EAI Endorsed Transactions
on Pervasive Health and Technology

Research Article **EAI.EU**

Detection of Misinformation Related to Pandemic Diseases using Machine Learning Techniques in Social Media Platforms

J. Naeem[1], O. M. Gul[1,2,*], I. B. Parlak[3], K. Karpouzis[4], Y. B. Salman[5], and S. N. Kadry[6,7,8,9]

[1] Department of Computer Engineering, Bahcesehir University, Istanbul, Turkey; jawarianaeem19@gmail.com, omermelih.gul@ou.bau.edu.tr
[2] Informatics Institute, Istanbul Technical University, Istanbul, Turkey; omgul@itu.edu.tr
[3] Department of Computer Engineering, Galatasaray University, Istanbul, Turkey; bparlak@gsu.edu.tr
[4] Panteion University of Social and Political Sciences, Greece, kkarpou@panteion.gr
[5] Department of Software Engineering, Bahcesehir University, Istanbul, Turkey; batu.salman@bau.edu.tr
[6] Department of Applied Data Science, Noroff University College, Kristiansand, Norway; skadry@gmail.com
[7] Artificial Intelligence Research Center (AIRC), Ajman University, Ajman, 346, United Arab Emirates
[8] Department of Electrical and Computer Engineering, Lebanese American University, Byblos, Lebanon
[9] MEU Research Unit, Middle East University, Amman 11831, Jordan

Abstract

INTRODUCTION: The advent of the COVID-19 pandemic has brought with it not only a global health crisis but also an infodemic characterized by the rampant spread of misinformation on social media platforms.
OBJECTIVES: In response to the urgent need for effective misinformation detection, this study presents a comprehensive approach harnessing machine learning and deep learning techniques, culminating in ensemble methods, to combat the proliferation of COVID-19 misinformation on Facebook, Twitter, Instagram, and YouTube.
METHODS: Drawing from a rich dataset comprising user comments on these platforms, encompassing diverse COVID-19-related discussions, our research applies Support Vector Machine (SVM), Decision tree, logistic regression, and neural networks to perform in depth analysis and classification of comments into two categories: positive and negative information. The innovation of our approach lies in the final phase, where we employ ensemble methods to consolidate the strengths of various machine learning and deep learning algorithms. This ensemble approach significantly improves the model's overall accuracy and adaptability.
RESULTS: Experimental results underscore the efficacy of our methodology, showcasing marked improvements in detection performance compared to individual models. After applying ensemble learning, we achieve an accuracy of 91% for Facebook data, 79% for Instagram data, 80% for Twitter data and 95% for YouTube data.
CONCLUSION: Our system not only aids in curbing the dissemination of COVID-19 misinformation but also provides a robust framework for addressing misinformation across various contexts on social media platforms.

Keywords: COVID-19, machine learning, deep learning, sentimental analysis, social networks, Term Frequency – Inverse Document Frequency (TF-IDF), Recurrent Neural Network (RNN)

Received on 17 May 2024, accepted on 25 June 2024, published on 28 June 2024

doi: 10.4108/eetpht.10.6459

*Corresponding author. Email: omgul@itu.edu.tr

1. Introduction

Thanks to technological improvements, social media has made knowledge easier to get, changing people's lives. These days, people prefer to learn things through the media. In 2020, the global media user base surpassed 3.6 billion. The amount of people utilising social media will have significantly expanded by 2025 [1]. Numerous advantages come with social networking, such as fast and trouble-free communication, business marketing, and online education. On social media sites, however, there are certain concerns associated with open access. One well-known example is misinformation. Misinformation is a decisive combination for the community. For people who use social media, this is a major problem. Pandemics are the root cause of the worldwide issue of fake news. Social media platforms help people communicate and share creative ideas. By resharing content, social media users can spread ideas and news, which could lead to the spread of misleading information. Social media networking gives people a way to spread misleading information fast. Media outlets are seen to be useful for disseminating vast volumes of unfiltered content, permitting deceit, and therefore raising the possibility that the public will be misled by the transmission of misleading information.

Restricting the spread of false information is a challenging task in the contemporary society. This implies that a sizable audience is being misinformed by ministry members. Thus, false information has affected nearly every aspect of our lives. The dissemination of misleading information during the coronavirus outbreak has been the most worrisome development of the past year. In the contemporary era, we are encountering one of the most contagious pandemics in history, incited by the SARS-CoV-2 contagion, generally renowned as the coronavirus. The World Health Organization (WHO) has officially designated this pandemic as COVID-19 (PAHO, 2020). The term "pandemic" now signifies that this epidemic has transcended international borders, affecting numerous countries, continents, and a substantial portion of the global population. Even today, its impact remains most pronounced in Europe. 2019 saw the discovery of the deadly COVID-19 virus in China, and since then, it has spread over the entire world. This virus has claimed the lives of many people. At first, it was believed that animals were the source of this ailment in humans. Through touch, droplets, and fomites, this virus can transfer from one person to another. There were 4.4 million coronavirus cases as of May 14, 2020; of those, 1.6 million were successfully treated, and 298,000 resulted in death. According to the latest research, there is a widespread spread of misleading news and rumours about COVID-19. It's challenging to distinguish false information from data whose accuracy cannot be questioned [2]. Consequently, misleading information disseminated via social media has significantly exacerbated the coronavirus pandemic situation. When addressing disease outbreaks, the term "infodemic" was used to highlight the dangers of disseminating misleading information.

Even though research into a COVID-19 therapy has started, false information has spread widely on social media, which scientists say is a main source of the epidemic's hazards. False health information constantly puts people's health at danger. In view of the quick development of misinformation spreading, especially during times of global crisis, a great deal of research is thus required to understand the nature and motivations behind the dissemination of false information through social media. Information regarding the proliferation of bogus news on social media is still scarce. Additionally, there is a dearth of study on fake news, but there is a growing body of evidence on false information. It is yet unclear what is causing incorrect information to proliferate on social media platforms. The latest data has shown that inaccurate information on health-related matters is not brand-new. However, the emergence of media platforms that provide unlimited data exchange has resulted in an increase in the medical community's production of incorrect information [3]. Due to the vast quantity of misleading material and its quick dissemination, users are perplexed and look for reliable and accurate information about the outbreak. People from all around the world may now interact thanks to social media sites like Facebook, Instagram, and Twitter. However, because of the sheer volume of data, some of it is made up and some is real. False news misleads users.

By distorting and affecting societal responses, this false information has the potential to expedite epidemic spread [4]. For instance, CNN recently pre-empted a rumor about the potential lockdown of Lombardy, a northern Italian region, as a preventive measure against pandemics. This news broke hours before the official announcement by the Italian Prime Minister. Consequently, a surge of individuals flocked to trains and airports, attempting to leave Lombardy for the southern regions before the lockdown came into effect. This sudden exodus disrupted the government's efforts to contain the epidemic and posed a potential risk of increased contagion. Consequently, a critical research challenge lies in understanding how individuals seek or avoid information and how these decisions impact their behavior, especially when the news cycle, characterized by the rapid and direct spread of information, reshapes the way information is consumed and disseminated [5].

Social media platforms have been introduced to foster communication among human communities, enabling them to exchange ideas, information, knowledge, and various forms of electronic data through these digital channels. Among the most widely used social media applications globally are Facebook, Twitter, Instagram, and YouTube. Despite the vast amount of data available on cybermedia platforms, the materials within them can hold varying outcomes, spanning from destructive psychological impacts over individuals to affirmative psychological influences on their lifespan. These platforms are responsible for providing wrong information during the era of Covid19 which made the situation worse. Misinformation concerning the pandemic poses grave danger to community health and foreign affairs. This threat encompasses a wide spectrum, from the dissemination of harmful health advice, such as the dangerous suggestion of consuming bleach, to agenda-driven conspiracy theories regarding the infection origin [6].

Currently, false information is discovered through manual verification, and cases that raise suspicions are forwarded to specialists for confirmation. It takes too long and is a laborious process for us to get a timely response. Simultaneously, the process of creating counterfeit is becoming more complex. False news is updated often and comes from a wide range of sources. A number of countries have set up fact-finding agencies, such PolitiFact and Agence France-Presse Fact Check, to counteract the dissemination of false information. Companies pay knowledgeable individuals to analyse news data from the media and try to quickly find items that are purposefully false. Erroneous information regarding deadly diseases has proliferated to an epidemic level. These results indicate that their families mistreat others, harbour resentment towards others, and disparage others, which contributes to a host of societal issues [7].

Comprehensive analyses of Facebook data concerning vaccine hesitancy disclose that anti-vaccination crew presently constitute a small group. Regarding COVID-19, the latest study of the highly seen YouTube channels related to the disease revealed that beyond 25\% of the popular videos embrace misleading data, accumulating a staggering 62 million views worldwide. Additionally, there is substantial evidence suggesting that vulnerability to fabrication regarding the disease may be extra prevalent as compared to commonly presumed. For instance, study conducted by Ofcom in the United Kingdom demonstrated that nearly a substantial portion (46\%) of the British populace reported being exposed to false information about the virus. Of specific concern, among those vulnerable, almost two-thirds (66\%) claimed to encounter such misinformation day-to-day. This is challenging because frequent exposure is identified as reinforcing faith in false information. The rationale for undertaking this thesis lies in the ever-evolving landscape of digital. communication dominated by internet communities, consisting of Facebook, Instagram, YouTube, and Twitter. Digital social platforms have become epicentres for the exchange of ideas, opinions, and emotions, reflecting a rich tapestry of human sentiment. In a time when social media has a significant impact on public opinion, policy choices, and social discourse, it is critical to interpret the feelings that are conveyed on these platforms. Sentiment analysis provides insights that go beyond academic curiosity, and its impact on many aspects of life emphasizes the need to learn more about it. Decisions in politics, public relations, marketing, and healthcare can be made using sentiment analysis. Furthermore, the capacity to discriminate between truthful and deceptive emotions is essential for public safety and health in pandemic situations where false information and emotive speech are rife. The primary objectives of this thesis are multi-faceted. First and foremost, the aims to develop and apply advanced sentiment analysis techniques to social media comments across the selected platforms. This involves the extraction and classification of sentiments into two types: positive and negative.

Researcher introduced novel and extensive sentiment dataset named "COVIDSENTI," comprising 90,000 tweets related to COVID-19. These tweets were gathered during the initial phases of the pandemic, spanning throughout February to March 2020. To enhance the utility of the dataset, we meticulously categorized these tweets into three sentiment classes: positive, negative, and neutral. Our analysis encompassed a comprehensive examination of the collected tweets, focusing on sentiment classification. This analysis involved the utilization of diverse feature sets and classifiers. Notably, observed the influential role of negative sentiment in shaping public opinion. During the early stages of the pandemic, for instance, noted a preference for lockdown measures among individuals. However, as anticipated, sentiment dynamics evolved, and by mid-March, observed a notable shift in sentiment patterns [8].

In [9], researchers Carried out analysis of textual data from 13 Reddit communities related to the COVID-19 vaccine, spanning from December 1,2020 to May 15, 2021, using sentiment analysis and Latent Dirichlet Allocational topic modeling. The data was then merged and examined monthly to identify variations in sentiment and uncover latent issues. [10] proposes an approach that involves the analysis of sentiments within the gathered tweets, employing diverse feature sets and classifiers for sentiment classification. This early recognition of COVID-19 reactions within the tweets serves to enhance our comprehension and pandemic handling. We categorize the tweets into three sentiment classes: positive, negative, and neutral. [11] introduced a sentiment analysis strategy to assess public reactions to these videos, employing text mining techniques and machine learning. This study applied various machine learning algorithms to categorize public sentiments and utilized text mining methodologies to uncover latent themes within the comments gathered from YouTube.

