of social media data and sentiment analysis in election prediction," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1-27, 2020.

[17] X. Ji, S. A. Chun and J. Geller, "Monitoring Public Health Concerns Using Twitter Sentiment Classifications," *2013 IEEE International Conference on Healthcare Informatics*, Philadelphia, PA, USA, pp. 335-344, 2013.

[18] G. Coppersmith, M. Dredze and C. Harman, "Quantifying Mental Health Signals in Twitter," *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, Baltimore, Maryland, USA, pp. 51-60, 2014.

[19] D. Georgiou, A. MacFarlane and T. Russell-Rose, "Extracting sentiment from healthcare survey data: An evaluation of sentiment analysis tools," *2015 Science and Information Conference (SAI)*, London, UK, pp. 352-361, 2015.

[20] D. L. Mowery, A. Park, C. Bryan and M. Conway, "Towards Automatically Classifying Depressive Symptoms from Twitter Data for Population Health," *Proceedings of the Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media (PEOPLES)*, Osaka, Japan, pp. 182-191, 2016.

[21] Y. Zhang, M. Chen, D. Huang, D. Wu and Y. Li, "iDoctor: Personalized and professionalized medical recommendations based on hybrid matrix factorization," *Future Generation Computer Systems*, vol. 66, pp.30-35, 2017.

[22] A. Vij and J. Pruthi, "An automated psychometric analyzer based on sentiment analysis and emotion recognition for healthcare," *Procedia computer science*, vol. 132, pp. 1184-1191, 2018.

[23] J. Lee, J. Kim, Y. J. Hong, M. Piao, A. Byun, H. Song and H. S. Lee, "Health information technology trends in social media: using twitter data," *Healthcare informatics research*, vol. 25, no. 2, p.99 , 2019.

[24] L. Abualigah, H. E. Alfar, M. Shehab and A. M. A. Hussein, "Sentiment analysis in healthcare: a brief review," *Recent Advances in NLP: The Case of Arabic Language*, pp.129-141, 2020.

[25] P. Arora, H. Kumar and B. K. Panigrahi, "Prediction and analysis of COVID-19 positive cases using deep learning models: A descriptive case study of India," *Chaos, Solitons & Fractals*, vol. 139, p.110017, 2020.

[26] I. Arpaci, S. Alshehabi, M. Al-Emran, M. Khasawneh, I. Mahariq, T. Abdeljawad and A. E. Hassanien, "Analysis of twitter data using evolutionary clustering during the COVID-19 pandemic," *Computers, Materials & Continua*, vol. 65, no. 1, pp.193-204, 2020.

[27] S. Das and A. K. Kolya, "Predicting the Pandemic: Sentiment Evaluation and Predictive Analysis of Large-Scale Tweets on Covid-19 by Deep Convolutional Neural Network," *Evolutionary Intelligence*, vol. 15, pp. 1913-1934, 2022.

[28] Md. Y. Kabir and S. Madria, "CoronaVis: A Real-time COVID-19 Tweets Data Analyzer and Data Repository," 2020. [Online]. Available: arXiv preprint arXiv:2004.13932,`https://doi.org/10.48550/arXiv.2004.13932`

[29] J. Samuel, G. G. Ali, M. Rahman, E. Esawi, and Y. Samuel, "Covid-19 public sentiment insights and machine learning for tweets classification," *Information*, vol. 11, no. 6, p.314, 2020.

[30] H. Jelodar, Y. Wang, R. Orji and S. Huang, "Deep sentiment classification and topic discovery on novel coronavirus or covid-19 online discussions: Nlp using lstm recurrent neural network approach," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 10, pp.2733-2742, 2020.

[31] V. K. R. Chimmula and L. Zhang, "Time series forecasting of COVID-19 transmission in Canada using LSTM networks. Chaos," *Solitons & Fractals*, vol. 135, p.109864, 2020.

[32] N. Zheng *et al.*, "Predicting COVID-19 in China Using Hybrid AI Model," in *IEEE Transactions on Cybernetics*, vol. 50, no. 7, pp. 2891-2904, July 2020.

[33] K. Chakraborty, S. Bhatia, S. Bhattacharyya, J. Platos, R. Bag and A. E. Hassanien, "Sentiment Analysis of COVID-19 tweets by Deep Learning Classifiers—A study to show how popularity is affecting accuracy in social media," *Applied Soft Computing*, vol. 97, p.106754, 2020.

[34] M. A. Al-Shaher, "A hybrid deep learning and NLP based system to predict the spread of Covid-19 and unexpected side effects on people," *Periodicals of Engineering and Natural Sciences (PEN)*, vol. 8, no. 4, pp.2232-2241, 2020.

