Detecting COVID-19 vaccine hesitancy in India: a multimodal transformer based approach

Anindita Borah[1]

Abstract

COVID-19 has emerged as the greatest threat in recent times, causing extensive mortality and morbidity in the entire world. India is among the highly affected countries suffering severe disruptions due this pandemic. To overcome the adverse effects of COVID-19, vaccination has been identified as the most effective preventive measure globally. However, a growing amount of hesitancy has been observed among the general public regarding the efficacy and possible side-effects of vaccination. Such hesitancy may proved to be the greatest hindrance towards combating this deadly pandemic. This paper introduces a multimodal deep learning method for Indian Twitter user classification, leveraging both content-based and network-based features. To explore the fundamental features of different modalities, improvisations of transformer models, BERT and GraphBERT are utilized to encode the textual and network structure information. The proposed approach thus integrates multiple data representations, utilizing the advances in both transformer based deep learning as well as multimodal learning. Experimental results demonstrates the efficacy of proposed approach over state of the art approaches. Aggregated feature representations from multiple modalities embed additional information that improves the classification results. The findings of the proposed model has been further utilized to perform a study on the dynamics of COVID-19 vaccine hesitancy in India.

Keywords COVID-19 · Social media · Social network analysis · Feature learning · Multimodal learning

1 Introduction

Initially identified in Wuhan, China, the Coronavirus Disease (COVID-19) has spread exponentially throughout the world affecting millions of lives. The World Health Organization (WHO) has declared the COVID-19 outbreak a pandemic, owing to its severity and alarming levels of spread and infectivity. India is among the worst affected countries in terms of mortality and number of cases. The nation has suffered extensive economic,

✉ Anindita Borah
  anindita01.borah@gmail.com

[1] Indian Institute of Technology Guwahati, Assam, India

political and social crisis due to this sudden outbreak. The Indian government has taken extraordinary measures to manage and control this emergency situation. Measures like social distancing and wearing of masks have been implemented along with nationwide lockdown to restrict the spread of the disease. These measures however, have caused extensive devastation to the global economy and inimically affected people's social, physical and mental health. Vaccination therefore, has been identified as the most effective preventive measure to control the spread of the virus. Considering the significance of COVID-19 vaccine, the government has started vaccination drives all over the country to achieve herd immunity among the population. With the onset of vaccination programmes in 2021, the spread of the virus is expected to get reduced considerably. Figure 1a illustrates the number of COVID-19 cases identified in the year 2021 and Fig. 1b elicits the number of doses administered in India from 16th January 2021 to 31st December 2021. The statistics have been collected from the data posted in the statistics portal of Statista Research Department.

The preventive measure of vaccination towards COVID-19 adopted by the government can be hampered by the hesitancy of people regarding the same. WHO has listed the issue of vaccine hesitancy as one of the top 10 global health challenges. The Strategic Advisory Group of Experts Working Group (SAGE) has defined vaccine hesitancy as the "refusal to get vaccinated or delay in acceptance of vaccine by the public despite its availability". Timely maintenance of vaccine hesitancy is an important issue to be addressed for effective control of severe diseases like COVID19.

Social media platforms like Twitter have been flooded with enormous content associated with Coronavirus, as the world discovers itself in the middle of this crisis. Diverse use of social media platforms expedite the process of information and opinion sharing on health crisis like COVID-19. Indian citizens explicitly use Twitter for propagating information and expressing their views related to COVID-19. However, reluctance in getting vaccinated and anti-vaccination ideas expressed by the citizens towards COVID-19 vaccine may hinder the control of the disease. Exposure to negative sentiments expressed on Twitter regarding COVID-19 vaccination may increase hesitancy among the users. Such behavior can be a threat to any economy particularly to worst COVID-19 affected countries like India. It is therefore, essential to analyze the orientation of Indian Twitter users towards COVID-19 vaccination to identify their reluctance and hesitancy. Considering the serious implications of vaccine hesitancy on highly affected country like India, this study attempts to identify the hesitancy of Indian public regarding COVID-19 vaccination through their Twitter activities.

The problem of COVID-19 vaccine hesitancy detection can be formulated under a classification setting. As a binary classification problem, it can be predicted whether a user is hesitant or non-hesitant regarding COVID-19 vaccination based on their Twitter activities. The paradigm of deep learning has received considerable attention in recent times, owing to its capability of solving complex prediction tasks. For challenging deep learning based classification problems, information conveyed by a single modality may not be sufficient for an accurate prediction. Multimodal deep learning is an emerging paradigm that attempts to combine multiple distinct features for learning common feature spaces and patterns. Incorporating the notion of multi-view learning in deep neural networks greatly enhances the generalization performance. Among the neural network models, the recently developed transformer models have been identified to be most effective for Natural Language Processing tasks. Transformer models offer several benefits over the rest of the neural network models including accuracy and faster processing.

