Machine Learning
https://doi.org/10.1007/s10994-024-06562-7

Imbalanced COVID-19 vaccine sentiment classification with synthetic resampling coupled deep adversarial active learning

Sankhadeep Chatterjee[1] · Saranya Bhattacharjee[2] · Asit Kumar Das[1] · Soumen Banerjee[3]

Received: 14 May 2022 / Revised: 22 January 2024 / Accepted: 26 April 2024

Abstract
Despite an abundance of scientific evidence supporting the effectiveness of COVID-19 vaccines, there has been a recent global surge in vaccine hesitancy, primarily driven by the spread of misinformation on social media platforms. It is crucial to address this issue and raise awareness about the importance of vaccination in combating the deadly COVID-19 virus. Predicting community sentiment through social media platforms can provide valuable insights into vaccine hesitancy, aiding health workers and medical professionals in taking necessary precautionary measures. However, the lack of high-quality labeled data presents a challenge for building an effective COVID-19 sentiment classifier. Additionally, the available labeled datasets suffer from severe class imbalance. To address these challenges, this article presents an effective COVID-19 sentiment prediction framework. Firstly, a deep adversarial active learning framework leverages abundant unlabeled data by training autoencoder and discriminator components adversarially to select the most informative unlabeled samples. Secondly, to mitigate the effects of imbalanced labeled datasets, a resampling phase is incorporated into the adversarial training loop. The proposed framework, named Resampling Supported Deep Adversarial Active Learning (RS-DAAL), is rigorously evaluated using two different datasets comprising social media posts from Twitter and Reddit. Various resampling techniques, including undersampling, oversampling, and hybrid methods, are assessed, with oversampling techniques further tested at different levels of resampling. Comparative studies are conducted against a baseline model without any resampling layer and with current state-of-the-art methods as well. Experimental results and statistical analysis demonstrate the superiority of the proposed RS-DAAL method in identifying COVID-19 sentiments on social media platforms.

Keywords COVID-19 · Sentiment analysis · Deep active learning · Class imbalance

Extended author information available on the last page of the article

Published online: 15 July 2024

1 Introduction

In response to the rapid global transmission of the COVID-19 virus, country-wide lockdowns have had a significant influence on people's daily lives (Alamoodi et al., 2021b). Social media analysis has offered a deep insight into the trends and public opinion in light of the Coronavirus pandemic (Noor et al., 2020; Amjad et al., 2021). Reflecting the urgency to develop preventive measures to curb the spread of SARS-CoV-2, mass vaccination drives have proven to be remarkably effective (Sattar & Arifuzzaman, 2021). In the past twelve months, several researchers have studied and analyzed the general viewpoint of people regarding COVID-19 vaccines (Nwafor et al., 2021; Chakraborty et al., 2020). However, one of the challenges being faced by the research community, today, has been obtaining a large, superior-quality labeled dataset. Furthermore, to obtain such large datasets, manual annotation requires a certain level of expertise and is time-consuming. Hence, it has been critical to figure out a strategy to maximize the model's performance gain when annotating a limited number of instances. In the past few years, Deep Active Learning (DAL) has been demonstrated to improve labeling efficiency exponentially in the field of sentiment classification (Naseem et al., 2021; Longpre et al., 2022). Some studies have worked on Active Learning (AL) techniques for BERT-based classification (Dor et al., 2020). Also, prior studies have established that when dealing with an extreme imbalance in the labeled dataset, sentiment classification remains biased towards the majority class, ultimately leading to a deceptive analysis (Miller et al., 2020). To address this issue, we have explored class imbalance in feature space using a Bi-LSTM variational autoencoder (VAE) in an adversarial manner and therefore developed a Deep Adversarial Active Learning framework efficient enough to enhance the classification performance for an imbalanced COVID-19 vaccine sentiment analysis.

