

Sentiment Analysis for COVID-19 Vaccine Popularity

Muhammad Saeed¹, Naeem Ahmed¹, Abid Mehmood^{2*}, Muhammad Aftab¹, Rashid Amin³
and Shahid Kamal⁴

¹ University of Engineering and Technology Taxila, Pakistan

² Department of Management Information Systems, College of Business Administration,
King Faisal University, Al-Ahsa 31982, Saudi Arabia
[e-mail: aafzal@kfu.edu.sa]

³ University of Chakwal, Pakistan

Self-employed computer science researcher, D.I. Khan, KP, Pakistan

*Corresponding author: Abid Mehmood

*Received January 14, 2023; revised March 28, 2023; accepted May 15, 2023;
published May 31, 2023*

Abstract

Social media is used for various purposes including entertainment, communication, information search, and voicing their thoughts and concerns about a service, product, or issue. The social media data can be used for information mining and getting insights from it. The World Health Organization has listed COVID-19 as a global epidemic since 2020. People from every aspect of life as well as the entire health system have been severely impacted by this pandemic. Even now, after almost three years of the pandemic declaration, the fear caused by the COVID-19 virus leading to higher depression, stress, and anxiety levels has not been fully overcome. This has also triggered numerous kinds of discussions covering various aspects of the pandemic on the social media platforms. Among these aspects is the part focused on vaccines developed by different countries, their features and the advantages and disadvantages associated with each vaccine. Social media users often share their thoughts about vaccinations and vaccines. This data can be used to determine the popularity levels of vaccines, which can provide the producers with some insight for future decision making about their product. In this article, we used Twitter data for the vaccine popularity detection. We gathered data by scraping tweets about various vaccines from different countries. After that, various machine learning and deep learning models, i.e., naive bayes, decision tree, support vector machines, k-nearest neighbor, and deep neural network are used for sentiment analysis to determine the popularity of each vaccine. The results of experiments show that the proposed deep neural network model outperforms the other models by achieving 97.87% accuracy.

Keywords: COVID-19, vaccines, social media, sentiment analysis, machine learning

This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [Grant No. GRANT2578]

1. Introduction

Recently, the world has seen significant advancements in technology, which play an essential role in industrialized countries. Modern technical tools are now used in all parts of daily life, including academia, trade, commercial, military, networking, architecture, and medical services. From recognizing symptoms to exact diagnosis and automated patient triage, the medical insurance system is a key sector that must significantly rely on current technology [1]. In December 2019, Wuhan, China, reported the first symptom of the new coronavirus COVID-19. SARS-CoV-2 continued to infect people all over the world, prompting the World Health Organization (WHO) to declare the virus a global pandemic as the number of affected persons continues to rise [2].

The virus posed a global threat, and the World Health Organization designated it COVID-19 on February 11, 2020. (Wu 2020). Some businesses have developed a number of medication combinations that are usually formed up of ethanol, isopropyl alcohols, and hydrogen peroxides in various configurations that demonstrate a strong reaction to the new pathogen and have been proven and accepted by WHO for usage worldwide [3].

COVID-19 was proclaimed a pandemic on March 11, 2020. Attempts to create a vaccine against SARS-CoV-2 were launched, with the first viral genomic sequence becoming accessible in early January. However, when the epidemic became more widely known, there was a rush to produce vaccinations in various countries [4]. As a result of the fast global spread of SARS-CoV-2 infection and the rising death toll, the development of an effective vaccine became a top priority. As a consequence, substantial breakthroughs in this research sector were made in a shorter period than predicted, and many vaccine candidates are presently in phase II/III in China, the United Kingdom, the United States, Russia, Brazil, and other countries [5]. The outbreak has put a tremendous strain on practically all countries, particularly those with weak health systems and slow reaction times. An international health catastrophe has been brought on by the COVID-19 pandemic of the coronavirus sickness. The virus is lethal for two reasons: it is novel and has yet to be discovered as a vaccine, and it is easily transferred through direct or indirect contact with the infected person. Around 1,179,035 people globally have died from the disease and 44,748,380 persons have been affected. Most COVID-19 victims come from the US. Brazil, India, Russia, South Africa, and a long list of other countries total 215 countries. Fever, dry cough, vomiting, diarrhea, and myalgia are among the symptoms that have been recognized and listed as indicators of this infection.

India's Delta wave lasted barely 90 days and faded quickly, subsequently topping in mid-May. Cases have been low since June 2021. Low vaccination rates at the time 4.2% of the population was fully immunized at the end of June 2021, led to severe viral loads. 18 months ago, the most recent COVID-19 pandemic cycle in Europe started. Million confirmed cases have been reported in Europe overall. Since January 2021, there have been fewer than 80,000 new cases reported per week in the United States. China's quasi-safe status for domestic COVID-19 infections ended in May. Since then, the number of new confirmed cases has climbed to 758.3 million in 2023. During the week concluding May 2021, there were roughly 2500 new cases recorded in China. On November 9, 2021, the 'Omicron' variety was first found in South Africa. A great amount of concern was raised by preliminary findings due to the abnormally high number of mutations. It might increase its contagiousness and cause immune evasion, which would increase the risk of reinfection. The 32 changed locations in the spike proteins of the Omicron version are associated with the virus's ability to spread and evade immune defense. Researchers warn its immune escape ability may exceed all prior SARS-CoV-2 variants,

including the prevalent Delta variant. Vaccines have long been used in public health to help prevent illnesses such as mumps, polio, rubella, and yellow fever. However, we still don't have a good understanding of their long-term viability. Several concerns have been expressed about the new vaccines being researched for SARS-CoV-2, particularly about their effectiveness and side effects [6].

As various countries are manufacturing the COVID-19 vaccine, it would be interesting to find which vaccine is best according to the people around the world. The aim of this paper is to extract the social media data about the popularity of the COVID-19 vaccine. The contribution of the proposed work are as follows:

1. Development of various machine learning models for the popularity prediction of various covid-19 Vaccines.
2. An analysis of the popularity of COVID-19 vaccines using Twitter data.
3. Investigation of public perception of different COVID-19 vaccines based on Twitter data.
4. The comparison of different machine learning models on popularity predictions of vaccines.

The Data is extracted from twitter and preprocessed for converting it into a standard form. After that machine learning techniques are performed on the data in order to find the most popular Covid-19 vaccine among people [7, 8]. The COVID-19-related deaths that have reduced in the United Kingdom and Wales are illustrated in Fig. 1.