Multimodal data captured from different modalities have been widely explored in several domains including social networks. Twitter data can have multiple views capturing
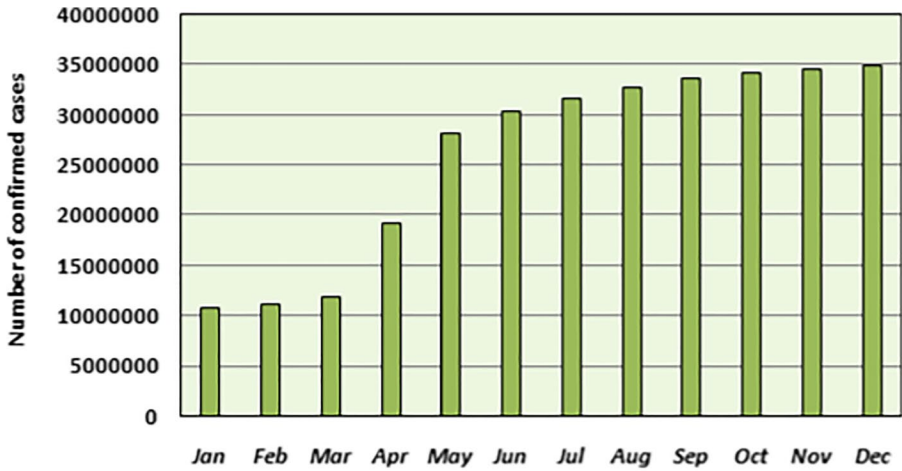
(a) Confirmed cases in 2021
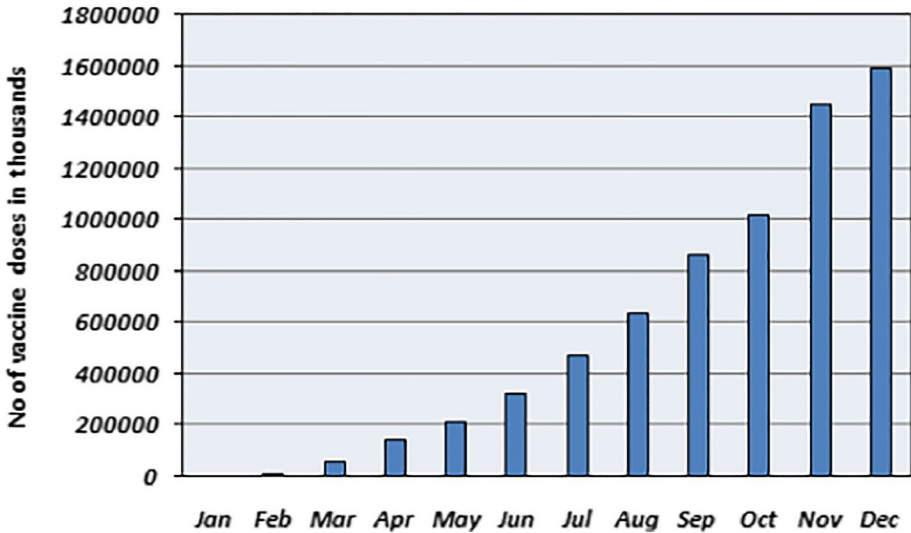
(b) Doses administered till 31st December 2021

Fig. 1 Number of confirmed cases and doses in India in 2021

different types of information regarding the users: tweet content or features extracted from tweets like hashtags, retweets and mentions. Most of the approaches consider either the textual content of the tweets or the network relationship between the users for solving any Twitter based prediction task. Network-based approaches are better than their content-based counterparts in terms of language independence. However, a prominent weakness of the network-based approaches is the inability to identify the labels for disconnected nodes. A possible solution in this regard is to incorporate the textual information along with the users graph information for predicting the labels for isolated nodes. Exploiting information

from multiple features and modalities have resulted in significant improvements in the performance of prediction models.

Keeping in view, the significance of multi-modal transformer based deep learning, this paper presents a credible approach integrating the advantages of transformer models, BERT (Devlin et al., 2018) and GraphBERT (Zhang et al., 2020) and multimodal learning. The proposed approach leverages both content-based and network based information for predicting COVID-19 vaccine hesitancy among Indian Twitter users. To the best of our knowledge, this study is the first attempt towards detection of COVID-19 vaccine hesitancy in India using a computational approach. Moreover, integrating the concepts of multi-modal transformer based learning has further enhanced the originality of the proposed method. To summarize, the prime contributions of this study are as follows:

- A generic transformer based multimodal architecture capable of exploiting both structural and textual modality is proposed.
- Comparative evaluation of the proposed multimodal approach has been performed with the state of the art approaches.
- An ablation study has been conducted to identify the importance of individual input features on model performance.
- An in-depth analysis on the dynamics of COVID-19 vaccine hesitancy in India has been performed.

The remaining paper is organized as follows: Section 2 elucidates some existing studies on Twitter based COVID-19 studies. The description of the data collected and used in the study is discussed in Section 3. The methodology adopted for this study is described in Section 4. The findings and observations of the analysis are discussed in Section 5. Analysis on the dynamics of vaccine hesitancy is provided in Section 6. The study is finally summarized and concluded in Section 7.

2 Related work

Identifying the orientation of social media users towards COVID-19 vaccination, can reveal significant information pertaining to COVID-19 vaccination awareness (Bhat et al., 2020; Chehal et al., 2020). Several studies based on social media usage after COVID-19 and vaccine hesitancy have been proposed in the literature (Shoaei et al., 2020; Jain & Sinha, 2020; Zeemering, 2021; Vicari & Murru, 2020; Umair & Masciari, 2022). This section elicits some of the relevant and important works on vaccine hesitancy and social media analysis during COVID-19.

2.1 Twitter-based studies on COVID-19

Wu et al., (2020) carried out a study to identify the number of infectious persons exported from Wuhan during 1st December 2019 to 25th January 2020. They utilized the flight booking data of COVID positive persons to predict the COVID-19 cases across the country and the rate of national and international proliferation. Medford et al., (2020) measured the frequency of keywords related to infection prevention, racial partiality and inoculation in the COVID-19 related tweets. They performed topic modelling and sentiment analysis to identify the important topics of discussion as well as the dominant emotions of the user during the pandemic.

Results of sentiment analysis revealed that nearly 49.5% of the tweets expressed fear while 30% expressed surprise. With respect to topic modelling, the most common topics were identified to be the financial and political effects of COVID-19. Li et al., (2020) performed online ecological recognition (OER) based modelling to analyze the tweets of 17865 Weibo users. They classified the sentiments of the tweets as depression, happiness, indignation and anxiety and also evaluated the word frequency in the tweets and other indicators like social risk assessment. Their results reflected the increase in negative sentiments and sensitivity to social risks. Pandey et al., (2022) developed a model for predicting the risk of misinformation by utilizing the Hindi language tweets. Their best performing combination produced a Cohens kappa value of 0.54. Gupta et al., (2020) performed sentiment analysis of the Indian public on lockdown. They used eight different machine learning classifiers and identified an overall positive response of the public towards lockdown. Nassem et al., (2021) carried out a study on a large scale dataset to analyze the sentiments of people towards COVID-19 lockdown. Their study revealed that people initially favoured lockdown but the sentiments gradually shifted to negative in the later stages of the pandemic.