1.1. Main Contributions of This Paper

This study assumes significance even when utilizing outdated datasets due to its potential to provide foundational insights into the dynamics of misinformation on social media during the earlier phases of the COVID-19 pandemic. This paper extends the work in our conference paper in [50]. The importance of this research with outdated datasets can be outlined in the following key aspects:

- Examining the patterns of disinformation during the early stages of the epidemic offers a historical perspective on the information-dissemination process and the building of public opinion. Understanding this historical context is crucial to understanding how the information landscape has evolved throughout time.
- Pre-release datasets provide an initial analysis of the distribution and categorization of fraudulent information at a specific point in time. Temporal variations can be observed by comparing this baseline with more recent data points.
- Research utilising outdated datasets offers methodological insights into the challenges and opportunities of post-event misinformation investigation. The findings can teach researchers about limitations and considerations to make when utilising data from the early phases of the pandemic.
- The study can shed light on early detection strategies for misinformation, offering insights into patterns and characteristics that were prevalent during the initial phases of the pandemic. These strategies can be foundational for developing proactive measures in future public health crises.
- Findings from the study can contribute to educational initiatives aimed at enhancing digital literacy and critical thinking skills. The retrospective analysis provides insights into how misinformation was perceived and responded to during a specific period, informing the design of educational programs.

1.2. Organization of This Paper

The rest of this paper is organized as follows. Section 2 provides related literature. Section 3 defines the problem precisely by considering a dataset from four different social media platforms, namely, Facebook, Instagram, Twitter, YouTube. In Section 4, we propose a novel ensemble learning-based approach by using logical regression, SVM, and decision tree. In Section 5, we evaluate the performances of the proposed approach for the four different social media platforms. Section 6 concludes the paper.

2. Related Work

While a formal definition of fake news is still lacking, it is generally classified as "misinformation" or "disinformation." The distinction between the two lies in the intentional creation of false information and the rumours it promotes, all with the intention of helping someone. Misinformation is the term for the spread of incorrect information unrelated to expert verification. False information found on the internet is first considered misinformation because the main objective of its dissemination is to mislead others. The purpose of disseminating misleading information is to create a broad debate among people that has the potential to alter people's ideas while hiding the source and good intentions. Misinformation on the internet can be categorised into several different things, including false news and rumours.

2.1. Fake News

The phrase "fake news" refers to information that has been purposefully misrepresented or created in order to be published as news. Rumours often aim to mislead or deceive people, spread propaganda or further a certain cause, or draw attention and financial support by fabricating sensational headlines and narratives.

Fake news can be presented in a variety of ways, including whole made-up articles, exaggerated or misleading headlines, edited images or videos, and carefully selected quotes. Additionally, it can be disseminated via a range of platforms, including blogs, websites, social media, and conventional news outlets. There are three distinct categories of false news: extensive fabrications, extensive hoaxes, and copious amounts of news. The huge fabrications that feature purposefully false images and information are the most common type of fake news. These pieces are typically disseminated across several social media platforms. More material is created to mislead the public than is found in regular news. Large-scale hoaxes usually include well-known individuals and contain more specific information than the normal news article [12]. To identify fraudulent material, a variety of web-based information is considered, such as online news, Facebook and Instagram posts, public announcements, and other social media content. In addition, there are other websites dedicated to content verification, such Polifact, Snopes, and Boomlive, that aggregate factual information about the accuracy of well-known claims. The websites excel at identifying false material regarding COVID-19 and other contentious subjects [13].

Fake news can have serious consequences such as spreading incorrect information, damaging people's reputations, and inciting social unrest. Additionally, it may diminish public trust in social media sites and the credibility of legal sources.

To stop false information from spreading, it is essential to apply critical thinking skills when assessing the information sources, we utilise and to verify the accuracy of assertions before transmitting them. Reputable news sites also employ fact-checking techniques for confirming their reporting accuracies, which strengthens their reputation as trustworthy sources of information.

2.2. Rumors

A rumour is an unofficially spread fact or tale that typically spreads via word of mouth and lacks supporting evidence. They could be real, imaginary, or a mix of the two. They could appear as urban legends, conspiracy theories, or gossip, among other forms. Rumours can spread quickly and cause significant harm to individuals, particularly when they centre on well-known individuals, occasions, or subjects. They can harm people individually, in groups, and within organisations. They can also provoke societal discontent and bloodshed. Rumours can be challenging to deal with because of their tenacity in the face of contradictory evidence and their difficulty being readily verified or refuted. This is partly because people's concerns, anxieties, and prejudices are sometimes aroused more by rumours than by real events. Online rumours are hard to categorise with precision. As of right now, it is known as "information that has been made public but has not been independently verified and appears to have been manipulated." [14].

According to more recent definitions, rumours are stories that are "told with a significant degree of certainty but are challenging to confirm and may incite fear even when they seem credible." [15]. Studies show that there are two types of rumours: those that initially break news and those that gain traction over time. Long-running rumours are hard to confidently confirm and don't always make the news. Conversely, speculations of breaking news, on the other hand, travel fast and originate from unexpected places. Some deliberately disseminate false information in an effort to perplex or mislead others. These stories are fake but often realistic enough to spread swiftly and pose a threat to public safety. It is critical to recognise these rumours and stop their dissemination. Fighting rumours requires an objective view of the facts and the willingness to refute claims that are not backed up by evidence. Verifying the accuracy of remarks also requires citing reliable sources of information.

2.3. Others

The use of captivating headlines or summaries on social media platforms to entice users to click through to the content that is not accurately reflected by the synopsis is a practice known as "clickbait," which is essential to the spread of fraudulent material [16].

Emails with attention-grabbing subject lines that aim to get the recipient to open or download files are referred to as "spam." The aim of spam is getting personal information.

"Content farm" is website which disseminates low-quality material in a variety of ways in an attempt to boost its financial objectives by attracting a large volume of people quickly [17]. These articles usually include false information and trick readers. These articles' high traffic and search engine optimisation ranking make people easily sidetracked.

A quote is considered out-of-context when it is utilised to bolster a particular argument or viewpoint. If a quote is used in part, its original meaning can be misinterpreted or distorted.

Conspiracy theories are defined as the dissemination of untrue or illogical claims that are typically meant to capitalise on people's anxieties or prejudices. Conspiracy theories can support a range of conflicting viewpoints and have complex explanations that are difficult to disprove.

False flag attacks can be used to exaggerate the seriousness of a situation or offer support for more stringent security protocols. Usually, they involve making up an occurrence or even blaming someone else for it.

The act of posing as a grassroots movement or campaign to further a particular objective or viewpoint is known as "astrofalling." The process of creating fake social media profiles or hiring people to post or leave comments on articles supporting a particular viewpoint is known as "astrurfing."

When someone surrounds themselves with things or people who share their values, it can create echo chambers that amplify prior opinions and make it more challenging to consider competing viewpoints.

2.4. Covid-19 Disinformation

During the COVID-19 outbreak, unprecedented amounts of false information were disseminated. As a result, scholars have investigated the relationship between social media sites and the dissemination of false information throughout pandemic [18]. False information concerning COVID-19 has become much more prevalent online. Fact-check reveals that the messages are not wholly made up; for instance, material is combined from multiple sources to convey fraudulent information by adding a phoney caption to a real photo, greatly enhancing the data's credibility.

Misinformation regarding COVID-19 is often distributed on social media platforms including Instagram, Facebook, and Youtube. These social media platforms started using this method to detect and delete fraudulent content, even if some fake content still shows up on them.

2.5. Detection of Fake News

In recent years, fake news has spread widely across media channels, making it more difficult to distinguish between the two. The most recent research looks on how ML and DL can be used to distinguish between information that is true and untrue.

Two natural language processing (NLP) approaches that can be used to assess news article text and identify any erroneous or misleading information are sentiment analysis and word embeddings. For instance, a neural network can be taught to identify linguistic patterns that are commonly linked to misleading news reports, such as the use of emotive language or the repetition of particular phrases. Another method called metadata analysis can be used to assess the metadata of news items, including the story's source, publication date, and social media interactions. These factors could be evaluated by a neural network to determine the likelihood that the news item is reliable.

A study indicates that one support vector machine (SVM) and some humorous language can be used for detecting phoney content. With a 97% accuracy rate, SVM is useful for detecting fraudulent content in languages other than English [19]. In a different study, SVM is used to detect fraudulent information on Twitter [20]. While random forest and term frequency-inverse document frequency were used in certain experiments to identify false news with an accuracy of 85\%, other studies used random forest tree classifiers to identify fake news [21]. [22] offers a strategy that combines databases of linguistic attributes, category tags, parts-of-speech tags, and sentence features to identify fraudulent content from posts on social media networks.

Furthermore, ethnic tensions resulting from the COVID-19 pandemic have been made worse by the "infodemic" of false material circulating online that targets certain communities for discrimination, causing grave worries among academic and national government institutions. A number of research have proposed an approach to evaluate COVID-19 data [23], [24] using three transformer models (XLNET, ALBER, and BERT; bidirectional encoder representations transformers) to identify fake news. The F1-score for this approach was found to be 98.55\% [25].

In a different study, information fusion is employed to retrieve precise data from websites pertaining to government, health, and media. However, false information can be found on social media platforms. Thirty-nine features were taken out of multi-media content using deep learning techniques, and these features were then used to detect false information about coronavirus. This method increased the results from 59\% to 86\% [26]. Using the CTF and COVID-19 databases, researchers investigated misleading news on Twitter in 2021 [27]. Moreover, they proposed the cross-SEAN model, a cross-stitching-based semi-supervised end-to-end neural attention model. This approach acknowledged the massive volume of unlabeled data. The cross-SEAN method detects fake news by utilising data from external sources. In 2021, the BERT model was presented as a way to detect misleading information about social bot behaviour related to the COVID-19 pandemic. Researchers created the BERT technique to recognise misleading information. Using the transfer learning method, they found bot accounts in a COVID-19 dataset. In order to discriminate between various types of misinformation, a study by [28] included a psycho-linguistic aspect to demonstrate disinformation. The study also shows how this characteristic increases the rate at which false information concerning COVID-19 is recognised on Twitter.

[29] offered a two-step process for identifying fake news: First, we fine-tune a language model based on a transformer with powerful loss functions; second, we use carefully calibrated computations to remove undesirable training occurrences. The initial technique's accuracy was 98\%; however, after data cleaning, accuracy rose to 99\%, indicating the importance of generalisation models.The transformer-based model that had been trained beforehand for COVID-19 misinformation detection was evaluated by the researcher in the same year. The study being given has addressed a variety of transformer and recurrent models along with several related word implanted models. In addition, the effectiveness of the model is evaluated by substituting a different loss function for the study's traditional loss function. It proved that utilising associated language models and bespoke loss functions improves the result. [30] and [31] suggested a mechanism for locating COVID-19 disinformation on social media networks. It employs CNN, LSTM, bi-directional LSTM, and CNN fusion, among other DL techniques. This technique can automatically recognise and classify fake COVID-19 data on social networking sites. To train and evaluate models for detecting misleading information, 21,379 real and fake news items were included in the COVID-19 dataset.