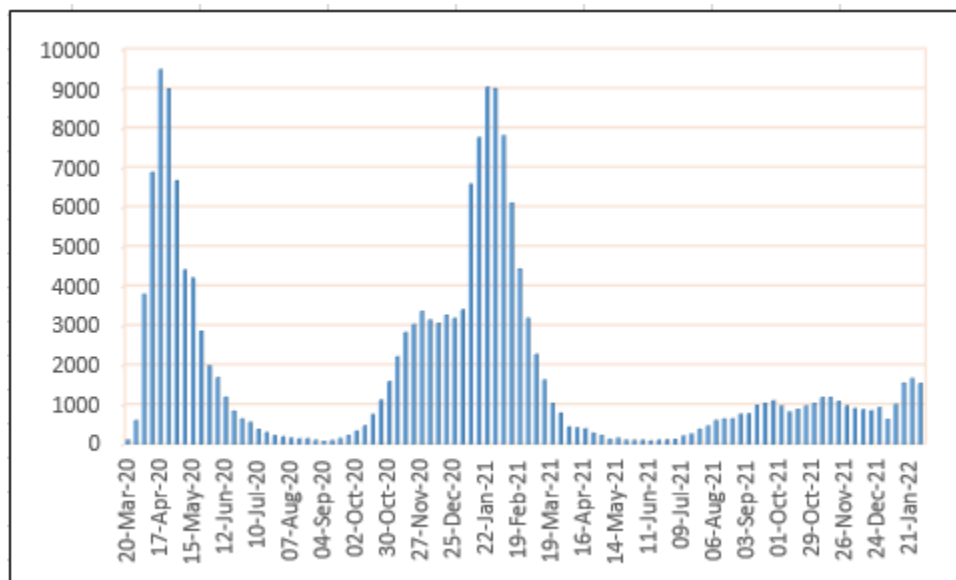


Fig. 1. Reduction of Covid-19 deaths in UK and Whales

Scientists have made various vaccines for COVID-19, all have their own advantages and limitations. Several countries attempted to develop their own corona virus vaccine for according to their needs. Vaccines in all countries achieved the same goals, such as providing immunity to the virus and attempting to stop the spread of the corona virus. There are four different vaccines available for COVID-19 [8]: (i) whole virus, (ii) protein subunit, (iii) viral vector, and (iv) nucleic acid (RNA and DNA). Several vaccines are the substance (antigen) from the body which is used for the immune system and other vaccine are used the body cells for making viral antigen [8].

A. Whole Virus

Whole virus vaccines are used to trigger an immune system in different conventional vaccines. Two main approaches have been used in the whole virus vaccine. The first approach is the live attenuated vaccine. It is used against the weaker corona virus because it can replicate by itself. The second approach is the inactivated vaccine. It is used against those viruses which destroy their genetic material [9]. So, they cannot replicate themselves. Both procedures also use technical and government approval processes, although live attenuated poses a danger of sickness in those with weakened immune systems. It is very difficult for low-resource countries because they do not have resources for cold storage and their careful use. The second vaccine, inactivated virus vaccine, also needs cold storage, and this vaccine is also given to affected immune system people. The inactivated virus vaccine cannot affect the cells [10].

B. Protein subunit

It works by triggering the immune response using infection bits and protein fragments. Although the protein subunit vaccination has reduced the danger of side effects, it has also resulted in a lesser immune system. Adjuvants are required to increase the immunological response [11]. A subunit vaccination, such as the hepatitis B and acellular pertussis vaccine, is an example. The Protein subunit vaccine has several types. It uses specific isolated proteins for bacterial and viral diseases. The Polysaccharide vaccine has a chain of sugar molecules found in the cell wall of some bacteria. The Conjugate vaccine uses two things for boosting the immune response [12]. The first one is carrier protein, and the second is polysaccharide chain. The protein subunit is the only vaccine that is against the virus which is caused by the COVID-19 [13].

C. Viral vector

Vaccines made with viral vectors also work by genetically instructing cells to make proteins. It differs from copy DNA (cdna) vaccinations because it delivers such information to the body via a disease that is not the vaccine's target. Another type of disease that is used as a vector is viral, which brings the cold virus. To induce an immune reaction, our own biological machinery is hacked, exactly just as with nucleic acid vaccines, to build the antibody from those instructions [14]. Immunization depending on viral vectors can mimic natural viral infection, eliciting a strong immune response. However, several humans may be immune to the vaccine because they've been exposed to the viruses used as vectors, making it less effective [15]. Viruses infect the cells of their hosts and take control of the protein-making machinery, which scans the pathogen's genetic information and makes new viruses. These virus particles include antigens, which are molecules that might trigger an immune response [15].

D. Nucleic acid (RNA and DNA)

Genomic DNA vaccines utilize genetic data from an infectious disease (a pathogen) to stimulate an immunological reaction against it. Based on the immunization, the DNA sample in the vaccine may be either DNA or RNA; in either case, it includes directions for creating a particular viral protein which the antibodies will recognize as alien (an antigen) [16]. The human body's normal protein-making system reads this genetic link when it happens in the

host cell body and utilizes it to produce antigens, which triggers an immune reaction. Although DNA and RNA immunizations are being researched for a number of diseases, including Hepatitis c, Dengue, and COVID-19, none have ever been licensed for humans due to the tech's youth [17]. **Table 1** shows the types of COVID-19 vaccine along with their pros and cons.

Table 1. Pros and cons of different vaccines

Platform	Vaccine	Group	Pros	Cons
Inactivated Virus	SARS-CoV vaccination that has been inactivated (ISCV)	Sinovac	Immune response is stronger; the virus is less dangerous than a live attenuated virus.	The inactivation process has the potential to change epitopes.
DNA	VRC-SRSDNA015-00-VP	NIAID	It is both risk-free and well-received. It is stable at room temperature. Flexible to adapt to novel viruses in a big way; Expression of native antigens	Humoral immunity is lower, medication is complicated, and genetic integration is a possibility.
Viral vector vaccine	MVA-MERS-S	CTC North GmbH & Co. KG	Stronger immune response; native antigen preservation; replication of natural infection	Response is slowed by pre-existing vector immunity; the production process is more difficult; and there is a possibility of being involved in the process.
Protein subunit	NVX-CoV2373	Novavax	It is completely risk free and is very well.	Pathogenicity is diminished, causing the use of an auxiliary or conjugation to increase immunogenicity.
RNA	mRNA-1273	Moderna/NIAID	Expression of original protein; secure and very well; adaptable to new viruses.	Immunogenicity is reduced; low-temperature storage and transportation is required; and the danger of an RNA-induced interferon response exists.

2. Related Work

On English-language Twitter, Liu et al. [18] outlines a sentiment study of popular sentiment toward COVID-19 immunizations. The study's major goal is to seek theme and chronological patterns in tweets for COVID-19 vaccines, as well as to investigate differences in opinions in the United States at the global, regional, and state levels. Between Nov 1, 2020, and Jan 31, 2021, the author looked at English-language tweets about COVID-19 vaccines. The Nucleon Knowledgeable Thesaurus and Attitude Arguer tools were used to evaluate whether each tweet expressed a positive (compounded 0.05), neutral (-0.05 compounded 0.05), or negative (-0.05

compounded 0.05) attitude (compounded 0.05). (0.05 compounded) (compounded-0.05). To find fundamental themes in tweets with favorable and unfavorable sentiment, we used latent semantic evaluation. Next author did a time series analysis to see if there were any trends over time, and a geographic analysis to see if there were any differences in sentiment across tweets from different areas. Positive, neutral, & negative opinions made up 42.8 percent, 26.9%, and 30.3 percent of the 2,678,372 tweets about the COVID-19 vaccination, respectively. The author detected five positive emotion tweet patterns (study results, management, living, material, and power), as well as five negative emotion tweet patterns (test results, management, social lives, details, and potency) (test results, narrative, confidence, efficiency, and management). Public opinion on COVID-19 immunizations has shifted dramatically over time and across geographic borders. Attitude analysis can help public health policymakers build regionally targeted vaccine education initiatives by providing immediate information into world opinion regarding the COVID-19 vaccine [19].