1.1 Related work

Social media has a significant impact on how people perceive and decide about COVID-19 vaccines. Some people use social media to spread false or misleading claims about COVID-19 vaccines, which can lower public trust and confidence in vaccine safety and efficacy (Wilson & Wiysonge, 2020; Al-Hajri et al., 2021). People who get a lot of their news from social media tend to be more doubtful and reluctant about getting vaccinated (Mitchell et al., 2021). On the other hand, social media can also be used to share accurate and positive information about covid vaccines, and to communicate with people who have questions or worries about vaccination (Wilson & Wiysonge, 2020; Zhang et al., 2021). A considerable amount of statistical analysis has been conducted in the last two years since the first official case of COVID-19 was reported in Wuhan, China, in December 2019. Various researchers have studied the general sentiment around the globe to monitor people's presence of mind (Lwin et al., 2020; Müller et al., 2020; Xue et al., 2020). There have been various schools of thought that have originated since the approval of vaccination drives worldwide (Amjad et al., 2021; Rahman et al., 2022). Imran et al. studied the reactions of citizens of diverse cultures to COVID-19 with the help of deep LSTM models and analyzed the emotional response from extracted tweets, achieving state-of-the-art accuracy on the sentiment140 dataset (Imran et al., 2020). A novel fusion model for sentiment analysis of tweets was proposed by combining state-of-the-art transformer-based deep models (Basiri et al., 2021). In Naseem et al. (2021), the authors presented a comprehensive study on an optimized framework for automated COVID-19 detection. Wu et al. proposed a weakly supervised deep AL framework (COVID-AL) to diagnose COVID-19 with CT scans and patient-level (Xing et al., 2021). Alamoodi et al. Alamoodi et al. (2021a) documented an extensive review of vaccine hesitancy via sentiment analysis and illustrated a detailed study into the community sentiment of the public. The authors in Muqtadiroh et al. (2021) well-documented the effects of class imbalance in handling public opinion of the school-from-home policy during the pandemic. A thorough study on sentiment towards COVID-19 vaccines in the Philippines (Villavicencio et al., 2021) was undertaken using Naıve Bayes, attaining an accuracy of 81.77%. In Bhoj et al. (2021), the authors dealt with identifying tweets on social media that conveyed an anti-vaccine sentiment. With the highest F1-Score of 87.179, the research indicated that SVM was best fitted to identify negative tweets on a balanced dataset, while KNN showed the highest improvement after mitigating class imbalance using Edited Nearest Neighbour (ENN). Prior research (Amjad et al., 2021; Bhoj et al., 2021) has exhibited that since deep learning models are known to be data-driven, it has become crucial to alleviate the class imbalance problem in the dataset to obtain an honest analysis.

Deep Active Learning approaches have received a notable amount of traction in various fields, including image classification (Beck et al., 2021; Mittal et al., 2019), multimedia image processing (Dhiman et al., 2023), machine translation (Peris & Casacuberta, 2018; Stafanovičs et al., 2020), medical imaging (Liu et al., 2020a; Yang et al., 2017; Nam et al., 2019), speech recognition (Luo et al., 2021; Abdelwahab & Busso, 2019), visual tracking (Yuan et al., 2023), and object detection (Han et al., 2020; Li et al., 2021). Recent advancements in DAL in the field of text classification (Siddhant & Lipton, 2018; Liu et al., 2020b; Zhou et al., 2013) have proved to be an effective solution thereby reducing the labeling cost significantly. An exceptional thorough survey was conducted on DAL by the authors in Ren et al. (2021). In medium and large query batch sizes, Gissin et al. introduced the Discriminative AL (Gissin & Shalev-Shwartz, 2019) algorithm, which showed to be on par with state-of-the-art models. The authors of Huang et al. (2019) addressed vehicle type detection in surveillance images using deep AL and developed a solution that effectively decreased the annotation cost by up to 40%. Ash et al. proposed "Batch Active Learning by Diverse Gradient Embeddings" (BADGE) (Ash et al., 2019), enabling the predictive uncertainty and the variety of the instances, for each batch, to be considered at the same time. Zhu and Bento reported an extraordinary AL method via query synthesis approach, Generative Adversarial AL, Zhu and Bento (2017) that outperformed typical pool-based techniques. Further research by Tran et al. (2019) in another study proposed Bayesian generative active deep learning approach(BGADL) (Tran et al., 2019) incorporating Bayesian data augmentation, GAAL (Zhu & Bento, 2017), VAE (Kingma & Welling, 2013) and auxiliary-classifier generative adversarial networks (ACGAN) (Dash et al., 2017) algorithms. The authors Yoo and Kweon (2019), in 2019, demonstrated a novel task-agnostic AL approach that integrated a loss prediction module to a target network to assist it learn to anticipate target losses of unlabeled inputs. The authors Goudjil et al. (2018) selected a batch of informative samples based on the posterior probabilities given by a collection of multi-class SVM classifiers, resulting in a notable gain in classification accuracy while lowering labelling effort. A remarkable Deep Active Self-paced Learning (DASL) (Wang et al., 2018) approach was developed, and when tested on the publicly available LIDC-IDRI dataset, the Nodule R-CNN produced state-of-the-art performance in pulmonary nodule segmentation. In Yan et al. (2020), the authors demonstrated an AL strategy for text classification by automatically generating the most insightful instances based on the classification model. A novel pool-based AL strategy was proposed by the authors