The International Committee of the Red Cross, the World Health Organisation, and the United Nations Children's Fund are just a few instances of trustworthy, international websites from where the real news was gathered. On the other hand, false information is gathered from several fact-checking websites. The result showed that CNN is the best model for identifying misleading information. [32] proposed many DL methods to detect bogus COVID-19 information. The first step was data processing, which involves stemming, tokenization, replacing null values, and noise reduction. The data is represented by a model with inverted document and term frequencies. In the last step, four deep learning algorithms (CNN, RNN, LSTM, and GRU) and eight machine learning (ML) techniques (random forest, K-nearest neighbours, logistic regression, support vector machine, decision tree, Adaboost, naive Bayesian, and neural networks) were measured and estimated.

[33] research endeavors to enhance predictions of future Twitter view evolutions using ensemble deep learning techniques, surpassing previous approaches. Initially, we employ a supervised learning algorithm, the N-gram piled autoencoder, for trait retrieval. These extracted features are subsequently integrated into classification and estimation process, employing an ensemble integration approach that combines methods of machine learning, including decision

trees, support vector machines, random forests, and K-nearest neighbors. The outcomes from each individual method are amalgamated using mean and mode formulas to improve prediction accuracy. Novel approach, combining the N-gram piled encoder with an ensemble ML framework, excels competing techniques such as unigram autoencoders and bigram autoencoders. Experiments are based on a detailed evaluation using dataset obtained from publicly available Twitter data filtered with keywords related to coronavirus.

[34] compiled a comprehensive database comprising 600 questions related to COVID-19, aiming to enhance the accuracy and reliability of the chatbot tailored for Brazilian users. An evaluation of Bot Covid was conducted, assessing its functionality, compatibility, and reliability with a sample of 52 users. The results revealed a high level of satisfaction among the users with the prototype. Additionally, a subset of 20 users was randomly chosen to assess the chatbot's usability, gauged through the System Usability Scale. The average final score, reaching 83.25, indicated excellent usability.

The work in [35] have compiled a substantial dataset comprising 10,742 hand operated sorted comments in the Albanian. Additionally, this work documents our endeavors in designing and constructing a sentiment analyzer, leveraging DL techniques. Hence, they present outcomes of our experiments conducted with our sentiment analyzer, employing Categorizing models, which includes fixed and word vectorization such as fastText and BERT. These models were instructed and authenticated using meticulously gathered dataset. Notably, results indicate that incorporation of Bidirectional Long Short-Term Memory with an concentration mechanism yielded the peak performance in our task, achieved impressive F1 score of 72.09\%.

In [36] the authors gathered a dataset comprising 15,000 comments related to Covid-19. Following a rigorous pre-processing stage, they extracted 7,309 comments for analysis. Subsequently, we employed three supervised ML algorithms, namely SVM, Naive Bayes, and Maximum Entropy, in conjunction with feature retrieval techniques like Bag of Words (BOW), TF-IDF, and word2vec, to classify sentiments expressed in these comments. Among these approaches, Naive Bayes using TF-IDF achieved remarkable results, attaining an accuracy rate of 83.3\% in sentiment classification. During the global coronavirus crises, sentiment analysis emerged as pivotal tool for extracting precious observations from massive data produced on social networking sites.

In [37], author present the work that used deep learning-based language models along with LSTM (long short-term memory) RNN for sentimental analysis during the time of Covid19 in India. This structure includes LSTM model along with global vector as well as BERT model. In this work Selective sentiments were analyzed from 2020 year which includes peak cases of coronavirus cases in India. In this work multiple labels of sentiment classification is used in which many sentiments are analyzed at a time. Results of this paper explained most of the tweets posted is positive and other were not that much in number. It was observed that optimistic and annoying tweets largely prevail in the monthly tweet, while negative sentiments constitute a significantly smaller portion. They used LSTM and global vector for the representation of word for making model of language. They also used BERT for the comparison of results of LSTM and BD-LSTM. They worked on three datasets that includes different states of India.

In paper [38] sentiment analysis is done on tweets based on United States by using a combination of Machine learning and lexicon analysis techniques. The datasets were gathered from RStudio software containing 11858 tweets. This data is from 30 Jan 2020 to 10th May 2020. They used the TextBlob for labelling the dataset as 'positive', 'negative' and 'neutral'. They utilized various feature methods for example bag-of-words (BoW) and TF-IDF in order to preserve the information of expression. Then they applied different classifiers which includes random forest, logistic regression, machine of gradient boosting, tree classifier and SVM to categorize the labelled data. They proposed that TF-IDF increased the performance of SVM and the best method which outperforms is gradient boost with accuracy of 96\%.

In [39], the data is collected from Dec 21 to July 21. This data includes the vaccination details which were commonly used in all over the world. The sentiment of people were analyzed related to all type of vaccines used in the world. For the purpose of analysis NLP and VADER (Valence Aware Dictionary for sentiment reasoner) are used. First they obtained the polarities as positive, negative and neutral which shows data contains 33.96\% positive, 48.49\% neutral and 17.55\% negative responses. An architecture centered around recurrent neural network (RNN), incorporating both long short-term memory (LSTM) and bidirectional LSTM (Bi-LSTM), was employed to evaluate predictive model performance. The results showed LSTM achieving an accuracy of 90.59\%, while Bi-LSTM performed slightly better at 90.83\%.

In [40], author focused on analyzing the reactions of people from various cultures to the. New deadly coronavirus and individual's sentiment regarding to subsequent measures taken by many countries. They utilized LSTM model for finding the polarity of sentiments and the emotions were extracted from tweets which is trained to get accuracy. The utilization of the emoticons provided a distinctive and innovative approach to test supervised deep learning model using Twitter-extracted tweets.

In [41], author proposed three computation techniques for the sentiment analysis for finding the hesitancy in people towards COVID-19 Vaccine. The methods are TextBlob, Azure Machine Learning and VADER. Author used five learning algorithms which are Random Forest, Naïve Bayes, Logistic regression, LinearSVC and Decision tree along with various fusion of three vectorization techniques. These vectorization methods are TF-IDF, Doc2Vec and CountVectorizer. The normalization of vocabulary took three forms: Porter stemming, lemmatization, and a combination of Porter stemming with lemmatization. For each normalization strategy, we created, implemented, and assessed 42 models. This analysis shows rate of people hesitancy towards vaccine decreased over the time. People started feeling warm and positive about Covid19 vaccine. In conclusion, experiment demonstrates that combining TextBlob, TF-IDF, and LinearSVC yields the optimal performance in classifying public sentiment into positive, negative, or neutral. The achieved accuracy, precision, recall, and F1 score are 0.96752,0.96921, 0.92807, and 0.94702, respectively. This indicates that the most effective performance is attained when employing the TextBlob sentiment score, coupled with TF-IDF vectorization and the LinearSVC classification model.

[42] presents the worldwide analysis of sentiments expressed in tweets about the coronavirus, examining how people's feelings in various countries have evolved over the time. In addition, for finding effect of coronavirus on lives, tweets regarding the works which people do at home and the system of online business were scrapped and sentimental change was noticeable. Machine learning algorithms were used for the classification of sentiments which includes LSTM (Long short-term memory), ANN (Artificial Neural Network). Exploratory data analysis was conducted on a dataset detailing daily confirmed cases in some of the severely affected countries. This analysis aimed to draw comparisons between evolution of sentiment and progression of cases from pandemic onset until June 2020.

[43] described how people sentiments are linked with pandemic by utilizing tweets of coronavirus and software of R statistical along with packages of sentimental analysis. They Illustrate insights into the evolution of fear sentiment as COVID-19 approached its peak levels in the United States, utilizing descriptive textual analytics and supported by essential textual data visualizations. Moreover, they provide the overview of methodology in which they used two important machine learning classification ways, in term of textual analysis and compared the different length of classified tweets. The strongest classification of short tweets obtained accuracy of 91\% with the method of Naïve Bayes. Logistic regression gave the accuracy of 74\% of short length tweets but both techniques showed poor performance with the length of longer text tweets.

[44] presents the four models of language based for the purpose of sentiment analysis. This research only contains the Covid19 tweets of Nepal. They used three algorithms which are BernoulliNB, Support Vector Machine and LSTM. The best accuracy given by BernoulliNB which is alpha value of 0.1.

[45] introduces a conversational AI system designed to combat misinformation through a dual strategy. Firstly, it facilitates user access to clear and reliable information synthesized from authoritative sources, using natural language via speech or text. Secondly, the system directly addresses and refutes prevalent myths related to the coronavirus. While initially tailored for use by university staff and students, the system holds promise for broader applications. Through tests evaluating the system's Natural Language Understanding (NLU), they attain an F1-score of 0.906. Additionally, they delve into ongoing research challenges within the realm of conversational Natural Language interfaces for health information.

The critical issue of finding fake short films regarding COVID-19 is examined in research [46], where the text, audio, and visual components of the videos are all heavily laden with misinformation. The majority of conventional techniques for detecting fraudulent films concentrate on looking at video manipulation or comparing the authenticity of content to AI algorithms (like deepfake). These methods, however, are inadequate for our specific problem because most movies are intentionally created and user-generated. We need to successfully extract pertinent information from the distracting and manipulating visual content included in TikTok videos in order to address our challenge. Additionally, we must effectively compile a variety of data from many modalities that are accessible in short films. To address these problems, we introduce TikTec, a multimodal framework for disinformation detection. TikTec uses captions to precisely extract crucial information from distracting video content, therefore masterfully learning the combined false information transmitted by visual and aural components. The TikTec evaluation uses a real-world dataset of COVID-19 films from TikTok. The results demonstrate that TikTec outperforms state-of-the-art baselines in detecting fraudulent COVID-19 short films.

[47] develops an application called Washkaro constitutes a comprehensive approach employing conversational artificial intelligence (AI), machine translation, and natural language processing (NLP) to counter misinformation. Through the utilization of AI, WashKaro delivers precise information aligned with WHO recommendations, presented in locally relevant languages for better comprehension. The principal objective of this research was to evaluate the effectiveness of neural models in text summarization and machine learning techniques in disseminating COVID-19 information consistent with WHO guidelines, thereby addressing the issue of misinformation. Additionally, a secondary goal was pursued, focusing on the development of a symptom assessment tool and segmentation insight to enhance information delivery.

In [48], the author tweets sentiments that were about fear and distress sentiment. This study proposed the notion of web application for filtering out the fearful sentiments from the user tweets, in return provides accurate information with that text analysis is also performed and provides visualization of textual data. In this Naïve Bayes is used which yields 91\% accuracy in the classification of respective user tweets and 74\% is achieved by logistic regression.

One of the first analyses of the images utilised in fraudulent content regarding COVID-19 may be found in [49]. A mixed-methods analysis of ninety-six examples of misinformation from the first three months of 2020 that were rated false or misleading by independent professional fact-checkers identifies and examines six frames and three distinct functions of visuals in misinformation: how they illustrate and selectively emphasise arguments and claims, pretend to present evidence for claims, and impersonate supposedly authoritative sources for claims. It's noteworthy to note that in over half of the purposefully misleading items that were analysed, photos were utilised to support false claims; most of these assertions were mistakenly tagged rather than purposely misrepresented. Our examination turned up a few changed photographs, but they were all produced using simple software; no "deepfakes" or other artificial intelligence-based techniques were discovered. By identifying the various functions of visuals in misinformation and drawing on recent literature on scientific visualisation, this work emphasises significance of paying attention to the visual content in misinformation and expanding the scope beyond a concern with only the representational aspects and functions of misinformation.