COVID-19 Immunization Twitter Sentiment Analysis and Opinion Mining was presented by Jahanbin et al. [20]. The following keywords were used in the search: vaccine, vaccination, Pfizer corona vaccine, Moderna, AstraZeneca, coronavirus pandemic, coronavirus outbreak, Epidemic, and COVID-19. Only or in combination, A total of 1,127,127 tweets for the Covid-19 immunization were gathered for a period of time from December 1 to December 30, 2020. When "tweets" were categorized as positive sentiment, it means that the discovery of the Covid-19 vaccines has given people a sense of relief, hope, and a restoration to life as it was before the pandemic. This population is more likely to be immunized when the vaccines are available. When "tweets" are categorized as negative, it means that people are afraid of the Covid-19 vaccine and are hesitant to get it even if it is available. The fact that "tweets" were classed as having a neutral attitude shows that people were neither excited nor concerned about the development of the Covid-19 vaccine. According to the statistics, 591053 tweets (52 percent) expressed positive feeling for the Covid-19 immunization, 382431 (34 percent) expressed neutral sentiment, and 152653 (14 percent) expressed unfavorable reaction.

DeVerna et al. [21] proposed the COVID-19 Vaccines: An assortment of tweets in the English language. In this paper, CoVaxxy dataset was utilized for analysis. This dataset contains tweets in English about to the COVID-19 vaccines. Using data of one week, the researcher offers details on the number of tweets sent out over time, the hashtags used, and the shared URLs. The author also shows how these data may be used by analyzing the predominance of high- and low-credibility sources, hashtag topic groups, and geographical distributions across time. The author created and displayed the CoVaxxy dashboard, which allows users to see the association between COVID-19 vaccination demand and location data posts in the United States in our dataset. With the help of our dashboard, this dataset may be used to explore how online information affects COVID-19 health outcomes (such vaccination uptake).

Bonnevie et al. [22] gave a presentation on how she used Twitter to track the surge in anti-vaccine sentiment during the COVID-19 vaccinations. A Twitter search gathered public tweets about vaccination criticism, which were then categorized into themes. A comparison was made between conversations 120 days before COVID-19 spread over the country (Oct-2019 to Feb 2020) and 120 days from (2/15/2020 to 6/14/2020). In the past year, anti-vaccine sentiment on Twitter has increased by 80%. The twelve communication themes included COVID-19, federal health agencies, vaccine elements, and research/clinical trials, with percentages of discourse about COVID-19, federal health agencies, and vaccine components increasing over time. This article tracked the surge in vaccination resistance on Twitter, claiming that vaccine opponents are fostering distrust in health officials by inciting anti-COVID-19 attitude. Vaccine opinions run the gamut, with proponents and opponents on opposite ends of the spectrum [23].

Lyu et al. [24] presented Topic Formulation and Sentiment Classification for the COVID-19 Vaccine Conversation on Twitter. In order to understand how people's opinions, considerations, and feelings may affect vaccine goals, this paper aims to distinguish between the major changes in aspects and attitudes over time and the sentiments expressed in community COVID-19 vaccine-related social networking site conversations. According to the study, the most common emotion is trust, which is followed by eagerness, worry, sorrow, and so on. The trust emotion peaked on November 9, 2020, when Pfizer said its treatment is 90 percent effective. During the discussion, a global perspective was also offered. COVID-19 medicines are expected to be more widely accepted than earlier immunizations, based on increased favourable attitude about the vaccines and the prevailing feeling of confidence displayed in online networking conversations.

Marcec et al. [25] Presented a sentiment study of AstraZeneca/Oxford, Pfizer/BioNTech, and Moderna COVID-19 vaccines using Twitter. From December 1, 2020, to March 31, 2021, the Twitter academic Application Programming Interface was utilized to obtain all English-language tweets mentioning AstraZeneca/ Oxford, Pfizer/BioNTech, and Moderna vaccines. The daily average mood of tweets was calculated over a period of four months using the AFINN vocabulary, and was contrasted for each vaccine after longitudinal analysis. Public opinion of the AstraZeneca/Oxford vaccine appears to be worsening, with a substantial drop in the months of December and March ($p=0.0000000001$, mean difference=0.746, 95 percent CI=0.915 to 0.577). Lexicon-based Tweets sentiment analysis is a useful and simple technique to track the opinion towards the SARS-CoV-2 vaccine. The public's view of the AstraZeneca/Oxford vaccine seems to be deteriorating over time, which is troublesome because it might cause scepticism towards this particular SARS-CoV-2 vaccine to grow [27].

3. Proposed Work

The design of the suggested scheme is depicted in Fig. 2. The following groups can be used to categorize the system's flow: To begin, the system retrieves COVID-19 vaccine related data using Twitter API, which includes text reviews for the specific Vaccine, company, Location, and date. The second stage of the system preprocesses the data and applies natural language processing techniques for sentence sentiment analysis. The final phase generates results from given data and stores them in a database for future use.

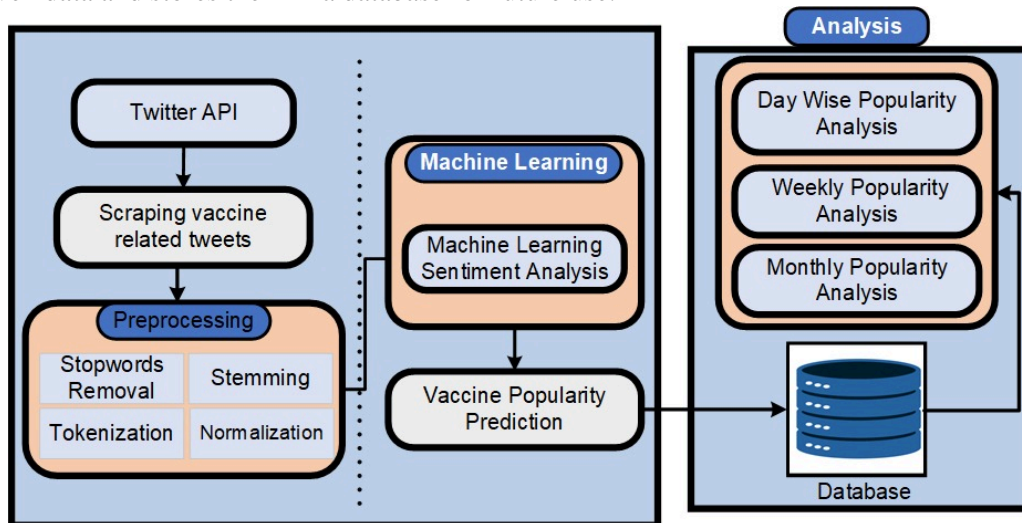


Fig. 2. Proposed model architecture

3.1. Data Collection

We gathered data by using Twitter API (Tweepy) for scraping hash tagged tweets from the Twitter platform. The gathered dataset containing 80k hash tagged tweets about COVID-19 vaccines. After that, we removed duplicate tweets, non-English tweets, and tweets without hashtags from the hash tagged dataset. We look at the distribution of hashtags in the remaining collection and aim to find sets of hashtags that are indicative of positive, negative, and neutral comments. These hashtags are used to filter tweets that are utilized for training and building machine learning models. Various locations/countries and specific hashtags, including "#COVID-19

#Vaccination", "#Vaccinated#Sinopharm", "#Sinovac", "#covid_vaccine", "#Moderna", "#Pfizer", "Moderna" and "Covid_vaccine" were used to filter the tweets. According to the number of tweets linked to the topic, many countries have been targeted for gathering data. The data of countries that have more ratio of tweets in the gathered dataset is illustrated in Fig. 3.