(Geifman & El-Yaniv, 2017), where data points were queried from the pool using traversals from the farthest point in the region of neural activity across a representation layer. In Sahan et al. (2021), a thorough comparative analysis on multiple AL approaches for several embeddings of the text was conducted. The study emphasized on the Bayesian AL methods and evaluated the approach on various datasets. Similarly, a BERT-based active learning strategy (Prabhu et al., 2021) was employed to explore the multi-class text classification problem. In Kwolek et al. (2019), the authors explored class imbalance in breast cancer histopathological images and proposed an effective deep convolutional neural network-based active learning framework, enabling weighted information entropy.

1.2 Motivation and objective

It is evident from Sect. 1.1 that class imbalance remains a severe obstacle that has a significant impact on classifier performance. Moreover, an imbalanced distribution of classes is expected when dealing with real-world datasets. Recent studies on COVID-19 sentiment classification (Alamoodi et al., 2021b; Amjad et al., 2021; Joloudari et al., 2023) have established the effects of this class bias problem. Over the past few years, research conducted on DAL frameworks has exhibited superior model performance and therefore garnered significant attention from the research community. In Ren et al. (2021), it has been observed how effectively DAL frameworks have improved the model performance while annotating as few instances as possible. In Bashar and Nayak (2021), the authors proposed a Mixed Aspect Sampling (MAS) framework, which remarkably performed better than random sampling and other state-of-the-art AL methods. Additionally, the MAS framework was efficient enough to deal with an imbalanced dataset. Despite the exponential model optimization, recent studies have revealed that DAL frameworks, too, suffer from class imbalance heavily. In an adversarial DAL setting, as studied in Sinha et al. (2019) and Kim et al. (2021), it has been detected that when dealing with a highly imbalanced labeled set, the training is biased towards the majority class. Yet, little effort has been made to alleviate the effect of imbalanced classes. Hence, in this article, a synthetic resampling-based deep adversarial AL framework has been proposed to tackle the imbalanced class distribution of the labeled set. Besides, the adversarial training of the components ensures that the most informative data samples from the unlabeled pool are selected for the query. In our approach, the synthetic resampling is embedded in the adversarial training loop to eventually force the model to select the most prominent unlabeled data for querying.