This work embarks on a comprehensive investigation into sentiment analysis across prominent social media which covers Instagram, Facebook, YouTube and Twitter, with a primary focus on discerning positive and negative information of COVID-19. The paper leverages a multifaceted approach, tailoring the choice of machine learning and deep learning algorithms suit specific characteristics of each platform. For Instagram and Facebook, machine learning techniques such as decision tree, logistic regression, and SVM are harnessed to extract sentiments from user-generated content.

This methodological choice is influenced by the kind of data that is usually found on these platforms and the need for results that are clear. However, because of the delicate and dynamic nature of speech on Twitter and YouTube, deep learning algorithms are needed. A novel approach to improve the resilience and accuracy of sentiment analysis is given. By stacking ensemble learning, it integrates deep learning and machine learning approaches. By exploiting the unique characteristics of each model, this ensemble technique produces a sentiment analysis framework that is more accurate and complete. Our objective is to contribute to the field of sentiment analysis research and offer valuable insights into the ways in which the public's view of COVID-19 is evolving on different social media platforms. In these historic times, we aim to understand the complex dynamics of emotions communicated online by tailoring our approach to each platform's unique requirements.

3. Problem Definition

Social media sites have become important information and opinion sources since COVID-19 pandemic. As pandemic spreads, it is more important than ever to comprehend the complicated public viewpoints expressed on social media platforms like Instagram, Twitter, Facebook, and YouTube. Absence of a robust and comprehensive sentiment analysis technique that can discern between positive and negative COVID-19-related data across these numerous social apps is one important problem that has to be solved.The challenge arises from the distinct characteristics of each platform, requiring tailored sentiment analysis methods. Reputable for its user-generated content and structured data, Facebook and Instagram employ standard machine learning methods including logistic regression, decision trees, and support vector machines (SVM). On the other hand, deep learning models like CNN, LSTM, RNN, and GRU must be utilised due to the fluid and unstructured nature of talks on YouTube and Twitter. Moreover, obtaining extremely accurate sentiment analysis on these networks might be challenging.

A novel approach to solve this problem is proposed: machine learning and deep learning approaches are combined with stacking ensemble learning. Through combining the benefits of multiple models, this approach aims to improve the precision and reliability of sentiment analysis. By addressing these problems and developing a comprehensive methodology for sentiment analysis tailored to the unique characteristics of each platform, our research seeks to improve our understanding of public perceptions regarding COVID-19 on the internet. The resulting insights can be very beneficial to several stakeholders, including public health agencies, lawmakers, and researchers, as they handle the present pandemic.

3.1. Dataset

The datasets used in this study include a historical data snapshot from specific COVID-19 pandemic times and were obtained from YouTube, Facebook, Instagram, Twitter, and Twitter. It is crucial to acknowledge that the datasets are out-of-date, and inferences made from them may not fairly represent current perspectives or developments about the ongoing pandemic. The results' applicability is limited by this temporal constraint. Despite this, historical data nevertheless provides valuable insights into the public discourse patterns during critical junctures of the pandemic. Future studies could build on and improve the knowledge gained from this inquiry by utilising real-time or more recent datasets.The choice to use earlier COVID-19 misinformation datasets is also influenced by the fact that internet platforms have actively removed or regulated a substantial percentage of disinformation concerning the epidemic.

Many fake postings and comments have been removed as a result of these platforms' prompt action to stop the spread of misleading information. Examining past datasets offers a singular chance to identify the characteristics of disinformation prior to implementing comprehensive content control protocols. By taking this method, the study aims to provide light on the kinds of false information that were common prior to the interventions, providing insights into the ways that misinformation is changing and the efficacy of content moderation techniques over the long run. For sentiment analysis and public opinion research on COVID-19-related content on YouTube, Facebook, Instagram, and Twitter, these datasets are useful resources. Researchers can use this data to look at sentiment trends, debates, and public opinion about different parts of the epidemic. The COVID-19 Dataset consists of user comments on numerous COVID-19 pandemic posts from Facebook, Instagram, Twitter, and YouTube. This dataset includes user opinions, remarks, and conversations around COVID-19.

Youtube Dataset

The size of the dataset is 42851.You can find dataset by following this link:
(https://www.kaggle.com/datasets/seungguini/youtube-comments-for-covid19 related-videos).
The dataset contains a query, which contains keywords used to identify and gather comments related to COVID-19, including terms like 'coronavirus', 'COVID-19', 'pandemic' and 'vaccine', URL,Title, Upload Date, Channel, Views, Likes, Dislikes, Comment count, CommentText, Comment Author, Comment Date.Fig. 1.and Fig. 2. Shows the sample of dataset and t-SNE of test and train data.

#	query	url	title	upload_date	channel	views	likes	dislikes	comment_count	comment_text	comment_author	comment_date	comment_likes	date
1	coronavirus|covid%7C19|pandemic|vaccine	https://www.youtube.com/watch?v=LyRolKjmk9o,Which Coronavirus	Which Coronavirus Vaccine Is More Effective? Comparison Of Vaccines That Are Ready For Use	2021-01-01T09:04:16Z	India Today	8276	58	7	5	OMG 😍💋😍❤️❤️	Brendan Eric	2021-01-01T09:23:59Z	0	2021-01-01 09:04:16+00:00
2	coronavirus|covid%7C19|pandemic|vaccine	https://www.youtube.com/watch?v=LyRolKjmk9o,Which Coronavirus	Which Coronavirus Vaccine Is More Effective? Comparison Of Vaccines That Are Ready For Use	2021-01-01T09:04:16Z	India Today	8276	58	7	5	I love my Indian Mumbai me welcome	Rathore rathore	2021-01-01T09:06:51Z	1	2021-01-01 09:04:16+00:00

Figure 1. Sample of Youtube Dataset

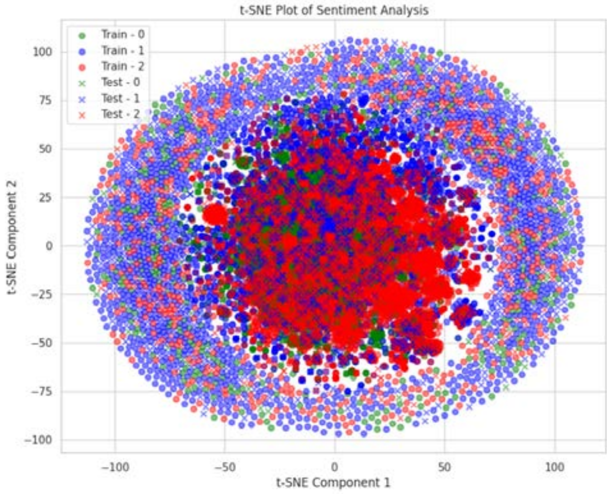

Figure 2. t-SNE of test and train data of Youtube

Twitter Dataset

The size of the dataset is 289414.You can find dataset from this link: https://www.kaggle.com/datasets/gpreda/covid19-tweets). Dataset contains user name: The username or display name of the twitter user who posted the tweet, user location, user created, user followers, user friends, User favorites,user verified, Date, Text. You can find the sample of Twitter Dataset in Figure 3 and Figure 4 shows the t-SNE of test and train data.

#	user_name	user_location	user_description	user_created	user_followers	user_friends	user_favourites	user_verified	date	text
	[illegible]	astroworld	wednesday addams as a disney princess keepin it [illegible]	2017-05-26 05:46:42	624	950	18775	False	2020-07-25 12:27:21	If I smelled the scent of hand sanitizers today on someone in the past, I would think they were so intoxicated…
	Tom Basile [illegible]	New York, NY	Husband, Father, Columnist & Commentator. Author of…	2009-04-16 20:06:23,2253	2253	1677	24	True	2020-07-25 12:27:17	Hey @Yankees @YankeesPR and @MLB - wouldn't it have made more sense to have the players…

Figure 3. Sample of Twitter Dataset.

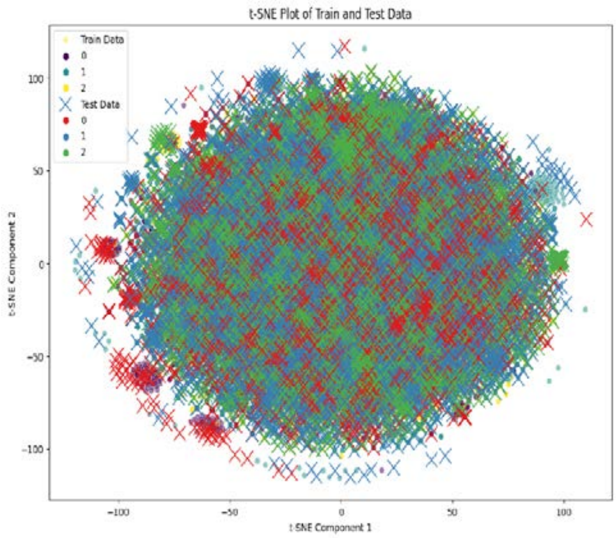

Figure 4. t-SNE of test and train data of Twitter

Instagram Dataset

The size of the dataset is 1327.You can find dataset from this source link: (https://www.kaggle.com/datasets/smartresearch19/instagram-comments). Dataset contains user id, username, comment id, Comment text,Profile URL, Avatar URL, Date. You can find the dataset sample of Instagram in Figure 5 and Figure 6 shows the tSNE.

#	user _id	user _name	comment_id	comment_text	profile_url	avatar_url	-date
	224570964	uniteagainstcovid19	17923841444073642	@aileenper Kia ora, when you are self-isolating, you can exercise outside your home without wearing a face mask...	https://www.instagram.com/uniteagainstcovid19	https://instagram.fhe2-2.fna.fbcdn.net/v/t51.2885-...	3/3/2022, 11:36:13 AM
	428036082	catcat_goose	17911449632450122	@mariaodea everything I believed is a lie...	https://www.instagram.com/catcat_goose	https://instagram.fhe2-2.fna.fbcdn.net/v/t5...	3/3/2022, 12:18:46 PM

Figure 5. Sample of Instagram Dataset.

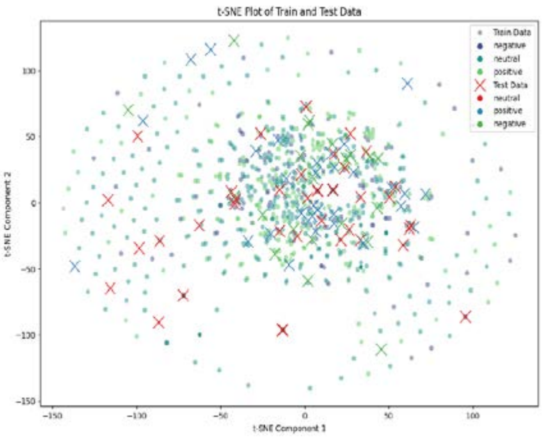

Figure 6. t-SNE of test and train data of Instagram

Facebook Dataset

The size of the dataset is 7029.Dataset is available on following link: (https://www.kaggle.com/datasets/smartresearch19/facebook-comments).

Dataset contains Date, Facebook ID, Likes count, Profile ID, Profile Name, Profile picture, Profile URL, Text. Figure 7 below shows the sample of data and Figure 8 is about t-SNE of Facebook test and train dataset.