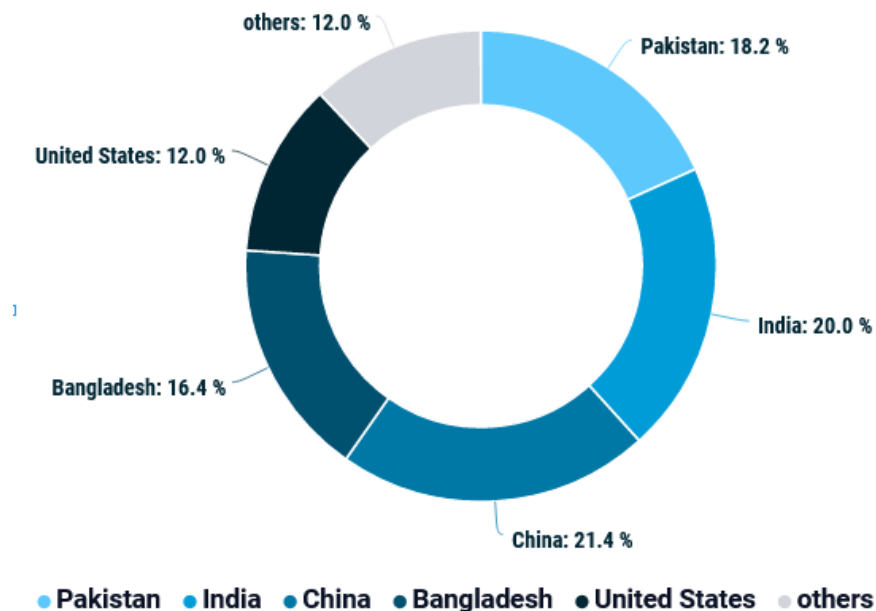


Fig. 3. Ratios of the collected data set

The posted reviews, user, location, and timestamp of the tweets are all included in the data. We preprocessed the dataset to remove various repetitive words, irrelevant information, stop words, and special characters to make it clean and appropriate for building the proposed models. This data set was collected from the Twitter platform over a period of time, ranging from September 2020 to December 2021. All tweets are annotated manually and stored in a .csv format.

Various researchers-built data sets for sentiment analysis of COVID vaccinations using different platforms, but a few have done work on popularity of vaccines. Reshi et al. presented a study on COVID-19 vaccination related sentiment analysis. They also scraped twitter data of 40k tweets from around the globe. Qorib et al. also published an article on the sentiment analysis for finding the COVID-19 hesitancy using twitter data. They used 40793 tweets related to COVID-19 vaccines. Griffith et al. also worked on vaccine hesitancy using COVID by sentiment analysis. They used 3915 tweets from Canadian people for conduction their

research. The comparison with some previous work of COVID vaccines data is illustrated in [Fig. 4](#).

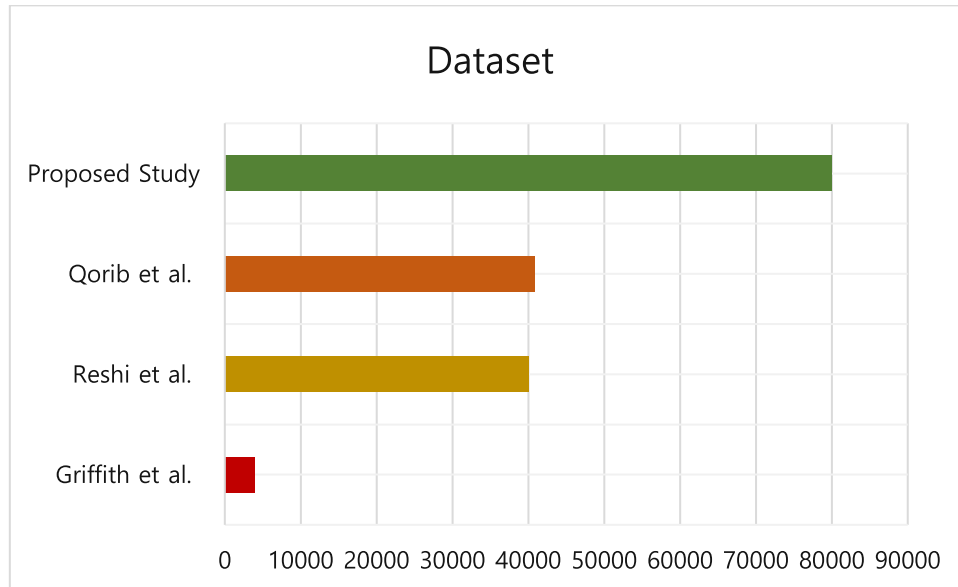


Fig. 4. Comparison of data set

2.2 Data Preprocessing

One of the most significant processes in text NLP and information retrieval is pre-processing. Data preprocessing includes various tasks like tokenization, normalization, stop words removal, Special characters' removal, etc. Data pre-processing is a technique for extracting useful and non-trivial information from unstructured data. Information retrieval is critical for determining which data in a set should be obtained in order to meet a user's information needs. We used this step for the above tasks to convert unstructured data into a structured format.

2.2.1 Tokenization

A text is tokenized when it is divided into words, signs, sentences, or other important parts. For additional processing, like text mining or parsing, the token collection acts as the beginning point. It disassembles sentences into their component words. Text data is initially just a string of characters. For each phase in the sentiment analysis process, the words from the data set must be used. It is the most important part of text mining, where all the terms from phrases are collected and their frequency counts are done. With the help of Count Vectorizer, we detect how often a single word appears in a tweet. Each phrase is given a unique number when tokens are created. That token comprises unique feature values that are utilized to create feature vectors.

2.2.2 Stop words Removal

Stop word removal is used in the preprocessing phase in various NLP tasks. The concept is to remove words that appear commonly in all of the documents in the dataset. After the Data has been converted into distinct tokens, the next step is to eliminate all terms that have no significance. For example, white spaces, brackets, punctuation marks, colons, full stops, and so on. We removed the stop words from the data by using NLTK module in python language.

2.2.3 Stemming

Stemming is the next step after token generation and stop words removal. Stemming is the process of returning the words that were derived from the data to their initial shapes. The prefixes and suffixes are removed from the basic terms. The stemming algorithm converts all altered or misspelt words into their matching basic or stem words. The data is stemmed using the NLTK standard Python module.

2.3.4 Feature Extraction

The feature extraction is a method of transforming the data in a format that machine learning models can work with. Depending on the base data, any attribute, feature, or class can be mined within this process. Feature extraction is an important stage in model training that helps the algorithm produce more trustworthy and precise outputs. Feature selection refers to choosing some key variables from data that adequately characterize data during the feature extraction stage. After the selection of important features, the machine learning models are trained and evaluated. The feature selection plays a vital role in the faster model training. Algorithm 1 provides a list of steps involved in the proposed methodology.

Algorithm 1: Using Twitter data for vaccine popularity detection

```

Input  $\leftarrow$  Dataset(Training "X"), Test_data (Y)
Let  $X=(features, Labels)$ 
Let  $X'=Preprocess(X[text])$ 

Tokenize( $X[text]$ )
Stop_Word_Removal( $X[text]$ )
Stemming( $X'[text]$ )

 $\beta= Feature\_Extract(X'[text])$ 
Data_splitting= $X\_train, X\_test, train\_y, test\_y \leftarrow train\_test\_split()$ 
Procedure Models:
Model  $\leftarrow$  (SVM, NB,DT,KNN,DNN) #initialize the model
Compile $\leftarrow Model.Compile()$  # SVM, NB,DT,KNN,DNN
Model= $Model.fit(\beta)$  # SVM, NB,DT,KNN,DNN
Model.evaluate(Test_data)
Model.predict(Tweets)

```

2.3 Sentiment Analysis

A common use of natural language processing (NLP) is sentiment analysis. We can use sentence sentiment analysis to examine comments on social media to determine how much a vaccine is popular in individuals on twitter platform. For this purpose, we can use various machine learning models. We used 4 machine learning models, i.e., naïve bayes, decision tree, kNN and support vector machine (SVM) along with deep neural network (DNN). The process of sentiment analysis for popularity detection is shown in [Fig. 5](#).