Borah and Singh, (2022) studied the presence of political polarization on Twitter during their discussions on Covid19. Rufai and Bunce, (2020) investigated the usage of Twitter by some world leaders during COVID-19 pandemic. They adopted a content analysis approach to categorize the tweets of 7 different world leaders into informative, moral boosting and political categories. Haman (2020) have examined the relationship between the usage of Twitter by state leaders during COVID-19 and the resultant increase in their number of followers due to the same. Their study revealed notable growth in the number of followers of the leaders after the pandemic has started than it was before. Haupt et al., (2021) explored the Twitter discourse of the supporters and opponents of COVID Liberate movement. Their analysis indicated the networks liberate supporters to be centralized while non-liberate supporters to be highly decentralized.

2.2 Machine learning based studies on vaccine hesitancy

Carrieri et al., (2021) performed a study based on area level indicators using several machine learning approaches. They identified communities with high risk of vaccine hesitancy based on indicators like waste recycling and employment rate. Bar et al., (2021) studied the effects of online vaccination contents from Facebook and Tapuz on vaccine hesitancy by employing machine learning strategies. They identified vaccine hesitancy and social media traffic to be highly correlated. Piedrahita-Valdés et al. (2021) employed sentiment analysis on Twitter posts to identify the factors contributing towards vaccine hesitancy. They identified a surge in positive sentiment tweets in the middle of the week that tends to get reduced during weekends. Cotfas et al. (2021) performed a study to identify public opinion regarding COVID-19 vaccination post one month of its initiation in UK. Sentiment analysis has been performed on the tweets posted by public on Twitter using different machine learning and deep learning approaches. They identified the hesitancy to get reduced as duration from the start of vaccination increases.

3 Data collection

This section discusses the data collected and the methodology used for the study. The COVID-19 vaccination drive started in India on 16th January 2021. The tweets therefore, have been considered after that duration. Popular hashtags on COVID-19 vaccination have

been identified. A retrospective search was conducted to identify the tweets containing the considered hashtags and keywords related to COVID-19 vaccination during the duration of 16th January 2021 to 30th November 2021. The Twitter search API has been used to collect the tweets posted by Indian users on COVID-19 vaccination irrespective of the language used. A total of 36,231,457 tweets by 51,682 users have been collected. The geographical metadata in the tweets have been used to identify the users. The tweet location information shared by the user at the time of tweet and the account location information from their twitter profiles that the user themselves specify have been utilized. For collecting COVID vaccination related data, the most popular hashtags on COVID vaccination have been manually identified. Hashtags like #CovidVaccine, #Covaxin, #Covishield, #ReadyToVaccinate etc, have been considered. After identifying the Indian users, these hashtags have been considered to filter the Covid vaccine specific tweets.

This study assumed 2 different classes of the users pertaining to hesitancy about COVID vaccination: hesitant and non-hesitant. To assign the users to a particular class, their average sentiment towards COVID vaccination has been analyzed. To identify the sentiments of the users, SentriStrength tool has been employed. The non-English tweets were translated to English for making it suitable for use by SentiStrength. SentiStrength assigns the tweets a score in the range -4 to +4, where -4 indicates very negative and +4 very positive. The user is assumed to be non-hesitant and have a positive sentiment towards Covid vaccination if the average sentiment is greater than 0 and hesitant if the average sentiment is equal to or less than 0. A user with 0 or neutral sentiment is also considered to be hesitant as lack of surety too indicates hesitancy. After class assignment based on average sentiment, 62% users belonged to the non-hesitant class and 38% to the hesitant class.

Some pre-processing has been performed on the data before computing the features considered in the study. Punctuation symbols, URL's, emoticons and special characters were further removed and all the characters were converted into lower case. NLTK Tweet Tokenizer has been used for tokenization in case of the existing models considered in the study. For tokenization of text used in BERT in the proposed model, the WordPiece transformer tokenizer has been used.

4 Proposed method

This section introduces the proposed multi-modal transformer based model for predicting COVID vaccine hesitancy among Twitter users. The model exploits the advantages of BERT and GraphBERT to learn the respective feature representations of text and network modalities (Fig. 2).

4.1 Multimodal features

Network based features In context of Twitter, network between users may involve different types of proximities like retweet, mention and follower-following relationships. Retweets signify endorsement or support and mostly happen between like-minded users. A user retweeting another user's tweets on COVID vaccination is likely to share the same opinion on it. Mentions and follow relationships on the other hand, do not necessarily represent shared opinion. Hence, in this study, the retweet network information of the users has been utilized as one of the modalities. The content of the tweets have been used for generating an undirected user graph by utilizing the retweet connections between users.
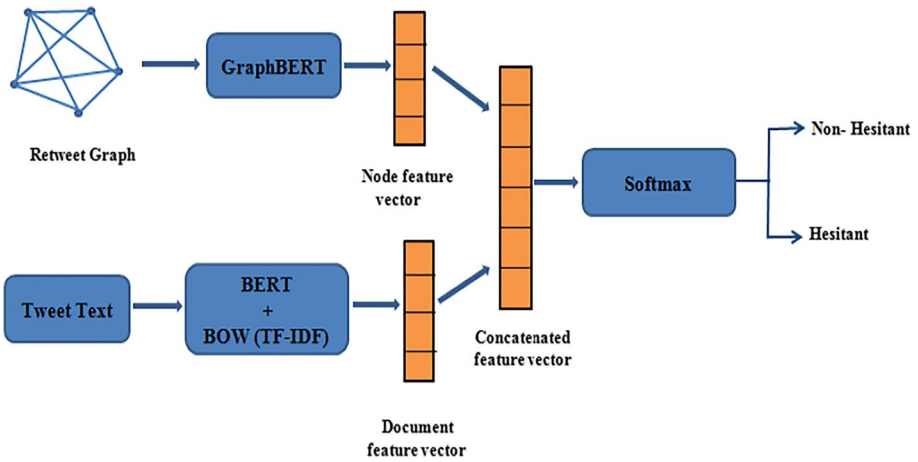
Fig. 2 Proposed Transformer based Multimodal Architecture

The nodes in the graph represents the users and an edge is created among two users if one user directly retweets the other user. Figure 3 shows an example retweet graph.