#	date	facebookId	likescount	profileId	profileName	profilePicture	profileUrl	text
	2023-04-15T21:20:11.000Z	6462252273820101	882	pfbid02fHTQMpV9qKuBDcLn6GjUBjEoZh...	CaesarPlasencia	https://scontent.ftpf1-1.fna.fbcdn.net/v/t1.6435	https://www.facebook.com/Caesar	I'm just here for all comments of people who don't have an advanced degre
	2023-04-25T18:45:18.000Z	6462252273820101	204		Mark Murphy	https://scontent.ftpf1-1.fna.fbcdn.net/v/t39.	https://www.facebook.com/murphytravels	Should have left Covid off that list

Figure 7. Sample of Facebook Dataset.

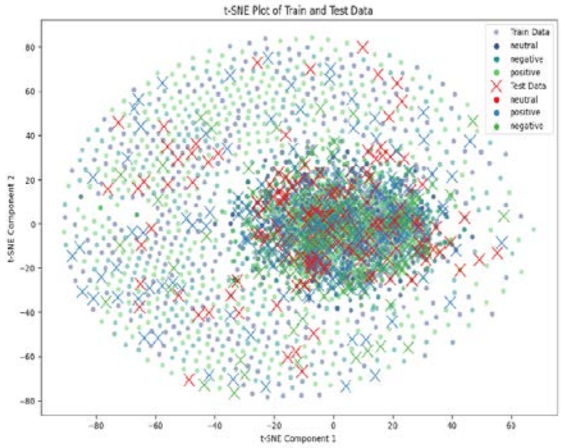

Figure 8. t-SNE of test and train data of Facebook

4. Methodology

Amidst the exponential rise in coronavirus cases, a frantic race is underway among researchers and medical experts to develop quick diagnostics to effectively manage the spread of this disease. The situation has become increasingly leading to a surge in fear and anxiety among individuals who are left wondering about the uncertain future. This atmosphere of uncertainty has given birth to a wave of panic and unease, leaving people feeling helpless and exacerbating the overall situation. Unpredictability and restlessness have become intertwined, as individuals seek answers to pressing questions: When will this deadly disease come to an end? When can we expect a vaccine to become available? When will schools resume? When will it be safe to meet our loved ones?

While COVID-19 has undoubtedly led to heightened levels of anxiety and depression, the internet has developed as a first-hand source for the public to seek solace and information, to some extent. However, the circulation of false information on the internet only serves to further exacerbate people's anxiety and distress which caused the need of figuring out the misinformation and way to remove the negative sentiments about Coronavirus. As this disease killed millions of people because of guidance of wrong information from social media which result people resist to get vaccination. In our study, we took different datasets from social media to find the misinformation as positive, negative and neutral sentiment. We collected four datasets from mostly used social media by people which includes Facebook, Twitter, Youtube and Instagram. We visualize the data to understand the complexity and flexibility. Size of the datasets are big which was the challenging task. So, for that we did some preprocessing data after then we applied deep learning and machine learning algorithms to obtain the model accuracy, moreover we used Ensemble learning approach for enhancing our results.

4.1. Pre-processing

After the collection of datasets, before applying models preprocessing of datasets is done. You can find the figure 9 below which includes the pre-processing steps.

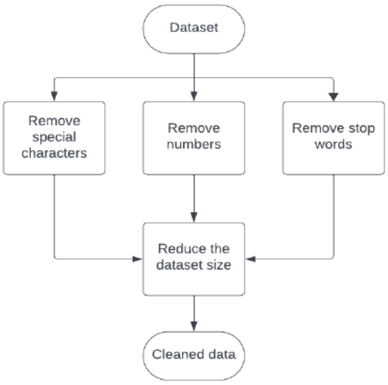

Figure 9. Pre-processing steps

Pre-processing data includes three main components, ‘routine’,’ parser’ and ‘regex’.

Routine Routine is the series of standard and repetitive actions or procedures that are performed on a given input. We applied ‘clean_text’ function, which is a routine that processes a given text string in several steps to clean and pre-process. It performs a sequence of operations to transform the text into a cleaner and more structured format.

Parser Comment text is parsed by using ‘.apply()’ method. This method is used to ‘clean_text’ function to each element (comment text) in the ‘comment_text’ column. It effectively processes and ‘parses’ each comment through the ‘clean_text’ function, transforming the text data.

Regex Regex is the short form of regular expression. We applied ‘re.sub()’ to perform text manipulation. Regex are patterns define a search for, or manipulation of, strings based on certain rules. We used two regular expressions to remove specific elements from the text:
– ’re.sub(r’+. ’,”, str(text))’: This regex is used for removing all the digits from the text.
– ’re.sub(r’[ˆ]’,”, str(text))’: It is used to remove all the characters that are not word characters or spaces. It also removed special characters and punctuation from the text.

4.2. Deep Learning

In deep learning methodology, we build a neural network using TensorFlow and Keras, consisting of several key layers.

Figure 10. Layers to build Neural Network

The process starts with the Input Layer, which acts as the initial point for data entry, specifying the shape of the incoming data and setting the stage for the flow of information through the network. Analogous to the sensory organs of the network, each neuron in the input layer corresponds to an element in the input data, such as pixels in image data. The definition of the input data’s dimensionality is paramount, providing essential information about the structure and layout of the incoming data. While the input layer typically lacks an activation function, it plays a fundamental role in weight initialization, adjusting connections to subsequent layers during training.

Normalisation and other preprocessing steps may be carried out before the input data being fed into the network for optimal training conditions. The input layer in Keras, which denotes the format of the input data, specifies the input layer in code. Understanding the nuances of the input layer is essential since it establishes the basis for the neural network's capacity to analyse and learn from the input data during training. After that, the ReLU activation function is applied to a Convolutional Layer (Conv1D) with 64 filters and a 3 kernel size in order to extract features. Next, a Max Pooling Layer (MaxPooling1D) with a pool size of two is used to down sample the features. The Convolutional Layer (Conv1D) is a crucial part of convolutional neural networks (CNNs), which are made to analyse one-dimensional sequential input, which is commonly encountered in applications like signal processing or time series analysis. In this case, applying 64 filters with a kernel size of three means that each filter scans a region of three consecutive elements in search of local patterns in the input data. After convolution, the Rectified Linear Unit (ReLU) activation function is applied to provide non-linearity and enhance the network's recognition of intricate patterns by replacing negative values with zero.

Following that, a Max Pooling Layer (MaxPooling1D) with a pool size of two is turned on. This layer maintains the maximum value in each pair of adjacent values, allowing for feature downsampling. This process successfully decreases the spatial dimensions of the feature maps without losing any significant information. This particular architectural arrangement is especially well-suited for applications like as time series analysis, where learning effectively depends on the recognition of local patterns. It is excellent at extracting important features from sequential data. The Global Max.Pooling Layer (GlobalMaxPooling1D) further reduces the spatial dimensions to a single value for each feature map in order to extract the most significant properties. Neural networks are specifically designed to assess one-dimensional input, such as time series or sequences, and this component is crucial to their operation. Its primary goal is to reduce each feature map's spatial dimensions to a single value by selecting the largest value within each map. This method effectively condenses the most significant information from each feature map while drastically reducing its spatial dimensions. Based on the context you provided, this layer is commonly utilised in natural language processing (NLP) activities. In order to recover the most crucial information for further processing, the GlobalMaxPooling1D layer is applied after previous layers—such as convolutional or recurrent layers—have extracted features from sequences, such as words or tokens. The fact that you mentioned the Embedding Layer (Embedding) further highlights how important it is to NLP because it transforms categorical input—like words—into a continuous vector space. Parameters such as vocab_size and embedding_dim, which represent the vocabulary size and the dimensionality of the embedding vector for each word, respectively, define the mapping dimensions. When combined, these layers facilitate the effective processing of sequential data, particularly in applications where it is imperative to capture the most salient features. The neural network's capacity to produce high-level feature representations is then significantly enhanced with the addition of a Dense Layer with 128 neurons and ReLU activation. In neural network architecture, a Dense Layer exhibits that 128 neurons are connected to every other neuron in the layer above, enabling a comprehensive examination of potential patterns in the data.

In a calculated move, 128 neurons were added, increasing the layer's capacity to process complicated data and perhaps enhancing the network's capacity to generalise and pick up complex representations. Using the Rectified Linear Unit (ReLU) activation function increases the layer's efficacy. ReLU creates non-linearity and enables the neural network to identify complex correlations in the input by setting negative values to zero. In order to help the neural network better grasp the input data and improve performance, the primary aim by adding Dense Layer is to extract a high-level feature representation from prior layers. In the neural network, final layer design process, output layer, is critical to understanding how the model predicts or categorises input.

The number of classes in the classification job is precisely multiplied by the number of neurons that are specified during this layer's construction. Each neuron in the output layer corresponds to a unique class; this configuration is commonly employed in a one-hot encoding method where the activation of the corresponding neuron indicates the expected class. The output layer makes predictions by utilising the softmax activation function. Soft-max transforms the raw output scores into a probability distribution for each class, ensuring that the entire sum of the predicted values equals one. This approach is particularly helpful in situations involving multiclass classification since it provides a probabilistic interpretation of the model's results. Therefore, the accuracy and reliability of the neural network's final predictions are enhanced when the softmax activation is used in the output layer. The Model is created by specifying the input and output layers, defining the neural network’s architecture. This model can then be compiled with loss functions and optimizers, and subsequently trained on a dataset to perform various tasks, such as classification or regression, depending on the specific problem you are addressing.

4.3. Machine Learning

Arthur Sameul coined the defination of Machine learning as the domain of study that imparts computers with the capability to learn without explicit programming. Machine. Machine learning methodology is a structured approach to solving real world problems using machine learning techniques. At times, when examining the data, it becomes challenging to interpret and extract meaningful information. In such instances, the application of Machine learning proves beneficial. The increasing availability of datasets is driving up the demand for machine learning. Many industries are currently applying machine learning algorithms to extract pertinent information from these databases. Main goal of machine learning is to enable autonomous learning from encountered data. Various studies have delved into methods for machines to learn independently, without explicit programming.

Scholars explore diverse approaches to tackle this challenge, especially in scenarios involving extensive datasets. Machine learning hinges on using various methods to tackle diverse data-related problems. Data scientists emphasise that there isn't a single, universal algorithm that performs well in every circumstance. The specifics of the problem, such as its nature, number of variables, and model type that best matches its characteristics, determine which algorithm should be used. .It begins with a clear problem definition, understanding business objectives, and data collection. Data preparation follows, where data is cleaned, pre-processed, and split into train and test datasets.

In Model selection, a suitable algorithm or architecture is chosen, and model training entails teaching model to acquire patterns from the data of training. Hyperparameter tuning fine-tunes model for optimal result, and evaluation assesses how well it performs using metrics like accuracy or mean squared error. Once satisfied, the model is deployed in real-world settings, with ongoing monitoring and maintenance to ensure it continues to perform well. Ethical considerations and interpretability are vital aspects, as is thorough documentation and reporting to facilitate collaboration and model reproducibility. Machine learning methodology is an iterative process that ensures systematic development and deployment of machine learning solutions. In this work, we applied three machine learning classifiers named as Logistic regression, SVM and Decision tree on Facebook and Instagram dataset.In figure 11, you can find architecture of machine learning methodology.