1.3 Contribution

The current study deals with efficient COVID-19 vaccine sentiment prediction to understand public opinion toward vaccine hesitancy. In the absence of a good quality-labeled dataset, the authors have engaged a deep active learning framework to take the leverage of abundantly available unlabeled data. However, the traditional DAL-based methods lack in proper querying of the unlabeled data pools. This leads to the minimum or almost no improvement of the labeled dataset. Hence, the performance of classifiers could not be improved adequately. Thus, in the current study, an adversarial approach has been adopted to address this issue. Firstly, a BiLSTM-based variational autoencoder has been trained with data instances that contain both labeled and unlabeled data. The unsupervised training of the VAE enables it to learn the most efficient latent vector representation of input text data instances. However, in the presence of imbalanced labeled data, the subsequent querying step becomes biased toward the majority class. Therefore, the latent vectors are resampled to obtain balanced labeled latent vectors. Next, a discriminator similar to the one used in generative adversarial models has been engaged to distinguish between true labeled and unlabeled data. The unlabeled set as obtained from the discriminator is queried and subsequently labeled by an Oracle. In the current study, a separately trained multi-layer perceptron model is used as the Oracle. The newly labeled data instances are added to the existing labeled pool to train the task learner for COVID-19 sentiment prediction. The autoencoder and discriminator are trained in an adversarial loop to force the querying stage to select the most informative samples from the unlabeled pool. To improve the discriminator performance, various resampling techniques have been explored with detailed analysis. The task learner is simulated by using eight different well-known classifiers in terms of F1-score, Geometric mean, and Balanced Accuracy. To establish the ingenuity of the results obtained by various configurations used in the current study, the receiver operating characteristics curve has been used to compare the performance of the models. Two different datasets consisting of social media posts from Twitter and Reddit are used to test the proposed RS-DAAL model which achieved significant improvement over the baseline model. To understand the effect of different levels of imbalance ratio, separate experiments are conducted in case of oversampling methods with various degrees of resampling. Furthermore, a separate comparative analysis is conducted with current state-of-the-art methods in terms of F1-score. Overall the following are the major contributions of the current study:

1. Adversarial training of deep active learning framework has been explored for effective prediction of COVID-19 sentiment.
2. The effects of the imbalanced labeled dataset in the DAL model have been addressed by involving the resampling of latent vectors inside the adversarial loop.
3. To achieve an unbiased training of the Oracle, top-k most confident labeled data samples have been used to train it.
4. Wide range of resampling techniques have been thoroughly tested using two different datasets from two different social media platforms to better understand the suitability of undersampling, oversampling and hybrid methods in the context of COVID-19 vaccine sentiment classification.

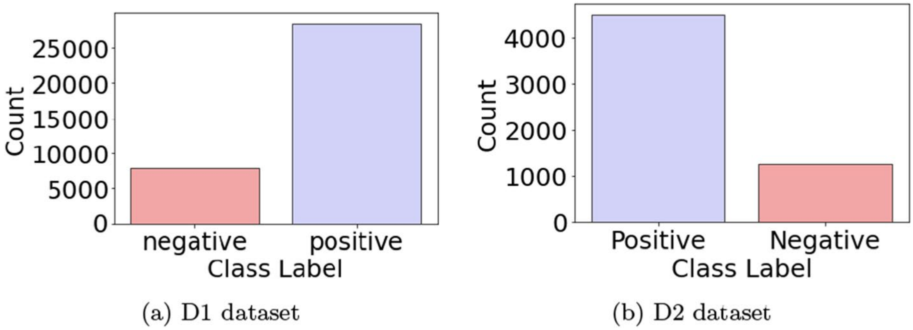
The rest of the article is arranged as follows: Sect. 2 describes the dataset used in the current study. Next, in Sect. 3, the deep active learning framework has been explained in detail. This is followed by Sect. 4 which introduces the proposed RS-DAAL method to predict COVID-19 vaccine sentiment. Section 5 describes the baseline model, resampling techniques, task learners along with all parametric setups. In addition, various performance metrics used to measure the performance are also described. Finally, Sect. 6 reports the experimental results and analysis.