It is necessary to identify how well the Twitter users are connected to each other in the retweet graph generated for the study. The connectivity of the retweet graph has been analyzed using different graph connectivity measures. Table 1 elicits the statistics of the graph connectivity measures computed for the retweet graph.

The graph has 35,432 number of nodes with 42,763 edges. The average degree of the graph is calculated as the ratio of total number of edges to total number of nodes. Since, the retweet graph has been considered as an undirected graph, the edges connecting the nodes represent reciprocal relationships. Therefore, the number of edges have been considered as twice the edges actually present in the graph. For computing the average clustering coefficient of the graph, the local clustering coefficient of each node is initially calculated. The local clustering coefficient of a node signifies the probability of its distinct neighbors

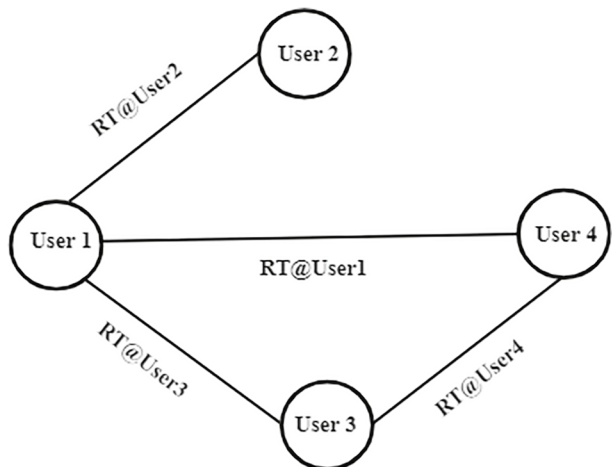**Fig. 3** Example Twitter Retweet Graph

Table 1 Graph metrics for retweet graph of the users

Metric	Value
No. of Nodes	35,432
No. of Edges	42,763
Average Degree	2.41
Average Clustering Coefficient	0.023
Network Density	2.4E-03
Transitivity	0.084

being connected. The values of local clustering coefficient of the nodes are then averaged to obtain the average clustering coefficient of the graph. Network density of the graph is calculated as the ratio of number of edges present in the graph to the total number of edges possible. Transitivity of the graph is computed as the ratio of total number of triangles to the total number of connected triplets of vertices. The network obtained is moderately connected.

Content based features To identify the opinion of the users on COVID vaccination, it is essential to analyze their posted tweet contents. The tweet content of the users have been considered as separate documents. For each user U_i, a document D_i is created containing the content published by them. Two different types of features are identified from the documents: TF-IDF feature and context feature. The set of terms used by the users is obtained from documents $D_1,....,D_N$ and the term frequency of each term t is computed. Term Frequency $TF_i(t)$ for a term t signifies its number of appearances in a document D_i. The document frequency $DF(t)$ for each term t is calculated, that represents the number of documents in which t appears. A Bag of Words (BoW) vector is constructed for each document $D_1,....,D_N$, where each value in the vector is the value of a term t from the set of terms:

$$TFIDF_i(t) = TF_i(t) \cdot log \frac{N}{DF(t)}$$

L2 normalization is applied to the TF-IDF representation of the documents to obtain a unit length. In addition to the TF-IDF features, the context feature of the documents is also captured by mapping it to a continuous valued vector.

4.2 Model architecture

The proposed architecture is depicted in Fig. 2. The model exploits both network based and content based features extracted from the tweet content of the users. Transformer encoder based models BERT and GraphBERT have been employed to learn the feature representations of the textual and network modalities. BERT has been utilized due its capability of capturing the meaning of a word based on both its right and left context. This bidirectionality of BERT makes it superior than other directional models. GraphBERT on the other hand, has been employed keeping in view the problem of suspended animation and over-smoothing faced by traditional graph neural network models. GraphBERT can effectively avoid these problems while performing any graph representation learning task. BERT and GraphBERT have been utilized for feature representation learning as discussed below:

- **BERT-BOW:** The tweet content posted by each user is considered as a separate document. BERT generally accepts a sentence as an input which may not be suitable while considering documents. Since BERT primarily accepts one or two sentences as input, the entire document has been considered as a single sentence. A feature based approach of BERT has been used instead of fine tuning, similar to the study proposed in (Tanaka et al., 2019). The document is divided into morpheme units and the tokens are given as an input to BERT for generating the word embeddings. The word embeddings generated by BERT are combined together to obtain an average vector, which is then considered as the feature vector of the document.

  BERT generates the word embeddings based on the left and right context of the target word. However, word frequency can also be an important feature with respect to documents. Therefore, the Bag of Words (BOW) model has been used to compute the feature vector of the document using the TF-IDF method. The feature vector generated by BOW model is then concatenated with the output feature vector of BERT after normalizing both the feature vectors to unit vectors. The concatenated feature vector is then used as the final feature vector for the document.

- **GraphBERT:** From the input retweet graph with nodes representing the users and the edges representing the retweet interactions between them, GraphBERT initially generates the linkless subgraph batches for all the nodes. For every node in each sampled subgraph, GraphBERT computes the different input vector embeddings: raw feature vector embedding, Weisfeiler-Lehman absolute role embedding, relative positional embedding and relative distance embedding. The initial input vector for all the nodes is an aggregation of all these four generated embeddings. The graph transformer based encoder generates the final representation of the nodes by iteratively updating their initial representations.