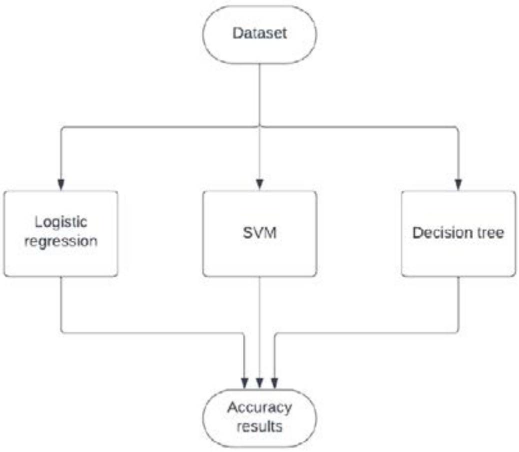

Figure 11. Algorithms used in Machine Learning

Logistic Regression

Logistic regression is a fundamental and broadly used statistical and machine-learning technique for binary and multi-class sorting problems. It's particularly valuable when you need to predict the chance of belonging to a specific class or category. This algorithm is totally different from its own name, it is classed logistic regression but worked as classification model. The primary objective of logistic regression is to model the link between set of independent variables (features) and dependent variable (target) by calculating the probability of an observation belonging to one of the classes. This probability then transformed using logistic function (sigmoid function) to ensure it falls within the range [0, 1]. In binary classification, this probability is typically interpreted as the likelihood of belonging to the positive class. In this paper, the logistic regression model trained by utilizing the training dataset. Following the training phase, the model is employed to make predictions on the test dataset.

Support Vector Machine (SVM)

Support Vector Machine (SVM) is a versatile machine learning algorithm applicable to tasks involving classification and regression. The primary concept revolves around finding a hyperplane that maximizes the margin between classes or achieves the optimal fit for data points.SVMs can efficiently work in high-dimensional spaces. In classification, SVM seeks to separate classes while minimizing misclassifications with the regularization parameter C. SVMs handle linear and nonlinear problems and are valuable in, text analysis. In this paper, SVM is used to classify the COVID-19 comments.The first dataset is divided into features and target labels and then split into testing and training data. The svc object is then trained by employing the training dataset with the fit method. This phase finds the optimal decision boundarythat separates the data points into separate categories while optimizing the margin between them. After training the SVM classifier, predictions are generated for the test dataset.

Decision Tree

A decision tree is straightforward algorithm suited for both classification and regression purposes. A decision tree organizes data in a hierarchical tree-like format,with nodes stands for features and branches indicates decisions. Decision tree simplicity and ease of interpretation make them effective for elucidating complex decision-making procedures. In this paper, a decision tree classifier is used to analyze the comment of COVID-19. First, feature matrix and target matrix are created from dataset. Employing the fit method, decision tree classifier is trained with the training data. During training, the Decision Tree algorithm constructs a tree structure that divides the data iteratively according to the most informative features, to minimize impurity or maximize information gain. After training the Decision Tree classifier make prediction on test data.

4.4. Proposed Approach

We used the stacking Classifier Ensemble Methodology. We used Logistic Regression classifier as the meta-model. The primary function of the meta-model is to merge predictions from base models, resulting in a final prediction. Logistic Regression is selected for its simplicity and interpretability, making it an appropriate choice for aggregating predictions. The stacking classifier is trained on the training data. Throughout the training process, each of the base models is fitted to the training data to generate individual predictions. When the training process is finished, the stacking classifier is tasked with predicting outcomes on the test dataset. The ensemble combines the predictions from the base models using the meta-model (Logistic Regression) to produce the final predictions. The accuracy of the stacking classifier's predictions is evaluated by comparing them to the true labels in the test dataset. The accuracy score is calculated using the accuracy score function. Ensemble learning methodology is given in Figure 12 and 13 below.

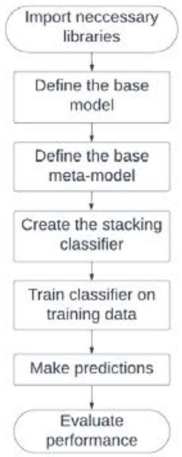

Figure 12. Ensemble learning Methodology

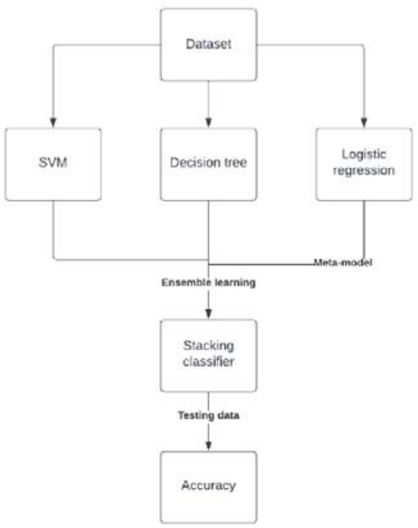

Figure 13. Ensemble Learning Framework

5. Numerical Results

We share the outcome of our analysis of the COVID-19 sentiment dataset gathered from diverse social networking sites like Instagram, Facebook, Twitter, and YouTube. Our analysis focused on identifying positive and negative sentiments related to COVID-19.We applied deep learning on Twitter and YouTube and machine learning on Facebook and Instagram.

5.1. Facebook

The Facebook dataset contains 7029 comments. For analysis, three machine learning including Decision Tree, Logistic Regression, and SVM are applied to the Facebook dataset. Our aim is to evaluate the performance of these algorithms. To do so, datasets are partitioned into training and testing subsets. The training set is 90\% and the testing set is 10\% of the dataset.

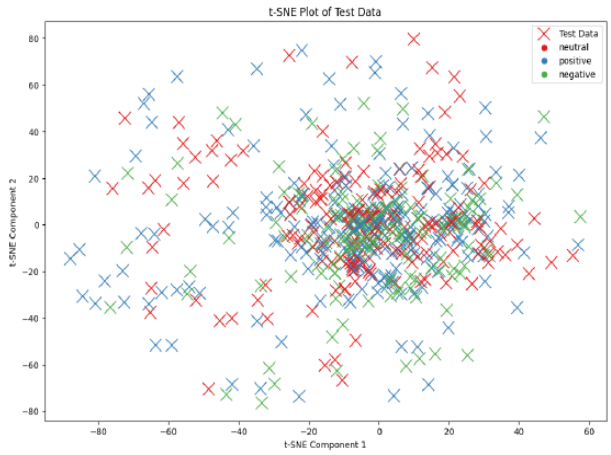

Figure 14. t-SNE test data of Facebook.

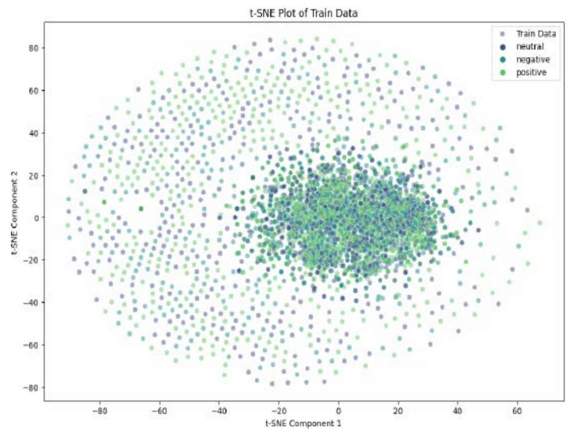

Figure 15. t-SNE train data of Facebook.

Logistic Regression

By applying logistic regression, we get an accuracy of 84.7%. The classification report can be found in Table 1.

Table 1. Classification Report of Facebook Posts Using Logistic Regression

	precision	Recall	F1	score
Negative	0.88	0.75	0.81	167
Neutral	0.75	0.96	0.84	187
Positive	0.93	0.83	0.88	236
Accuracy	0.85	590		
Macro avg	0.86	0.85	0.84	590
Weighted avg	0.86	0.85	0.85	590

In the case of negative class, the precision is 0.88, indicating that model's predictions of negative are accurate 88% of the time. In the case of the 'neutral' class the recall score is 0.96, indicating the model accurately detected 96%

of the actual ‘neutral’ instances. In the ‘positive’ class F1-score is 0.88 signifying a favorable equilibrium between precision and recall. The model’s overall accuracy stands at 85% indicating that it made correct classification for 85% of the total instances of the dataset.

Support Vector Machine (SVM)

By applying SVM, we get an accuracy of 76.9%, as shown in Table 2.

Table 2. Classification Report of Facebook posts using SVM

	precision	Recall	F1	score
Negative	0.53	0.89	0.66	99
Neutral	0.93	0.71	0.80	244
Positive	0.82	0.78	0.80	247
Accuracy	0.77	590		
Macro avg	0.76	0.79	0.75	590
Weighted avg	0.81	0.77	0.78	590

For the “negative” class, precision is 0.88. This implies that when model makes a prediction of ‘negative’ for an instance, it is accurate roughly 88% of the time. For the “neutral” class, the precision is 0.75. For ‘neutral’ predictions, the model is accurate in roughly 75% of cases. Regarding the ‘positive’ class, a precision score of 0.93% demonstrates model’s accuracy in predicting ‘positive’ around 93% of the time. In ‘negative’ class a recall of 0.75 implies that model precisely pinpointed 75% of actual ‘negative’ instances. In ‘neutral’ class a recall score is 0.96 indicates that the mode accurately recognized 96% of actual ‘neutral’ instances. For the “positive” class, recall is 0.83, meaning that the model correctly finds 83% of the actual “positive” instances. For the negative class, F1-score is 0.81. For neutral class, F1-score is 0.84. For positive class, F1-score is 0.88.The result of the model is 0.85, it correctly classified 85% of the total instances in the dataset.

Decision Tree

By applying Decision tree classifier, we get an accuracy of 76.9\%. Table 3 shows the classification results.

Table 3. Classification Report of Facebook posts using Decision trees

	precision	Recall	F1	score
Negative	0.80	0.78	0.79	167
Neutral	0.87	0.98	0.93	187
Positive	0.88	0.81	0.85	236
Accuracy	0.86	590		
Macro avg	0.85	0.86	0.85	590
Weighted avg	0.86	0.86	0.86	590

Precision of 0.80 for the ‘negative’ class indicates that the model anticipates ‘negative’. It is precise in roughly 80\% of instances. Model predicts ‘negative’ for an instance in ‘negative’ segment, it is accurate approximately 80\% of time. For the ``neutral" section, precision is 0.87, clarify that it is correct about 87\%. Precision value for ‘positive’ class is 0.88\%. In case ‘negative’ class, 0.78 is the recall score, indicating 78\% of the actual ``negative" instances. For the ``neutral" class, recall is 0.98, suggesting 98\% of the actual ``neutral" instances. For the ``positive" section, recall is 0.8. The F1-score for negative, neutral, and positive are 0.79, 0.93 and 0.85. With an overall accuracy of 0.86 the model correctly classified 86\% of total instances in the dataset. After applying ensemble learning, achieved accuracy increases to 91\%.

5.2. Instagram

The Instagram dataset contains 1327 comments. Machine learning techniques including Decision Tree, Logistic Regression, and SVM are applied to Instagram dataset. Our aim to evaluate the performance of these algorithms. To do so, In the dataset 90\% is allocated for training, and 10\% is designated for testing.

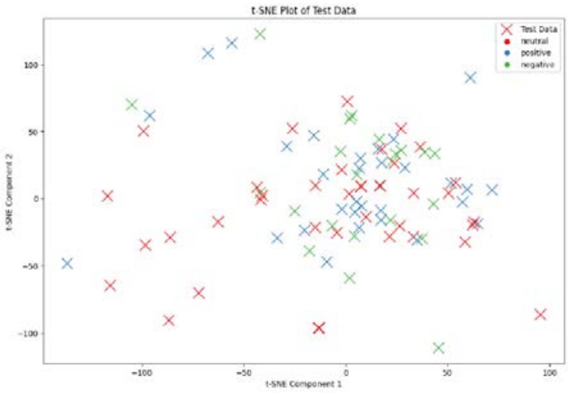

Figure 16. t-SNE test data of Instagram.