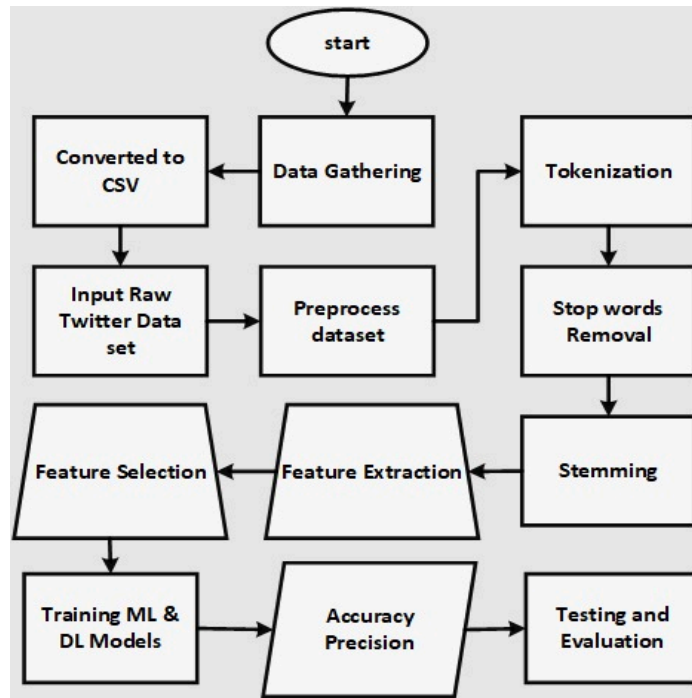


Fig. 5. Flow chart of proposed work

Naïve bayes is a supervised ML algorithm that can be used for various purposes. It is focused mostly on the ability to distinguish between various things based on predetermined criteria. The Naive Bayes algorithm detects a phrase or event that occurred previously and predicts the chance of it happening again in the future. We used naive bayes model with default parameters on training data. SVM is another supervised model that can be used for classification and regression. It performs very well in the case of small datasets. This algorithm uses hyperplanes for dealing with the data and producing its output. We use the following values of SVM parameters while taking other with default values. Table 2 presents the parameters and their values for SVM model.

Table 2. SVM parameters and their values

Parameter	Value
Kernel	Linear
C	1.0
Degree	2
Gamma	Auto

The supervised machine learning method decision tree is employed for both categorization and regression. A decision tree model based on the training data was developed using the Gini criterion and the best splitter approach. We utilized RFE to minimize the number of training features because the decision tree is prone to over-fitting on a high number of features. K-Nearest Neighbors (KNN) is a non-parametric classification and regression algorithm. In a KNN model, there is only one parameter: the number of nearest neighbors (K) that are used to make the prediction. The value of K determines how sensitive the model is to the input data. A smaller value of K means that the model is more sensitive to the input data and may overfit the training data, while a larger value of K means that the model is less sensitive to the input data and may underfit the training data. We used $k=15$ for our dataset while implementing KNN model.

The proposed models are trained on 70% of total data while 30% data is used for the evaluation phase. The trained models and their data are stored in a database that it can be used for future tasks. Also, the outputs from the models are stored so that we can generate the reports of the outputs of models on user data. All models are developed using Python language and after successful training phase models are evaluated using standard metrics, i.e., accuracy and precision. For the popularity detection, we built three options. The first option allows us to check the daily popularity of a specific vaccine, while the other option can be used for the weekly popularity detection. The last one is used for the monthly detection of the popularity of COVID-19 vaccines.

3. Results and Analysis

The proposed algorithms predict whether a COVID-19 vaccine will become popular. The result is computed using different evaluation measure, i.e., accuracy and precision. All proposed machine learning models perform well in the training and evaluation phase. However, the DNN model outperforms other models in term of accuracy and other evaluation metrics on the proposed dataset. Accuracy measures how often the model makes correct predictions, which is defined as the number of correct predictions divided by the total number of predictions made. It is a measure of how well the model is able to classify instances into the correct class.

Precision, on the other hand, measures how often the model correctly predicts a positive instance (i.e., true positive), out of all the instances that it predicted as positive. It is calculated as the number of true positives divided by the total number of predicted positives. Precision is a measure of the model's ability to avoid false positives. The training and validation accuracy proposed DNN model is shown in [Fig. 6](#). As illustrated in [Fig. 6](#), the proposed DNN model achieves 99% accuracy in the training phase while 97.12% in the validation phase.

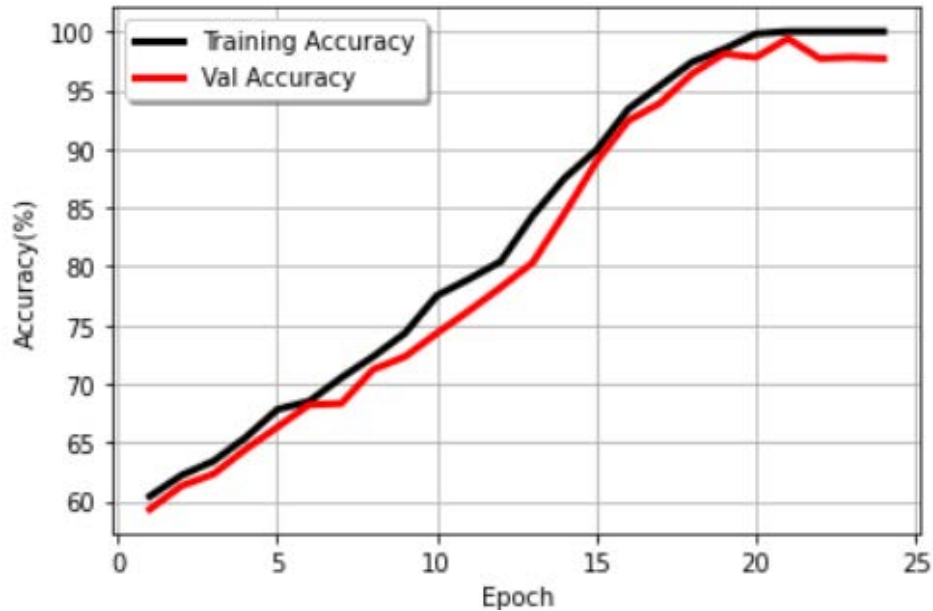


Fig. 6. Training and validation accuracies of DNN model

The proposed model achieved an optimal level of loss value in training and validation phases of the study. The proposed model was trained for 25 epochs to achieve a reliable accuracy rate. The **Fig. 7** present the training and validation loss for the proposed model.