The document feature vector generated by BERT with BOW and node feature vector obtained from GraphBERT are then combined using vector concatenation to generate the final concatenated feature vector. The concatenated feature vector is then used for predicting the category of the Twitter users as hesitant and non-hesitant. A feed forward neural network combined with a softmax activation layer is employed for the final prediction of the users with respect to the considered classes. Binary cross-entropy loss function has been considered to calculate the error and optimization of the model is performed using Adam optimizer. L2 regularization has been further performed to prevent the model from overfitting.

5 Results and discussion

This section discusses the experiments performed and the results obtained for the COVID19 vaccine hesitancy prediction task. Pre-trained models BERT with BOW and GraphBERT have been used to generate the intermediate representations for the graph and text based features used in this study. Selection of hyper-parameters can greatly affect the classification performance of the model. Several experiments were performed to identify the optimal values of hyper-parameters. Results with the most optimal hyper-parameter combinations have been reported. The hyper-parameters have been chosen with respect to 5-fold cross validation experiments performed on the dataset. Since the dataset is time related, for using 5-fold cross validation a forward chaining or rolling approach has been adopted. The dataset contains Twitter users data for almost a year. i.e 12 months. The data is divided in such a manner that the first set of training samples contains 2 months data and the testing

sample contains next 2 months data, the second set contains 4 months data in training sample and next 2 months data in the testing sample. Thus the different sets are:

Fold 1: Train[1,2]; Test[3,4]
Fold 2: Train[1,2,3,4]; Test[5,6]
Fold 3: Train[1,2,3,4,5,6]; Test[7,8]
Fold 4: Train[1,2,3,4,5,6,7,8]; Test[9,10]
Fold 5: Train[1,2,3,4,5,6,7,8,9,10]; Test[11,12]

Here, the number inside brackets represent the respective months. This process has been adopted to ensure that the training set does not contain information that occurs after the test set. The hyper-parameters considered for the pre-tained transformer models are elicited in Table 2.

5.1 Models considered for evaluation

The proposed approach employs pre-trained transformer models BERT and GraphBERT for learning the low-dimensional feature representations of the textual and network modalities considered in the study. A comparative study has been carried out by using different traditional deep learning models for capturing the feature representation of the textual and network features considered. For an unbiased comparison, several experiments were conducted with different variations of unimodal and multimodal approaches of existing traditional models and the proposed approach. A ratio of 80:20 has been adopted to split the labeled data into train and test set. The different combinations of unimodal and multimodal based approaches used for comparison are discussed below:

- **LSTM** (Hochreiter and Schmidhuber, 1997): The model uses textual modality for classification. The first layer of the model is an embedding layer representing each word as a real-valued feature vector. The embedding layer is followed by an LSTM layer with 100 memory units. An output dense layer combined with softmax activation function has been used for classification. A batch size of 64, drop out of 0.3 and cross-entropy loss function has been considered.

Table 2 Hyper-parameters considered for pre-trained models

Model	Parameter	Value
GraphBERT	Hidden Layer Number	2
	Subgraph Size	7
	Learning Rate	5e-4
	Hidden Size	32
	Hidden Dropout Rate	0.5
	Attention Head Number	3
	Attention Dropout Rate	0.4
	Weight Decay	0.01
BERT	Number of Transformer Blocks	12
	Training Batch Size	32
	Learning Rate	5e-4
	Weight Decay	0.01

- **BiLSTM** (Graves & Schmidhuber, 2005): The model consists of an initial embedding layer for mapping the text into vectors, followed by a two-way LSTM layer (BiLSTM). The output of the BiLSTM is then fed to the output dense layer for classification. Binary cross-entropy loss function with a drop out of 0.3 and batch size of 64 have been considered.

- **BERT Base** (Devlin et al., 2018): A sequence of tokens (words) along with CLS and SEP tokens are provided as an input to BERT Base model after tokenization of the text. Truncation is done for the tokens in the sequence, longer than 512. The output of BERT is a vector embedding of size 768 in each token. The output vector from CLS token is then passed as an input to the linear layer with ReLU activation function. Adam optimizer, binary cross-entropy loss and a learning rate of 1e-6 has been used for training the model.

- **GCN** (Kipf & Welling, 2016): The input graph represented by adjacency matrix and node feature matrix is given as an input to the GCN layers. Two graph convolution layers having 64 units and relu activation function have been used. The learned node representation is then passed to a fully connected layer which is followed by a final output layer with softmax activation function. Binary cross-entropy loss and Adam optimizer has been used for model training.

- **GraphBERT** (Zhang et al., 2020): Pre-trained GraphBERT has been utilized to obtain the node feature representation of the input retweet graph. A subgraph size of 5, hidden size of 32, and 2 attention head has been considered. Hidden drop out rate of 0.5 and attention drop out rate of 0.3 with cross entropy loss and Adam optimizer has been used for model training. The node feature representation generated by GraphBERT is then fed to a feed forward network with softmax activation layer for classification.

- **Doc2Vec+Node2Vec**: The document feature vector is computed using pre-trained Doc2Vec model and the node feature vector is obtained using the Node2Vec model. The feature representations generated by both the models is then concatenated and passed to a feed-forward network with softmax activation for final prediction.

- **TF-IDF+Doc2Vec+Node2Vec**: Along with the Doc2Vec and Node2Vec features, TF-IDF features are also being considered. All the feature representations obtained from TF-IDF, Doc2Vec and Node2Vec are concatenated and passed to a final output layer with softmax activation for classification.

5.2 Evaluation results

The performance of the models considered in the study are evaluated using the measures of accuracy, precision, recall and F1-score. The results obtained for different models are depicted in Table 3. The first section of the table reports the results on traditional models leveraging single modalities. Second section on the other hand, illustrates the results obtained by the proposed model along with the traditional models employing multiple modalities. Among the models exploiting only textual modality, BiLSTM obtained the highest precision value while BERT achieved the best performance in terms of accuracy, and recall. Both the models performed almost similar in terms of F1-score. The models leveraging network features alone displayed better results over textual modality based models. Among the network feature based models, GCN achieved better precision while Graph BERT performed better in terms of accuracy, recall and F1-score. Among the models exploiting multiple modalities, the proposed approach achieved better results in terms of accuracy, precision, recall and F1-score.