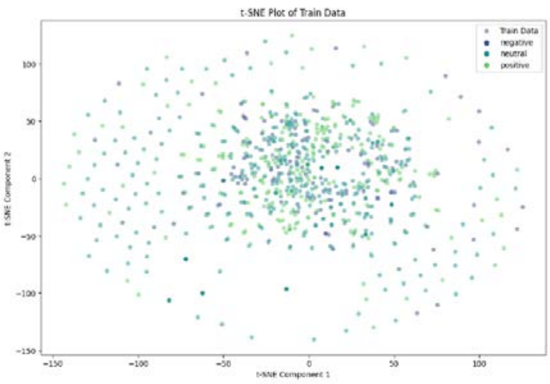

Figure 17. t-SNE train data of Instagram.

Logistic Regression

By applying logistic regression, we get an accuracy of 74.3%. The classification report is shown in Table 4.

Table 4. Classification Report of Instagram Posts Using Logistic Regression

	precision	Recall	F1	score
Negative	0.14	0.50	0.22	4
Neutral	0.98	0.72	0.83	82
Positive	0.59	0.85	0.70	27
Accuracy	0.74	113		
Macro avg	0.57	0.69	0.58	113
Weighted avg	0.86	0.74	0.78	113

For the “negative” class, the precision is 0.14. This shows that prediction for ‘negative’ instance is correct about 14% of the time. The precision score for “neutral”, and “positive” is 0.98, and 0.59 indicates that the efficiency of model finding positive is 98% and 59%. In the “negative” class, the model achieved a recall of 0.50, meaning it accurately identified half of the actual “negative” instances. When considering the “neutral” class a recall score of 0.72 indicates that the model accurately recognized 72% of the “neutral” instances. The “positive” class, recall is 0.85 demonstrating model identifies 85% of “positive” occurrences.

Support Vector Machine (SVM)

By applying SVM, we get an accuracy of 69.9%, as shown in Table 5.

Table 5. Classification Report of Instagram Posts Using SVM

	precision	Recall	F1	score
Negative	0.00	0.00	0.00	0
Neutral	0.90	0.69	0.78	78
Positive	0.64	0.71	0.68	35
Accuracy	0.70	113		
Macro avg	0.51	0.47	0.49	113
Weighted avg	0.82	0.70	0.75	113

The precision result for ‘negative’ is 0.50 which explains that model predicts an instance as "negative," is valid about 50\%. Precision of ‘Neutral’ class is 0.93, shows when model is predicted as "neutral," 93\%. “Positive" class attained a precision of 0.62 which means 62\%. In the case of ‘negative’ class, the recall is 0.44, signifying the model accurately detected 44\% of negative cases. The recall score for ‘neutral and ‘positive is 0.80, 0.89. It shows, model accurately identified 80\% of ‘neutral’ occurrences, 89\% of "positive" instances. the "negative" class, The F1-score of ‘neutral’ and ‘positive’ are 0.78 and 0.68. The total accuracy of the framework is 0.77, which means that it correctly identified 77\% of the total instances in the dataset. The accuracy obtained after ensemble learning is 79\%.

Decision Tree

By applying Decision tree classifier, we get an accuracy of 76.9%. Table 6 shows the results for this algorithm.

Table 6. Classification Report of Instagram Posts Using Decision Tree

	precision	Recall	F1	score
Negative	0.50	0.44	0.46	16
Neutral	0.93	0.80	0.86	70
Positive	0.62	0.89	0.73	27
Accuracy	0.77	113		
Macro avg	0.68	0.71	0.69	113
Weighted avg	0.80	0.77	0.77	113

The precision of “negative” class is 0.14, it suggests that the model is correct about 14% of the time when “negative”. Regarding to “neutral” class, score of precision is 0.98 demonstrates model’s accuracy in predicting ‘neutral’ about 98% while “positive” class, precision is 0.59, suggesting, prediction of “positive,” case is 59%. Recall for “negative”, “neutral”, and “positive” class, recall is 0.44,0.93, and 0.89. The F1-score for classes, ‘Negative’, ‘Neutral’, and ‘Positive’ is 0.46, 0.86 and 0.73. There are 4 instances of “negative.” The efficiency of the model is 0.74, shows that it correctly identified 74% of the total instances in the dataset.

5.3. Twitter

The Twitter dataset contains 289414 tweets; after applying Deep Learning, we obtain the results shown in Table 7.

Table 7. Classification Report of Twitter Posts

	precision	Recall	F1	score
Negative	0.78	0.74	0.74	5406
Neutral	0.76	0.81	0.81	8703
Positive	0.85	0.81	0.81	8037
Accuracy		0.79	0.79	22146
Macro avg	0.78	0.78	0.78	22146
Weighted avg	0.80	0.79	0.79	22146

There are three classes tagged as 0, 1 and 2; the recall score for Class 0 is 0.74, while for Class 1 recall is 0.86. The third class (Class 2) gets a 0.78 score of recall, meaning that the model correctly identified 78% of the tweets. Class 0, 1, 2 generate an F1 score equal to 0.74, 0.81 and 0.81, respectively. The final accuracy obtained by the model is 79%. After applying the ensemble learning, we get 80% accuracy.

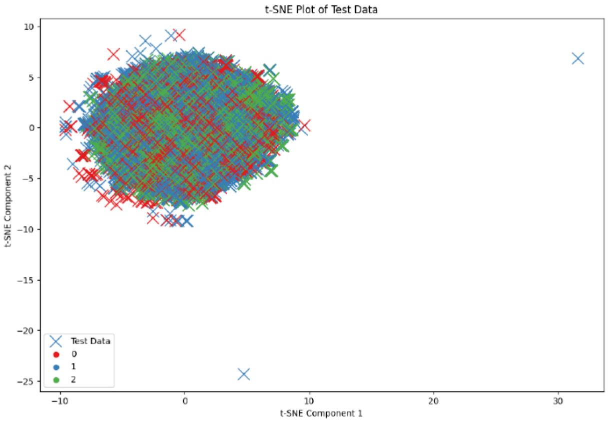

Figure 18. t-SNE test data of Twitter.

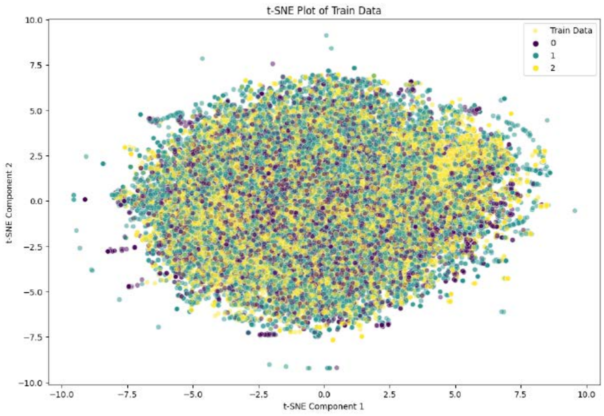

Figure 19. t-SNE train data of Twitter.

5.4. Youtube

The YouTube dataset contains 42851 comments. After removing empty or nonsensical entries, the Deep Learning network is applied. We obtain the results shown in Table 8.

Table 8. Classification Report of Youtube Posts

	precision	Recall	F1	score
Negative	0.91	0.89	0.90	1190
Neutral	0.96	0.96	0.96	1931
Positive	0.93	0.95	0.94	1997
Accuracy			0.94	5118
Macro avg	0.94	0.93	0.78	5118
Weighted avg	0.94	0.94	0.79	5118

When the model identifies Class 0, accuracy is about 91%, while for Class 1, precision is 0.96, suggesting that Class 1 is correct for about 96% of the posts. Finally, for Class 2 precision is 0.93, meaning that the model is correct at about 93%. Class 0 achieved recall 0.89, illustrating that the model identified 89% of the actual instances. Class 1 and 2 gained recall 0.96 and 0.95 suggesting that the model correctly identified 96% and 95% of the factual occurrences. Class 0 achieved an F1-score of 0.90, while class 1 attained 0.96 and Class 2 reached 0.94. The complete accuracy of the model is 0.94, revealing that it successfully identified 94% of the total instances in the dataset. After applying ensemble learning, accuracy is improved to 95%.

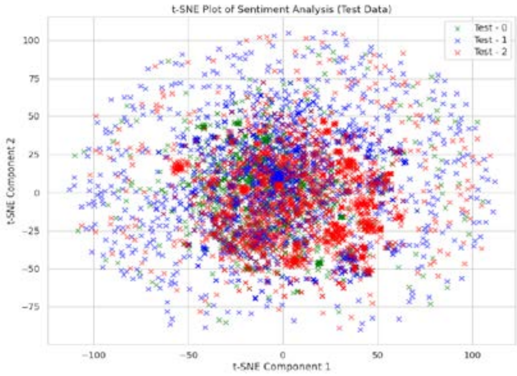

Figure 20. t-SNE test data of Youtube.

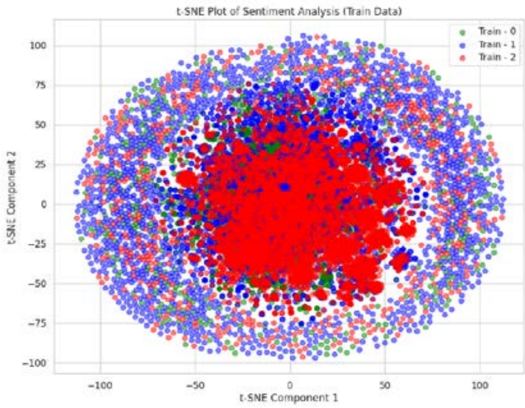

Figure 21. t-SNE train data of Youtube.

6. Conclusion

During the COVID-19 pandemic, this study analyzed sentiment on Instagram and Facebook using conventional machine learning methods and employed deep learning techniques for Twitter and YouTube due to their unstructured content. The research introduced stacking ensemble learning to enhance sentiment analysis accuracy by combining machine and deep learning models; this method proved to be the best method for improving the accuracy for Facebook, Twitter, Instagram and YouTube content, improving detection accuracy to 91\% on Facebook, 79\% for Instagram, 95\% of YouTube and 80\% for Twitter.

Future research can investigate the use of robotic agents or chatbots to automatically fact-check and verify information posted on social media; these robots could interact with users to provide real-time information verification. In the same context, research can investigate the role of social bots in spreading fake news within online social networks and analyze their impact on robotic communication systems, developing strategies to detect and mitigate their influence.

Acknowledgements.
This paper is extended from our conference paper [50] which was presented in the 7th EAI International Conference on Robotics and Networks (EAI ROSENET 2023).