[35] S. Behl, A. Rao, S. Aggarwal, S. Chadha and H. S. Pannu, "Twitter for disaster relief through sentiment analysis for COVID-19 and natural hazard crises," *International journal of disaster risk reduction*, vol. 55, p.102101, 2021.

[36] Q. Yang *et al.*, "SenWave: Monitoring the Global Sentiments under the COVID-19 Pandemic," 2020. [Online]. Available: arXiv preprint arXiv:2006.10842,`https://doi.org/10.48550/arXiv.2006.10842`

[37] M. M. Rahman and M. N. Islam, "Exploring the performance of ensemble machine learning classifiers for sentiment analysis of COVID-19 tweets," in *Sentimental Analysis and Deep Learning: Proceedings of ICSADL 2021*, Singapore: Springer Singapore, pp. 383-396, 2021.

[38] A. Timmaraju and V. Khanna, "Sentiment analysis on movie reviews using recursive and recurrent neural network architectures," *Semantic Scholar*, pp. 1-5, 2015.

[39] A. Yenter and A. Verma, "Deep CNN-LSTM with combined kernels from multiple branches for IMDb review sentiment analysis," *2017 IEEE 8th Annual Ubiquitous Computing, Electronics*

and Mobile Communication Conference (UEMCON), New York, NY, USA, pp. 540-546, 2017.

[40] Z. Shaukat, A. A. Zulfiqar, C. Xiao, M. Azeem and T. Mahmood, "Sentiment analysis on IMDB using lexicon and neural networks," *SN Applied Sciences*, vol. 2, no. 2, pp. 1-10, 2020.

[41] N. M. Ali, M. M. Abd El Hamid and A. Youssif, "Sentiment Analysis for Movies Reviews Dataset Using Deep Learning Models," *International Journal of Data Mining & Knowledge Management Process (IJDKP)*, vol.9, no.2/3, May 2019.

[42] J. Camacho-Collados and M. T. Pilehvar, "On the role of text preprocessing in neural network architectures: An evaluation study on text categorization and sentiment analysis," 2018. [Online]. Available: arXiv preprint arXiv:1707.01780 ,`https://doi.org/10.48550/arXiv.1707.01780`

[43] S. Minaee, E. Azimi and A. Abdolrashidi, "Deep-sentiment: Sentiment analysis using ensemble of cnn and bi-lstm models," 2019. [Online]. Available:arXiv preprint arXiv:1904.04206,`https://doi.org/10.48550/arXiv.1904.04206`

[44] A. M. Dai and Q. V. Le, "Semi-supervised sequence learning," 2015. [Online]. Available:arXiv preprint arXiv:1511.01432, `https://doi.org/10.48550/arXiv.1511.01432`

[45] L. Guner, E. Coyne and J. Smit, "Sentiment analysis for Amazon," *com reviews*, 2019.

[46] X.L. Dong and G. De Melo, "A helping hand: Transfer learning for deep sentiment analysis," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, vol. 1, Melbourne, Australia, pp. 2524-2534, 2018.

Arunava Kumar Chakraborty is working as an Analyst at an EdTech Organization. He completed his M.Tech degree in Computer Science and Engineering from RCC Institute of Information Technology, Kolkata, India in 2021. He was the first class first rank holder within his discipline. Besides this, he has several professional and advanced certifications in Data Science, Data Analytics, Machine Learning, Deep Learning, and Natural Language Processing. His research interests are in the area of NLP, Sentiment Analysis, Large-scale Event Analysis, Information Extraction, Machine Learning, Deep Learning, etc.

Dipankar Das is an Assistant Professor in the Computer Science and Engineering Department, Jadavpur University, Kolkata, India, and Visveswaraya Young Faculty, MeitY, GoI. He served as an Assistant Professor in the CSE, NIT, Meghalaya, GoI from 2012 to 2014. His research interests are in the area of NLP, especially in Emotion and Sentiment Analysis, Machine Translation, Social Texts Analytics, Information Extraction, Machine Learning, Data Science, HCI, etc. He is leading several research projects such as "Sevak" under DST and DRDO, UGC RUSA 2.0 funded by the Government of India. He has had more than 200 publications in top journals, conferences, and workshops.

Anup Kumar Kolya is an Assistant Professor in the Computer Science and Engineering Department, RCC Institute of Information Technology, Kolkata, India since 2014. He received his Ph.D. (Engg.) degree from Jadavpur University, India in 2015. He worked as an SRE in the research Project EILMT funded by DIT, India, GoI (2009-2014). His research interests are in the area of Natural Language and especially in Emotion and Sentiment Analysis, Social Texts Analytics, Information Extraction, ML, Data Science, etc. He has more than 50 publications in top journals, conferences, and workshops and editor and author of several Books and Chapters.