Table 3 Comparative analysis with unimodal and multimodal approaches

	Modality	Model	Accuracy(%)	Precision(%)	Recall(%)	F1-score(%)
Unimodal	Text	LSTM	77.24	78.36	81.52	79.92
	Text	BiLSTM	79.74	83.32	81.21	82.23
	Text	BERT	83.45	81.67	82.64	82.15
	Graph	GCN	86.42	87.34	85.57	86.44
	Graph	GraphBERT	87.36	86.56	87.79	87.19
Multimodal	Text+Graph	Doc2Vec+Node2Vec	90.17	89.64	88.75	89.19
	Text+Graph	TFIDF+Doc2Vec+Node2Vec	92.34	91.23	90.34	90.74
	Text+Graph	Proposed	95.24	94.67	92.34	93.48

Table 4 Results on different modalities

Modality	Accuracy(%)	Precision(%)	Recall(%)	F1-score(%)
Text	86.35	86.43	83.12	84.72
Graph	87.36	86.56	87.79	87.19
Proposed	95.24	94.67	92.34	93.48

The proposed model architecture is capable of exploiting different features with respect to the paradigm of multimodal learning. In this study, the model has been realized using textual and network features. It is essential to identify the individual contribution of each feature towards the discriminating strength of the model. Hence, in order to understand the importance of each modality, an ablation study has been performed considering the textual and network features separately. The results obtained from each individual modality is elicited in Table 4. Results indicate that utilizing only network modality leads to better performance than using only textual modality. However, utilizing multiple modalities, i.e concatenating both text and network features, enhances the performance of the model to a greater extent than leveraging either text or network modality alone. Results on textual modality indicates that BERT feature representation combined with TF-IDF representation achieved better results than utilizing BERT representation alone. This can be observed in Table 3 where, BERT without using TF-IDF obtained lower evaluation scores than BERT with TF-IDF.

To gauge the models ability for correct prediction, confusion matrices have been constructed with respect to different modalities. Error analysis has been performed to identify the miss-classification rate for different modalities. The comparative confusion matrices for textual, network and multi-modality are shown in Fig. 4. The confusion matrices in the figure are represented in terms of percentage. The x-axis in the matrices represent actual or true labels while the y-axis represents the predicted labels by the proposed model. The diagonal values in darker shades represent the percentage of users correctly identified as being hesitant and non-hesitant regarding COVID vaccination. For instance, using only textual modality, the model could predict the users as non-hesitant 78% of the time and as hesitant 83% of the time. With only network-based features, the model could identify the users as non-hesitant 82% of the time and as hesitant 87% of the time. Using both textual and network modality the model could correctly classify the users as non-hesitant 86% of the time and as hesitant 91% of the time. The non-diagonal values in lighter shade on the other hand, represent how often the users are being mis-identified as being hesitant or

non-hesitant. For example, using text as the only modality, the users will be mis-identified as hesitant 17% of the time and as non-hesitant 21% of the time. With network modality, the mis-classification rate for hesitant users is 13% and for non-hesitant users 18%. The lowest mis-classification rate of 9% for hesitant users and 14% for non-hesitant users is obtained while using both textual and network modalities.

6 Dynamics of vaccine hesitancy

After testifying the proposed model and carrying out a comparative study, the proposed approach has been used to identify the hesitant users for COVID vaccination. A time wise study has been performed using the proposed model to observe the hesitancy among users during different periods of the study. Tweets for every month starting from 16th Jan 2021 to 15th Dec 2021 have been considered separately and the corresponding retweet interaction networks for the users have been generated. Based on the tweet text and the retweet networks, the users are further classified by the model as either hesitant or non-hesitant. The hesitancy dynamics has been analyzed to demonstrate how the users hesitancy regarding COVID vaccine has evolved over the period of study.

Figure 5 illustrates the overall hesitancy dynamics of the users during the study period between 16th Jan-15th Dec 2021. *Hesitancy level* denotes the number of hesitant users
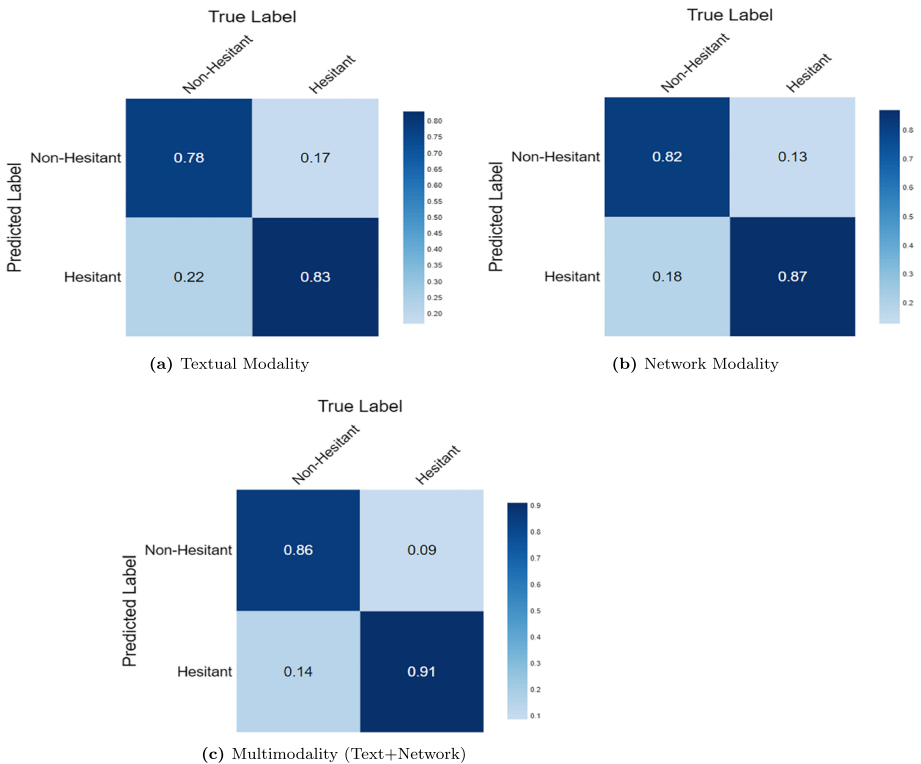