2 Dataset description

The first dataset denoted as the D1 dataset throughout the article has been obtained from ‘COVID-19 All Vaccines Tweets’, created by Gabriel Preda, and is publicly available on Kaggle (Preda, 2021). The tweets were extracted using the Twitter API, utilizing filter criteria based on vaccine-related hashtags. The original dataset features 2,28,207 unique tweets, acquired between 12th December 2020 and 24th November 2021 in its 113th version. Studies have been conducted on the subsequent dataset (Alam et al., 2021; Prabucki, 2021; Alanazi, 2021); however, class imbalance in a Deep adversarial AL scenario has not been explored yet. Preprocessing and cleaning the text field has been among the initial steps in our experiment, which involved lowercase conversion, removal of URLs, punctuation, emojis, double spacing, and stopword removal. We have developed a dataset of 68,035 tweets with the objective of building an unlabeled pool for batch-mode active learning. The rest of the tweets were manually annotated as positive or negative for the purpose of our study. A group of undergraduate students, and research scholars who were fluent in English and familiar with social media language voluntarily participated in annotating the tweets. We instructed them to label a tweet as neutral if it did not express any clear or strong sentiment, or if it expressed both positive and negative sentiment. We asked each tweet to be labeled by at least three different annotators, and we computed the majority vote for the final label. We found that 28,420 tweets were labeled as positive, 7990 tweets were labeled as negative, and 1,23,762 tweets were labeled as neutral by the majority of annotators. We decided to discard the neutral tweets from our dataset, as we were interested in the binary classification of positive and negative sentiment, which is a common and challenging task in sentiment analysis. Moreover, we observed that the neutral class was very subjective and difficult to define and annotate, as different annotators might have different interpretations of what constitutes a neutral tweet. For example, some annotators might consider a tweet neutral if it contains factual information or sarcasm, while others might consider it positive or negative depending on the tone or context. Therefore, we only kept the tweets that had a clear positive or negative sentiment, resulting in a final dataset of 36,410 tweets.

Accordingly, the labeled D1 dataset (as shown in Fig. 1a) is developed containing 28,420 positive and 7990 negative sentiment tweets. The imbalance ratio in our labeled D1 dataset is recorded to be 3.56. From Fig. 2, it is visible that Indian Twitter users have been most active during the pandemic. Evidently, in the labeled dataset (Fig. 2a), tweets originating from the United States, the United Kingdom, Canada, Russia, and China have been observed. Although, in the unlabeled set (Fig. 2b), tweets generated from India have outnumbered tweets originating from other countries.

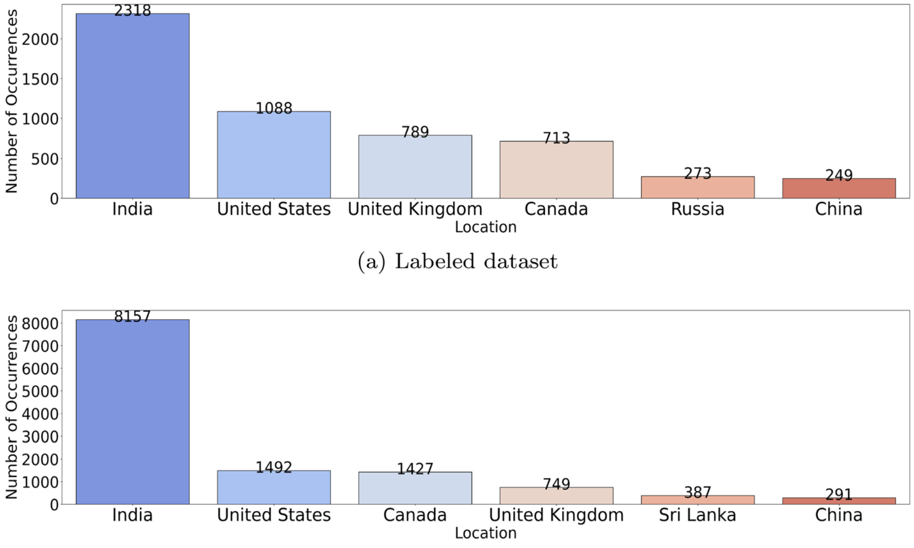
Word visualization for the D1 dataset based on the sentiment polarity has been shown in Fig. 3, offering an intuition into the most prevalent keywords. In (Fig. 3a), it has been noticed that along with vaccine dissatisfaction, there has been animosity towards Justin Trudeau and Doug Ford, owing to the 2021 Canadian federal election. Words like ’safe’,



(a) D1 dataset

(b) D2 dataset

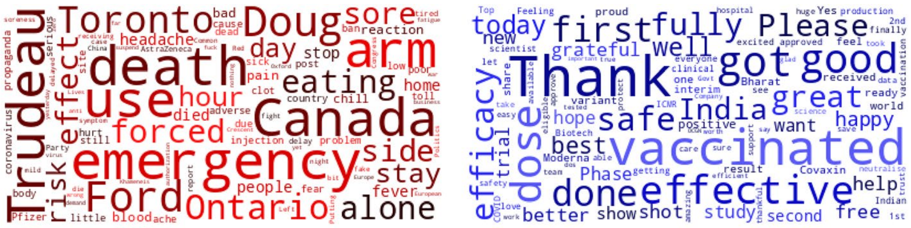
Fig. 1 Class distribution in labeled datasets