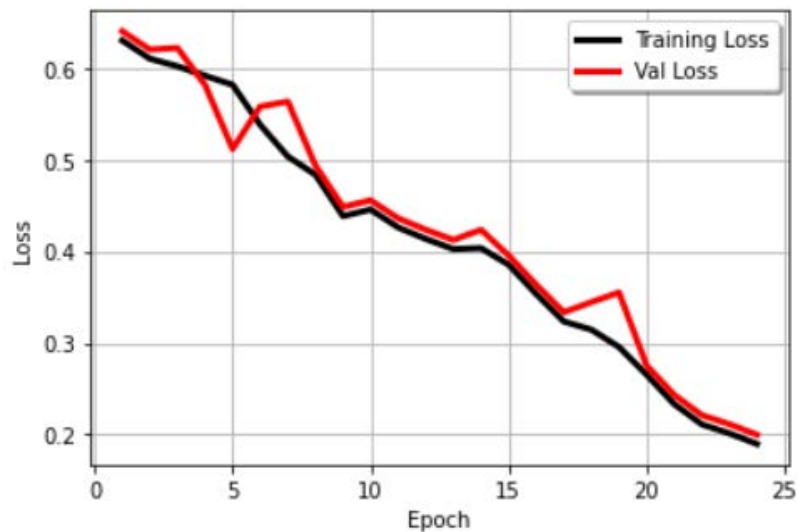


Fig. 7. Training and validation loss of DNN model

To determine which approach is the most effective, we examined the proposed methods using different evaluation metrics. The comparison of the offered methodologies' accuracy is shown in **Fig. 8**.

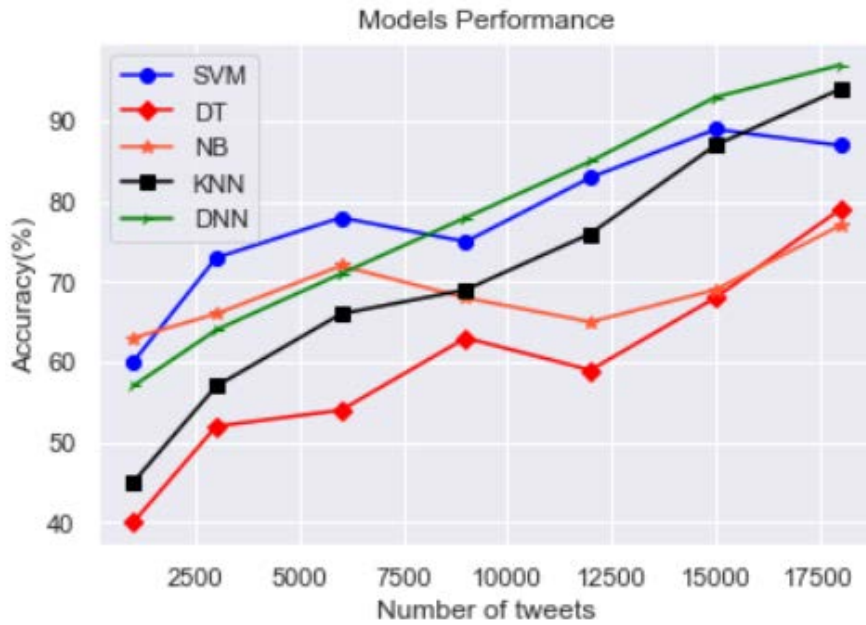


Fig. 8. Accuracy of proposed models

One measure of the learning model's effectiveness is precision, which is calculated as the ratio of true positives to all predictions made by the model. The precision of the proposed models was calculated and analyzed. The proposed DNN model also outperformed other models in term of precision. Fig. 9 presents the comparison of precision of all proposed models.

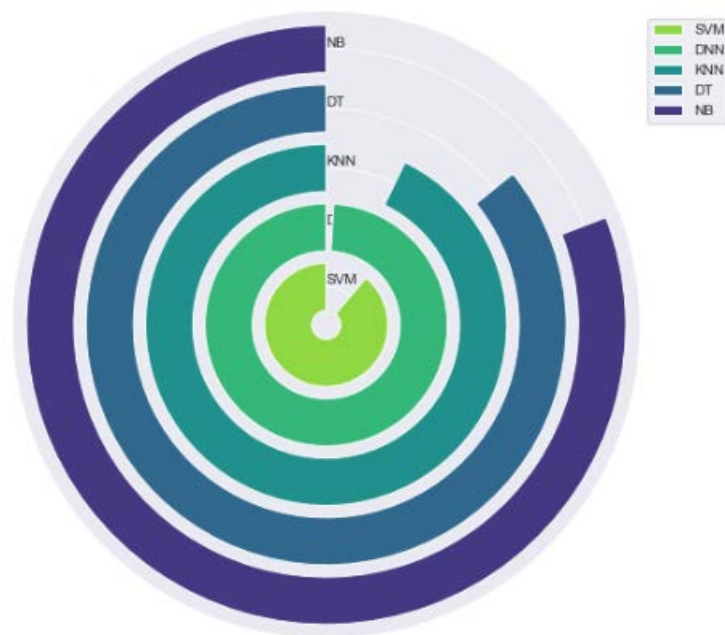


Fig. 9. Precision values of proposed models

Besides accuracy and precision, some other evaluation metrics are also used for finding the best model in term to popularity detection of COVID vaccines data. These metrics include F1-score and recall. The recall is calculated as the percentage of all positive instances that were properly classified as positive. The recall calculates how well the algorithm can distinguish between positive data instances. The harmonic mean of recall and precision is known as the F1-score. Equation (1) combines recall and precision into a single value to calculate F1-score:

$$2 * \frac{Precision * Recall}{Precision + Recall} \quad (1)$$

The comparison of models on all evaluation metrics is presented in [Fig. 10](#).

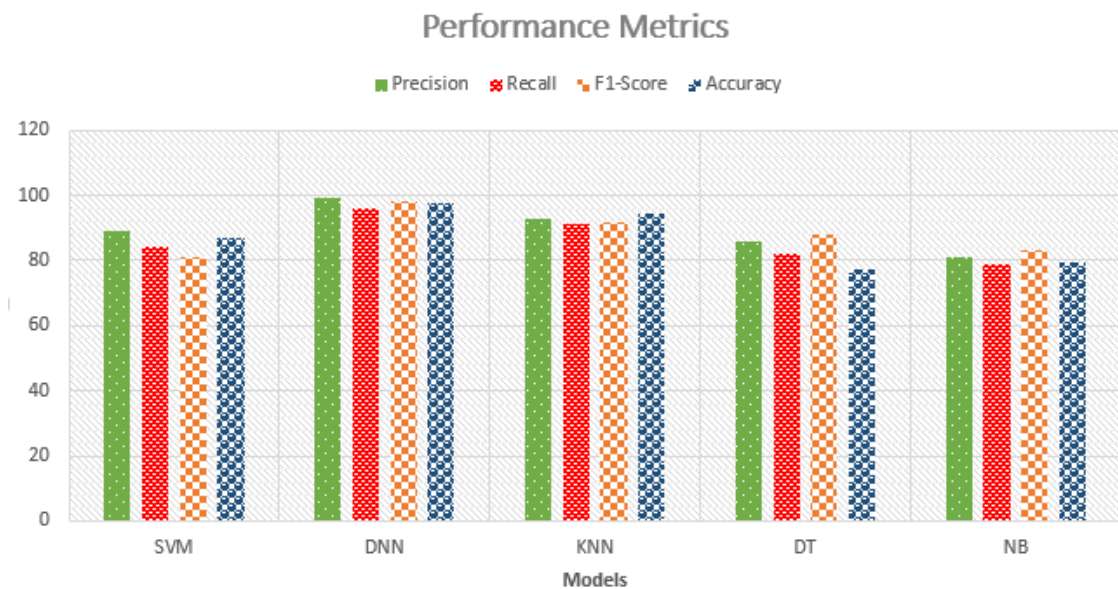


Fig. 10. Performance of proposed models

As shown in [Fig. 10](#) the DNN model outperform other models by achieving 97.87% accuracy while KNN achieved second highest accuracy of 83.23%. The decision tree model performs very bad on the proposed data set by achieving just 76.54% accuracy. [Fig. 10](#) shows the comparison of proposed models for vaccine popularity. The lowest accuracy is gained by the decision tree model as it is very sensitive to large number of features. The best model for the proposed problem was DNN as it outperformed other models. As the proposed models achieved handsome level of accuracy and other precision values so we can use the model for the prediction of popularity of vaccines on new data gathered from twitter or any other social media platform. By using the popularity prediction of vaccines from proposed models the producers of vaccines can take decisions about the quantity of production of their vaccines. The output of the DNN model for various COVID vaccines is illustrated in [Fig. 11](#).