(a) Textual Modality

(b) Network Modality

(c) Multimodality (Text+Network)

Fig. 4 Confusion matrix of proposed approach using different modalities
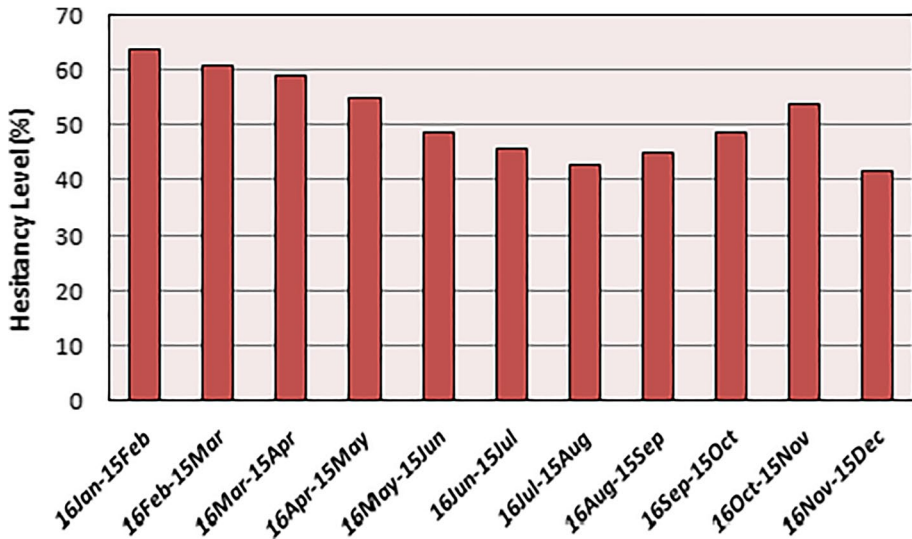
Fig. 5 Hesitancy Dynamics

over the total number of users each month. From the figure, it can be observed that the hesitancy level was highest at the start of the vaccination drive in 16th Jan 2021. The level remained almost consistent till the month of March-April. This high rate of hesitancy during the onset of vaccination drives is primarily due to the uncertainty regarding the side effects of the vaccine. A drop in the hesitancy level can be seen in the month of May after the start of the second wave of the pandemic in the country. The level continued to drop till the month of July. There is again a gradual rise in the level of hesitancy from the the month of August till November. The highest peak has been observed in the month of November. The prime reason for this increase is the reluctance of users in getting vaccinated due to their prevailing medical condition and their lack of confidence in the efficacy of the vaccine. It has also been observed that the hesitant users are not fully convinced that the currently administered vaccines could also provide protection from the existing and future variants. However, the highest drop in the hesitancy level has been observed in the month of December with the rise in the Omicron cases in the country. As the omicron cases are rising, more users have expressed their willingness in taking the vaccination.

7 Conclusion and limitations of the study

The COVID-19 outbreak has caused severe economic, political and social disruptions in several countries including India. Among several measures taken by the government to fight the disease, vaccination has been identified as the most effective preventive measure for controlling the emergency situation. Social media platforms have been disseminating both useful information as well as misinformation regarding important issues like the COVID-19 pandemic. With the advent of COVID vaccination drives in India, social media users are explicitly expressing their views on the outcome of vaccination

for eradicating the disease. Negative information and opinions regarding the effectiveness of vaccines may generate hesitancy among the masses which may hamper the prevention of this deadly disease. This study therefore, attempts to identify the hesitancy among Indian Twitter users based on their opinions about the effectiveness of COVID-19 vaccination.

The sentiments of the twitter users on COVID-19 vaccination have been identified and the users are assigned to classes non-hesitant and hesitant based on their average sentiments. A multimodal transformer based approach has been proposed that is capable of leveraging both textual and network modalities. Tweet content of the users and their corresponding retweet interaction structures have been utilized as the respective textual and network feature modalities. A comparative analysis of the proposed approach has been performed with state-of the art deep learning approaches. The proposed model outperformed the considered unimodal and multimodal approaches with respect to the evaluation measures of accuracy, precision, recall and F1-score. An ablation study has also been performed to identify the individual contribution of each modality. The results of the study suggest that although network modalities contribute more towards the performance of the model, combining both textual and network modalities further enhances the classification performance. After testifying the classification accuracy of the proposed model, it has been utilized to perform a study on the dynamics of COVID-19 vaccine hesitancy in India. The findings of the study reveal the hesitancy to be highest immediately after the onset of the vaccination programme which gradually reduced in the subsequent months. Slight increase in hesitancy has been observed in the month of August that gradually increased till November. The lowest hesitancy however, was observed in the month of December.

The current study has certain limitations which can be further addressed. The study is based on Twitter platform that represents only a portion of the native users. The platform is primarily used by the urban residents due to which the opinions of majority of the rural area residents could not be included. Moreover, the tweets collected for the study are entirely based on the hashtags used. Hence, usage of hashtags different from the current set might have generated different results.

Declarations

References

Bar-Lev, S., Reichman, S., & Barnett-Itzhaki, Z. (2021). Prediction of vaccine hesitancy based on social media traffic among israeli parents using machine learning strategies. *Israel Journal of Health Policy Research*, *10*(1), 1–8. https://doi.org/10.1186/s13584-021-00486-6.

Bhat, M., Qadri, M., Noor-ul Asrar Beg, M.K., Ahanger, N., & Agarwal, B. (2020). Sentiment analysis of social media response on the covid19 outbreak. *Brain, Behavior, and Immunity*, *87*, 136–137. https://doi.org/10.1016/j.bbi.2020.05.006.