References

[1] Patwa, P., Sharma, S., Pykl, S., et al. (2021). Fighting an infodemic: Covid-19 fake news dataset. In Combating Online Hostile Posts in Regional Languages during Emergency Situation: First International Workshop, CONSTRAINT 2021, Collocated with AAAI 2021, Virtual Event, February 8, 2021, Revised Selected Papers 1, Springer International Publishing, pp. 21–29.
[2] Huynh, T. L. (2020). The COVID-19 risk perception: a survey on socioeconomics and media attention. Economics Bulletin, 40(1), 758–764.
[3] Waszak, P. M., Kasprzycka-Waszak, W. and Kubanek, A. (2018). The spread of medical fake news in social media–the pilot quantitative study. Health Policy and Technology, 7(2), 115–118.
[4] John, T. Ben Wedeman, C. Italy prohibits travel and cancels all public events in its northern region to contain Coronavirus,(2020 (accessed April 9, 2020)).
[5] Alamoodi, A. H., Zaidan, B. B., Zaidan, A. A., Albahri, O. S., Mohammed, K. I., Malik, R. Q., ... \& Alaa, M. (2021). Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: A systematic review. Expert systems with applications, 167, 114155..
[6] Roozenbeek, J., Schneider, C. R., Dryhurst, S., Kerr, J., Freeman, A. L., Recchia, G., & Van Der Linden, S. (2020). Susceptibility to misinformation about COVID-19 around the world. Royal Society open science, 7(10), 201199.
[7] Velasquez, N., Leahy, R., Restrepo, N. J., et al. (2021). Online hate network spreads malicious COVID-19 content outside the control of individual social media platforms. Scientific Reports, 11(1), 11549.
[8] Naseem, U., Razzak, I., Khushi, M., Eklund, P. W., \& Kim, J. (2021). COVIDSenti: A large-scale benchmark Twitter data set for COVID-19 sentiment analysis. IEEE transactions on computational social systems, 8(4), 1003-1015.
[9] Melton, C. A., Olusanya, O. A., Ammar, N., \& Shaban-Nejad, A. (2021). Public sentiment analysis and topic modeling regarding COVID-19 vaccines on the Reddit social media platform: A call to action for strengthening vaccine confidence. Journal of Infection and Public Health, 14(10), 1505-1512.
[10] Jalil, Z., Abbasi, A., Javed, A. R., Badruddin Khan, M., Abul Hasanat, M. H., Malik, K. M., \& Saudagar, A. K. J. (2022). COVID-19 related sentiment analysis using state-of-the-art machine learning and deep learning techniques. Frontiers in Public Health, 9, 812735.
[11] Lekshmi, S., \& Anoop, V. S. (2022, June). Sentiment analysis on COVID-19 news videos using machine learning techniques. In Proceedings of International Conference on Frontiers in Computing and Systems: COMSYS 2021 (pp. 551-560). Singapore: Springer Nature Singapore.
[12] Rubin, V. L., Chen, Y. and Conroy, N. K. (2015). Deception detection for news: three types of fakes. Proceedings of the Association for Information Science&Techno.,52(1),1–4.
[13] Chakraborty, T., Shu, K., Bernard, H. R., Liu, H. and Akhtar, M. S. (eds.) (eds.). (2021). Combating Online Hostile Posts in Regional Languages during Emergency Situation: First International Workshop, CONSTRAINT 2021, Collocated with AAAI 2021, Virtual, February 8, 2021, Revised Selected Papers(Vol. 1402). Springer Nature.
[14] Chen, M. Y., Lai, Y. W. and Lian, J. W. (2022). Using deep learning models to detect fake news about COVID-19. ACM Transactions on Internet Technology.
[15] DiFonzo, N. and Bordia, P. (2007). Rumor, gossip and urban legends. Diogenes, 54(1), 19–35.
[16] Hua, W., Wang, Z., Wang, H., Zheng, K. and Zhou, X. (2016). Understand short texts by harvesting and analyzing semantic knowledge. IEEE Transactions on Knowledge and Data Engineering, 29(3), 499–512.
[17] Shu, K., Mahudeswaran, D., Wang, S., Lee, D. and Liu, H. (2020). Fakenewsnet: a data repository with news content, social context, and spatiotemporal information for studying fake news on social media. Big Data, 8 (3), 171–188.
[18] Huynh, T. L. (2020). The COVID-19 risk perception: a survey on socioeconomics and media attention. Economics Bulletin, 40(1), 758–764.
[19] Ahmed, H., Traore, I. and Saad, S. (2017). Detection of online fake news using n-gram analysis and machine learning techniques. In Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments: First International Conference, ISDDC 2017, Vancouver, BC, Canada, October 26–28, 2017, Proceedings 1. Springer International Publishing, pp. 127–138.
[20] Nikam, S. S. and Dalvi, R. (2020). Machine learning algorithm based model for classification of fake news on twitter. In 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC). IEEE, pp. 1–4.
[21] Jehad, R. and Yousif, S. A. (2020). Fake news classification using random forest and decision tree (j48). Al-Nahrain Journal of Science, 23(4), 49–55.
[22] Hua, W., Wang, Z., Wang, H., Zheng, K. and Zhou, X. (2016). Understand short texts by harvesting and analyzing semantic knowledge. IEEE Transactions on Knowledge and Data Engineering, 29(3), 499–512.
[23] Sharma, M. K., et al. "Post-symptomatic detection of COVID-2019 grade based mediative fuzzy projection". Computers and Electrical Engineering, vol. 101, 2022.
[24] Zhou, L., et al. (2023). Artificial neural network dual hesitant fermatean fuzzy implementation in transportation of COVID-19 vaccine. Journal of Organizational and End User Computing (JOEUC), IGI Global, 35(2).
[25] Gundapu, S. and Mamidi, R. (2021). Transformer based automatic COVID-19 fake news detection system. ArXiv Preprint ArXiv:2101.00180.
[26] Iwendi, C., Mohan, S., Ibeke, E., Ahmadian, A. and Ciano, T. (2022). Covid- 19 fake news sentiment analysis. Computers and Electrical Engineering, 101, 107967.
[27] Heidari, M., Zad, S., Hajibabaee, P., et al. Bert model for fake news detec- tion based on social bot activities in the covid-19 pandemic. In 2021 IEEE 12th Annual Ubiquitous Computing, Electronics \& Mobile Communication Conference (UEMCON). IEEE, 2021, pp. 0103–0109.

[28] Mahbub, S., Pardede, E. and Kayes, A. S. M. (2022). COVID-19 rumor detection using psycho-linguistic features. IEEE Access, 10, 117530– 117543.

[29] Bang, Y., Ishii, E., Cahyawijaya, S., Ji, Z. and Fung, P. Model generalization on COVID-19 fake news detection. In Combating Online Hostile Posts in Regional Languages during Emergency Situation: First International Workshop, CONSTRAINT 2021, Collocated with AAAI 2021, Virtual Event, February 8, 2021, Revised Selected Papers 1. Springer International Publishing; 2021. pp. 128–140.

[30] Hande, A., Puranik, K., Priyadharshini, R., Thavareesan, S. and Chakravarthi, B. R. Evaluating pretrained transformer-based models for COVID-19 fake news detection. In 2021 5th International Conference on Computing Methodologies and Communication (ICCMC).IEEE, 2021, pp. 766–772.

[31] Tashtoush, Y., Alrababah, B., Darwish, O., Maabreh, M. and Alsaedi, N. (2022). A deep learning framework for detection of COVID-19 fake news on social media platforms. Data, 7(5), 65.

[32] Bangyal, W. H., Qasim, R., Rehman, N. U., et al. (2021). Detection of fake news text classification on COVID-19 using deep learning approaches. Computational and Mathematical Methods in Medicine, 2021, 1–14.

[33] Kandasamy, V., Trojovský, P., Machot, F. A., Kyamakya, K., Bacanin, N., Askar, S., \& Abouhawwash, M. (2021). Sentimental analysis of COVID-19 related messages in social networks by involving an N-gram stacked autoencoder integrated in an ensemble learning scheme. Sensors, 21(22), 7582.

[34] Roque, G., Cavalcanti, A., Nascimento, J., Souza, R., \& Queiroz, S. (2021, October). BotCovid: Development and evaluation of a chatbot to combat misinformation about COVID-19 in Brazil. In 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC) (pp. 2506-2511). IEEE.

[35] Kastrati, Z., Ahmedi, L., Kurti, A., Kadriu, F., Murtezaj, D., & Gashi, F. (2021). A deep learning sentiment analyser for social media comments in low-resource languages. Electronics, 10(10), 1133.

[36] Tekle, E. (2022). Sentiment Analysis on Amharic Language-Based COVID-19 Discourse from Facebook social media comments (Doctoral dissertation, St. Mary's University)

[37] Singh, C., Imam, T., Wibowo, S., & Grandhi, S. (2022). A deep learning approach for sentiment analysis of COVID-19 reviews. Applied Sciences, 12(8), 3709.

[38] Khan, R., Rustam, F., Kanwal, K., Mehmood, A., \& Choi, G. S. (2021, April). US Based COVID-19 tweets sentiment analysis using textblob and supervised machine learning algorithms. In 2021 international conference on artificial intelligence (ICAI) (pp. 1-8). IEEE.

[39] Alam, K. N., Khan, M. S., Dhruba, A. R., Khan, M. M., Al-Amri, J. F., Masud, M., & Rawashdeh, M. (2021). Deep learning-based sentiment analysis of COVID-19 vaccination responses from Twitter data. Computational and Mathematical Methods in Medicine, 2021.

[40] Imran, A. S., Daudpota, S. M., Kastrati, Z., & Batra, R. (2020). Cross-cultural polarity and emotion detection using sentiment analysis and deep learning on COVID-19 related tweets. Ieee Access, 8, 181074-181090.

[41] Qorib, M., Oladunni, T., Denis, M., Ososanya, E., \& Cotae, P. (2023). Covid-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-19 vaccination Twitter dataset. Expert Systems with Applications, 212, 118715.

[42] Mansoor, M., Gurumurthy, K., \& Prasad, V. R. (2020). Global sentiment analysis of COVID-19 tweets over time. arXiv preprint arXiv:2010.14234.

[43] Kolluri, N. L., \& Murthy, D. (2021). CoVerifi: A COVID-19 news verification system. Online Social Networks and Media, 22, 100123.

[44] Samuel, J., Ali, G. M. N., Rahman, M. M., Esawi, E., \& Samuel, Y. (2020). Covid-19 public sentiment insights and machine learning for tweets classification. Information, 11(6), 314.

[45] Tripathi, M. (2021). Sentiment analysis of nepali covid19 tweets using nb svm and lstm. Journal of Artificial Intelligence, 3(03), 151-168.

[46] Gunson, N., Sieińska, W., Yu, Y., Garcia, D. H., Part, J. L., Dondrup, C., \& Lemon, O. (2021, September). Coronabot: A conversational ai system for tackling misinformation. In Proceedings of the Conference on Information Technology for Social Good (pp. 265-270).

[47] Shang, L., Kou, Z., Zhang, Y., \& Wang, D. (2021, December). A multimodal misinformation detector for covid-19 short videos on tiktok. In 2021 IEEE international conference on big data (big data) (pp. 899-908). IEEE.

[48] Pandey, R., Gautam, V., Pal, R., Bandhey, H., Dhingra, L. S., Misra, V., ... & Sethi, T. (2022). A machine learning application for raising wash awareness in the times of covid-19 pandemic. Scientific Reports, 12(1), 810.

[49] Khasnis, N. S., Sen, S., & Khasnis, S. S. (2021, August). A machine learning approach for sentiment analysis to nurture mental health amidst COVID-19. In Proceedings of the international conference on data science, machine learning and artificial intelligence (pp. 284-289).

[50] J. Naeem, O. M. Gul, I. B. Parlak, K. Karpouzis, Y. B. Salman, and S. N. Kadry, "Detection of Misinformation Related to Pandemic Diseases using Machine Learning Techniques in Social Media Platforms", 7th EAI International Conference on Robotics and Networks (EAI ROSENET 2023), Istanbul, Türkiye, pp. 1-12, 15-16 December 2023.