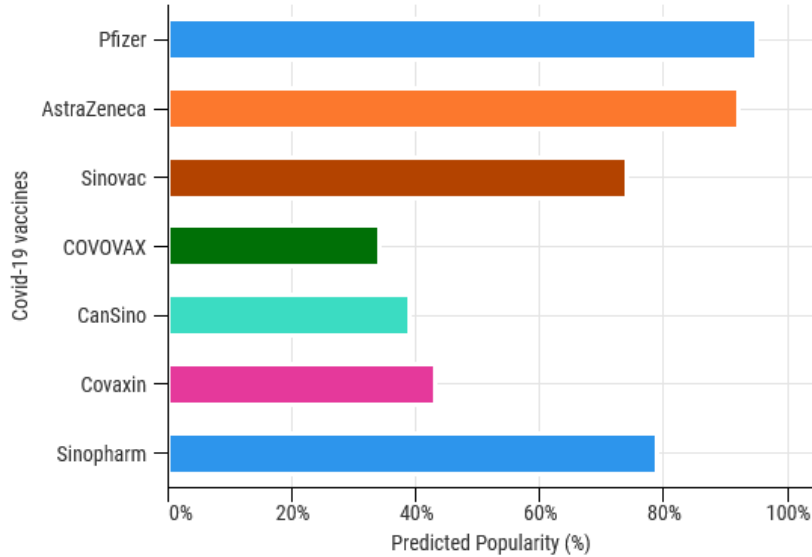


Fig. 11. Results of DNN model for vaccines

4. Comparison with Other Works

After the production of different Covid-19 vaccines, various studies have been conducted to predict the popularity of COVID-19 vaccines among different population groups using machine learning techniques. In this research paper, we compare our findings with some other studies to show the effectiveness of our proposed technique. By doing so, we aim to provide a comprehensive analysis of the effectiveness of machine learning in predicting COVID-19 vaccine popularity and its potential for informing public health policies and strategies. We compare the proposed system results with some previous related works on the basis of accuracy. The comparison of proposed work is illustrated in **Fig. 12**.

As shown in **Fig. 12**, Liu et al.[18] achieved 89% while Lyu et al.[24] got 90% accuracy of the sentiment analysis of different Covid-19 vaccines. The proposed model achieved 97% accuracy by using the proposed DNN model as discussed earlier.

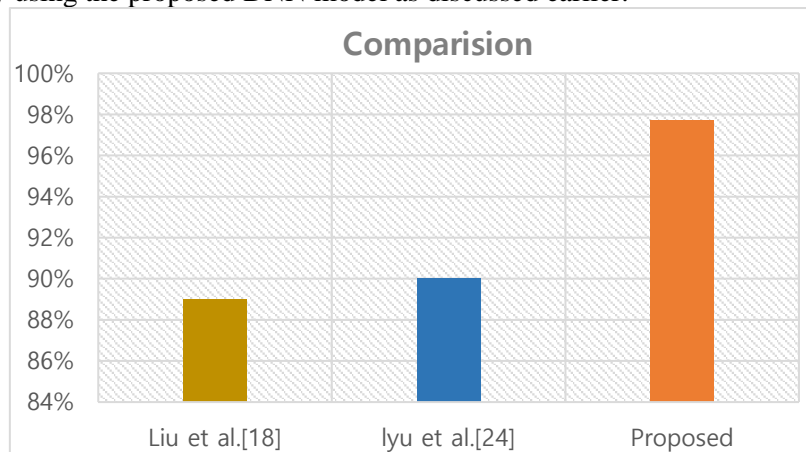


Fig. 12. Comparison with other works

5. Conclusion and Future Work

After the COVID-19 outbreak, various countries made vaccines for COVID-19 treatment. These vaccines are being used around the globe. Some vaccines are good, and people easily get vaccinated by that vaccines and some are not. In this work, we used machine learning techniques for the development a sentiment analysis model for COVID-19 vaccines popularity. The goal of this study was to examine people's feelings and emotions about Covid-19 vaccines during the COVID-19 pandemic. Machine learnings models were trained on twitter data that is gathered using tweepy API. After model training, all models were evaluated, and performance was compared. The proposed SVM model achieved the highest accuracy, while the decision tree model performs worst on the proposed data set. The SVM model achieved 94.87, naive bayes achieved 83.23% while the decision tree got 76.54% accuracy. The tweets used in this study were all in the English language, which could limit the scope of the research. Future research can be done using other languages' data. Moreover, some hybrid and Transfer learning-based models can be trained on a large dataset for improving the results.

References

- [1] Khamsi, R.J.N., "If a coronavirus vaccine arrives, can the world make enough," *Nature*, 580(7805), pp. 578-580, 2020. [Article \(CrossRef Link\)](#)
- [2] Koff, W.C. and S.F.J.S. Berkley, "A universal coronavirus vaccine," *American Association for the Advancement of Science*, vol. 371, pp. 759-759, 2021. [Article \(CrossRef Link\)](#)
- [3] Li, Y.-D., et al., "Coronavirus vaccine development: from SARS and MERS to COVID-19," *Journal of Biomedical Science*, 27(1), pp. 1-23, 2020. [Article \(CrossRef Link\)](#)
- [4] Nir Eyal, Marc Lipsitch, Peter G Smith, "Human challenge studies to accelerate coronavirus vaccine licensure," *The Journal of Infectious Diseases*, 221(11), pp. 1752-1756, 2020. [Article \(CrossRef Link\)](#)
- [5] Ella, K.M. and V.K.J.I.pp. Mohan, "Coronavirus vaccine: light at the end of the tunnel," *Indian Pediatrics*, 57(5), pp. 407-410, 2020. [Article \(CrossRef Link\)](#)
- [6] Dyani Lewis, "China's coronavirus vaccine shows military's growing role in medical research," *Nature*, 585(7826), pp. 494-495, 2020. [Article \(CrossRef Link\)](#)
- [7] Elgendy, M.O. and M.E.J.J.o.M.V. Abdelrahim, "Public awareness about coronavirus vaccine, vaccine acceptance, and hesitancy," *Journal of Medical virology*, 93(12), pp. 6535-6543, 2021. [Article \(CrossRef Link\)](#)
- [8] Freeman, D., et al., "Effects of different types of written vaccination information on COVID-19 vaccine hesitancy in the UK (OCEANS-III), a single-blind, parallel-group, randomised controlled trial," *the lancet public health*, vol. 6(6), pp. e416-e427, 2021. [Article \(CrossRef Link\)](#)
- [9] Lazarus, R., et al., "Immunogenicity and safety of inactivated whole virion Coronavirus vaccine with CpG (VLA2001) in healthy adults aged 18 to 55: a randomised phase 1/2 clinical trial," *medRxiv*, 2021. [Article \(CrossRef Link\)](#)
- [10] Medina-Pestana, J., et al., "Inactivated Whole-virus Vaccine Triggers Low Response Against SARS-CoV-2 Infection Among Renal Transplant Patients: Prospective Phase 4 Study Results," *Transplantation*, 106(4), 853-861, 2022. [Article \(CrossRef Link\)](#)
- [11] Richmond, P., et al., "Safety and immunogenicity of S-Trimer (SCB-2019), a protein subunit vaccine candidate for COVID-19 in healthy adults: a phase 1, randomised, double-blind, placebo-controlled trial," *The Lancet*, 397(10275), pp. 682-694, 2021. [Article \(CrossRef Link\)](#)
- [12] Mandolesi, M., et al., "SARS-CoV-2 protein subunit vaccination of mice and rhesus macaques elicits potent and durable neutralizing antibody responses," *Cell Reports Medicine*, 2(4), p. 100252, 2021. [Article \(CrossRef Link\)](#)