(a) Labeled dataset

(b) Unlabeled dataset

Fig. 2 The countries from which the most tweets about COVID-19 vaccines have emerged in D1 dataset



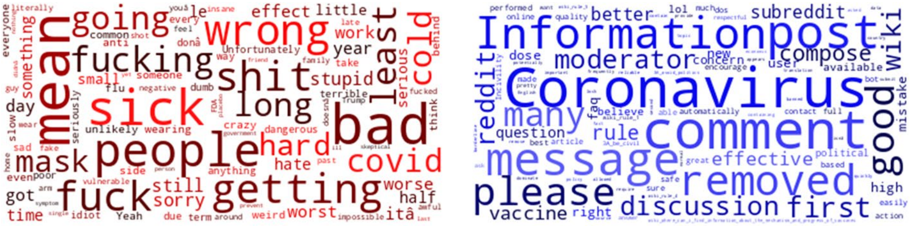
(a) Negative sentiment

(b) Positive sentiment

Fig. 3 Wordcloud visualization for D1 dataset

’effective’, ’vaccinated’, ’Thank’, ’best’, and ’happy’ in Fig. 3b have conveyed a positive opinion in regard to COVID-19 vaccines.

The second dataset which is referred to D2 dataset, has been developed by extracting Reddit comments from the subreddit Coronavirus. Each data record contains a comment collected from a Reddit post about COVID-19 vaccine news in ‘/r/Coronavirus’. Here, the Python Reddit API Wrapper was used for data collection. Further, preprocessing and cleaning of data were performed. The unlabeled D2 dataset, for our study, contains 19,201 Reddit comments, while similar to the procedure followed for the D1 dataset, manual labeling was conducted to obtain the labeled set. 4500 comments were labeled as positive and 1250 were labeled as negative, forming a labeled D2 dataset of size 5750 (as shown in Fig. 1b). Subsequently, for the D2 dataset, the Imbalance ratio is 3.6. The word cloud visualization for the D2 dataset is depicted in Fig. 4.



(a) Negative sentiment

(b) Positive sentiment

Fig. 4 Wordcloud visualization for D2 dataset

3 Deep adversarial active learning framework

In recent times, several DAL approaches have employed adversarial training for parameter optimization. It has been observed that through adversarial training, the model performance is much superior when compared to traditional AL methods. This is because, in the case of adversarial learning, it sets up a min-max game, permitting models to optimize their training parameters in a fully differentiable setting. The authors Shui et al. (2020) employed the Wasserstein distance and provided a hybrid query approach, balancing uncertainty and heterogeneity explicitly, and proposed Wasserstein Adversarial Active Learning (WAAL). Similarly, Adversarial Sampling for Active Learning (ASAL) (Mayer & Timofte, 2020) presented a novel sample selection process for multi-class problems, with an emphasis on obtaining high-quality synthetic samples. In Zhang et al. (2020), the state relabeling adversarial active learning model (SRAAL) was introduced that incorporated annotation as well as labeled/unlabeled state information to derive the most informative unlabeled data points. The authors in Liu et al. (2019) further worked on GAAL (Zhu & Bento, 2017) and proposed a novel Single-Objective Generative Adversarial Active Learning (SO-GAAL) framework for outlier detection. In another approach, Wang and Ren (2020) studied hyperspectral image classification with feature-oriented adversarial AL (FAAL) strategy, with the help of adversarially learned acquisition heuristic.