Borah, A., & Singh, S.R. (2022). Investigating political polarization in India through the lens of twitter. *Social Network Analysis and Mining*, *12*(1), 1–26. https://doi.org/10.1007/s13278-022-00939-z.

Carrieri, V., Lagravinese, R., & Resce, G. (2021). Predicting vaccine hesitancy from area-level indicators: A machine learning approach. *Health Economics*, *30*(12), 3248–3256. https://doi.org/10.1002/hec.4430.

Chehal, D., Gupta, P., & Gulati, P. (2020). Covid-19 pandemic lockdown: An emotional health perspective of indians on twitter. *International Journal of Social Psychiatry*, *67*(1), 64–72. https://doi.org/10.1177/0020764020940741.

Cotfas, L.A., Delcea, C., & Gherai, R. (2021). Covid-19 vaccine hesitancy in the month following the start of the vaccination process. *International Journal of Environmental Research and Public Health*, *18*(19), 10438. https://doi.org/10.3390/ijerph181910438.

Devlin, J, Chang, M-W, Lee, K, & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. https://doi.org/10.48550/arXiv.1810.04805.

Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional lstm and other neural network architectures. *Neural Networks*, *18*(5-6), 602–610. https://doi.org/10.1016/j.neunet.2005.06.042.

Gupta, P., Kumar, S., Suman, R., & Kumar, V. (2020). Sentiment analysis of lockdown in India during covid-19: A case study on twitter. *IEEE Transactions on Computational Social Systems*, *8*(4), 992–1002. https://doi.org/10.1109/TCSS.2020.3042446.

Haman, M. (2020). The use of twitter by state leaders and its impact on the public during the covid-19 pandemic. *Heliyon*, *6*(11), e05540. https://doi.org/10.1016/j.heliyon.2020.e05540.

Haupt, M.R., Jinich-Diamant, A., Li, J., Nali, M., & Mackey, T.K. (2021). Characterizing twitter user topics and communication network dynamics of the "liberate" movement during covid-19 using unsupervised machine learning and social network analysis. *Online Social Networks and Media*, *21*, 100114. https://doi.org/10.1016/j.osnem.2020.100114.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Jain, S., & Sinha, A. (2020). Identification of influential users on twitter: A novel weighted correlated influence measure for covid-19. *Chaos, Solitons & Fractals*, *139*, 110037. https://doi.org/10.1016/j.chaos.2020.110037.

Kipf, T.N., & Welling, M. (2016). Semi-supervised classification with graph convolutional networks. https://doi.org/10.48550/arXiv.1609.02907.

Li, S., Wang, Y., Xue, J., Zhao, N., & Zhu, T. (2020). The impact of covid-19 epidemic declaration on psychological consequences: a study on active weibo users. *International Journal of Environmental Research and Public Health*, *17*(6), 2032. https://doi.org/10.3390/ijerph17062032.

Medford, R.J., Saleh, S.N., Sumarsono, A., Perl, T.M., & Lehmann, C.U. (2020). An "infodemic": leveraging high-volume twitter data to understand early public sentiment for the coronavirus disease 2019 outbreak. *Open Forum Infectious Diseases*, *7*(7), ofaa258. https://doi.org/10.1093/ofid/ofaa258.

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. https://doi.org/10.1109/TCSS.2021.3051189.

Pandey, R., Gautam, V., Pal, R., Bandhey, H., Dhingra, L.S., Misra, V., Sharma, H., Jain, C., Bhagat, K., Patel, L., & et al (2022). A machine learning application for raising wash awareness in the times of covid-19 pandemic. *Scientific Reports*, *12* (1), 1–10. https://doi.org/10.1038/s41598-021-03869-6.

Piedrahita-Valdés, H, Piedrahita-Castillo, D., Bermejo-Higuera, J., Guillem-Saiz, P., Bermejo-Higuera, J.R., Guillem-Saiz, J., Sicilia-Montalvo, J.A., & Machío-regidor, F (2021). Vaccine hesitancy on social media: Sentiment analysis from june 2011 to april 2019. *Vaccines*, *9*(1), 28. https://doi.org/10.3390/vaccines9010028.

Rufai, S.R., & Bunce, C. (2020). World leaders' usage of twitter in response to the covid-19 pandemic: a content analysis. *Journal of Public Health*, *42* (3), 510–516. https://doi.org/10.1093/pubmed/fdaa049.

Shoaei, M.D., Dastani, M., & et al. (2020). The role of twitter during the covid-19 crisis: a systematic literature review. *Acta Informatica Pragensia*, *9*(2), 154–169. https://doi.org/10.18267/j.aip.138.

Tanaka, H., Shinnou, H., Cao, R., Bai, J., & Ma, W. (2019). Document classification by word embeddings of bert. In *International Conference of the Pacific Association for Computational Linguistics, Springer,* pp 145–154.

Umair, A., & Masciari, E. (2022). Sentimental and spatial analysis of covid-19 vaccines tweets. *Journal of Intelligent Information Systems*, 1–21. https://doi.org/10.1007/s10844-022-00699-4.

Vicari, S., & Murru, M.F. (2020). One platform, a thousand worlds: On twitter irony in the early response to the covid-19 pandemic in italy, (Vol. 6. https://doi.org/10.1177/2056305120948254.

Wu, J.T., Leung, K., & Leung, G.M. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, china: a modelling study. *The Lancet*, *395*(10225), 689–697. https://doi.org/10.1016/S0140-6736(20)30260-9.

Zeemering, E.S. (2021). Functional fragmentation in city hall and twitter communication during the covid-19 pandemic: Evidence from atlanta, san francisco, and washington, dc. *Government Information Quarterly*, *38*(1), 101539. https://doi.org/10.1016/j.giq.2020.101539.

Zhang, J., Zhang, H., Xia, C., & Sun, L. (2020). Graph-bert: Only attention is needed for learning graph representations. https://doi.org/10.48550/arXiv.2001.05140.