- [13] Yang, S., et al., "Safety and immunogenicity of a recombinant tandem-repeat dimeric RBD-based protein subunit vaccine (ZF2001) against COVID-19 in adults: two randomised, double-blind, placebo-controlled, phase 1 and 2 trials," *The Lancet Infectious Diseases*, 21(8), pp. 1107-1119, 2021. [Article \(CrossRef Link\)](#)
- [14] Klugar, M., et al., "Side effects of mRNA-based and viral vector-based COVID-19 vaccines among German healthcare workers," *Biology*, 10(8), p. 752, 2021. [Article \(CrossRef Link\)](#)
- [15] Crommelin, D.J., et al., "The science is there: key considerations for stabilizing viral vector-based Covid-19 vaccines," *J. Pharm. Sci.*, 110(2), pp. 627-634, 2021. [Article \(CrossRef Link\)](#)
- [16] Weixler, L., et al., "ADP-ribosylation of RNA and DNA: from in vitro characterization to in vivo function," *Nucleic Acids Research*, 49(7), pp. 3634-3650, 2021. [Article \(CrossRef Link\)](#)
- [17] Le, T.K., et al., "Nucleic acid-based technologies targeting coronaviruses," *Trends in Biochemical Sciences*, 46(5), pp. 351-365, 2021. [Article \(CrossRef Link\)](#)
- [18] Liu, S. and J.J.V. Liu, "Public attitudes toward COVID-19 vaccines on English-language Twitter: A sentiment analysis," *Vaccine*, 39(39), pp. 5499-5505, 2021. [Article \(CrossRef Link\)](#)
- [19] Amelia, L. and R.A.J.I.J.P.H.S. Syakurah, "Analysis of public search interest towards immune system improvement during the COVID-19 pandemic using google trends," *International Journal of Public Health Science (IJPHS)*, 9(4), pp. 414-420, 2020. [Article \(CrossRef Link\)](#)
- [20] JAHANBIN, K., et al., "Sentiment Analysis and Opinion Mining about COVID-19 vaccines of Twitter Data," 13, pp. 1-3, 2020.
- [21] DeVerna, M.R., et al., "CoVaxxy: A collection of English-language Twitter posts about COVID-19 vaccines," in *Proc. of the AAAI international conference on web and social media (ICWSM)*, 2021. [Article \(CrossRef Link\)](#)
- [22] Bonnevie, E., et al., "Quantifying the rise of vaccine opposition on Twitter during the COVID-19 pandemic," *Journal of Communication in Healthcare*, 14(1), pp. 12-19, 2021. [Article \(CrossRef Link\)](#)
- [23] Hussain, A., et al., "Artificial intelligence-enabled analysis of public attitudes on facebook and twitter toward covid-19 vaccines in the united kingdom and the united states: Observational study," *J Med Internet Res*, 23(4), p. e26627, 2021. [Article \(CrossRef Link\)](#)
- [24] E. Le Han, and G.K.J.J.o.m.I.r. Luli, "COVID-19 vaccine-related discussion on Twitter: topic modeling and sentiment analysis" *J Med Internet Res*, 23(6), p. e24435, 2021. [Article\(CrossRef Link\)](#)
- [25] R.J.P.M.J. Likic, "Using twitter for sentiment analysis towards AstraZeneca/Oxford, Pfizer/BioNTech and Moderna COVID-19 vaccines," *Postgraduate Medical Journal*, 98(1161), pp. 544-550, 2021. [Article \(CrossRef Link\)](#)
- [26] Lu, Y. and L.J.J.o.I. Zhang, "Social media WeChat infers the development trend of COVID-19," *Journal of infection*, 81(1), pp. e82-e83, 2020. [Article \(CrossRef Link\)](#)
- [27] Madanian, S., et al., "Twitter Sentiment Analysis in Covid-19 Pandemic," in *Proc. of 2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, IEEE, 2021. [Article \(CrossRef Link\)](#)



Muhammad Saeed is a computer scientist from Pakistan who has a strong academic background and expertise in a variety of computer science fields. He received the graduate degree in computer science from the University of Haripur, Haripur, Pakistan, in 2015, and the M.S. degree in computer science from the University of Engineering and Technology, Taxila, in 2021. Since January 2019, he has been working as a teacher in elementary and secondary education department KPK, Pakistan. His research interests are in Machine learning, IoMT and IOT.



Naeem Ahmed is a highly motivated computer scientist hailing from Pakistan, with a strong educational background and expertise in various fields of computer science. He received the bachelor's degree in computer science from the University of Haripur, Haripur, Pakistan, in 2019, and the M.S. degree in computer science from the University of Engineering and Technology, Taxila, in 2022. Since October 2022, he has been working as a Lecturer in Computer Science Department, at Government Post Graduate College Haripur, Pakistan. His research interests are in machine learning, information security, natural language processing, and the Internet of Things.



Abid Mehmood received the M.Sc. degree in computer science from Quaid-i-Azam University, Islamabad, Pakistan, in 2001, and the Ph.D. degree in computer science from Universiti Teknologi Malaysia, Johor Bahru, Malaysia, in 2014. Prior to entering academia, from 2001 to 2010, he worked at the software development industry in different roles and contributed to the design and development of various high-performance enterprise applications. He is currently an Assistant Professor with the Department of Management Information Systems, King Faisal University, Saudi Arabia. His research interests include neural networks and deep learning, the Internet of Things, dynamic and self-adaptive systems, model-driven engineering, and aspect-orientation.



Rashid Amin received a Ph.D. degree in computer science from the Comsats University Islamabad, Wah Campus, Pakistan. He received an MS in Computer Science (MSCS) and a Master of Computer Science (MCS) from International Islamic University, Islamabad. He has been working as an Assistant Professor at the Department of Computer Science, University of Chakwal, Pakistan. Before this, he worked as a Lecturer at the Department of Computer Science, University of Engineering and Technology, Taxila, Pakistan, for seven years and at the University of Wah, Wah Cantt, Pakistan, for four years. He supervised many MS level students' theses, and Five Ph.D. students are working under his supervision. His current research interests include machine learning, deep learning, IoMT, distributed systems, and cyber security. He is co-editing some special issues in some renowned journals. He has published several research papers on ML, DL, and SDN in well-reputed venues such as IEEE Communication Surveys & Tutorial, IEEE Access, and Electronics MDPI.



Muhamad Aftab received the MCS degree in Computer Science from the University of Haripur, Pakistan, in 2015 and the MS degree in Computer Science from the University of Engineering and Technology, Taxila, Pakistan, in 2021. His research interests include the Internet of Things, natural language processing, and bioinformatics.