One of the most remarkable approaches that have been introduced in AL, in recent years, has been Variational Adversarial Active Learning (VAAL) (Sinha et al., 2019). The pictorial representation of the algorithm proposed by VAAL is shown in Fig. 5. The learner interacts with an oracle, which is a source of labels for unlabeled data in active learning (Hacohen et al., 2022), usually a human expert or a ground-truth dataset. It is queried with a set of unlabelled data to obtain the labels of the same. The number of unlabelled data samples that can be queried to the Oracle is limited by ‘Budget’. However, in real applications, a human oracle could be unreliable, or too costly to access, and therefore a surrogate model can be used to approximate the oracle’s behavior. Deep learning models have been used as surrogates in active learning setup for code verification (Stark et al., 2015), Text classification (Zhang et al., 2017). In text classification, labeling queries by human oracles are too expensive (Figueroa et al., 2012) which motivated the authors to employ a surrogate model to act as the oracle. Using a VAE, optimized with both reconstruction and adversarial losses, VAAL learns the distribution of labeled data in latent space. Unlabeled instances

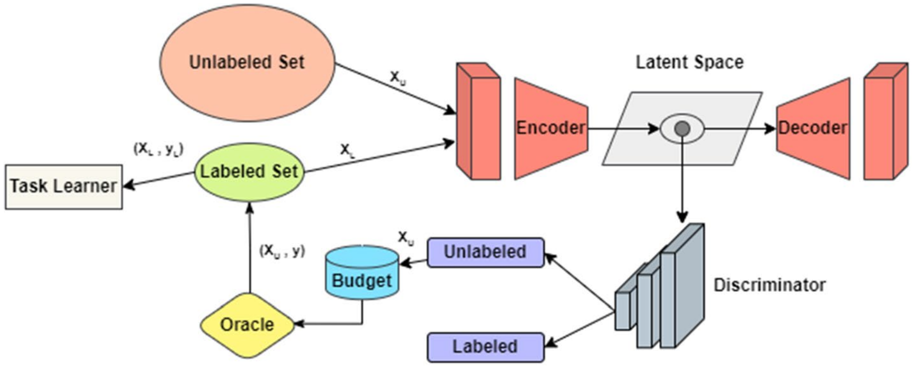


Fig. 5 Variational adversarial active learning

are predicted by a binary adversarial classifier (discriminator) and sent to an Oracle for annotation. The technique is task-agnostic since sample selection is done independently of the main-stream job of identifying data inputs. Furthermore, another method, task-aware variational adversarial AL (TA-VAAL) (Kim et al., 2021) is based on VAAL (Sinha et al., 2019) and Adversarial Representation AL (ARAL) (Mottaghi & Yeung, 2019), and takes into account a ranking loss prediction technique. Correspondingly, our proposed algorithm is greatly influenced by VAAL, where we studied the class distribution of the labeled set in latent space and introduced well-established resampling techniques with the aim of providing an unbiased classification on the binary COVID-19 vaccine sentiment dataset.

4 Imbalanced COVID-19 sentiment prediction

In the current article, randomly selected COVID-19 vaccine texts have been considered as an unlabeled set, represented as X_U, where the sentiment labels have been dropped off. The rest of the data samples have been included in the labeled set, denoted by (X_L, Y_L). The labeled set is imbalanced, where texts with positive sentiment are the majority class, while negative sentiment texts are the minority. According to recent literature (Aggarwal et al., 2020; Dong, 2021), it has been established that in case of extreme imbalance in the labeled set, the algorithm stays heavily biased towards the majority class and hence proves fatal to the entire intelligent system. However, there have been few efforts made to counteract the consequences of this problem. Thus, in the current study, it has become imperative to mitigate the effects of imbalanced classes. To achieve this goal, our model begins by learning the most effective compressed latent representation of the COVID-19 vaccine texts using both labeled and unlabeled data pools. Since the labeled set suffers from class imbalance, synthetic resampling strategies are implemented.

The proposed framework, illustrated in Fig. 6, learns a latent representation by mapping the labeled and unlabeled sets into the same latent space, similar to the methodology presented by VAAL (Sinha et al., 2019). The binary adversarial network is then constructed to distinguish one from the other. Bi-LSTM VAE and the discriminator have the structure of a two-player min-max game. The Bi-LSTM VAE is trained to learn a feature space in order to fool the adversarial network into believing that all data points from both the labeled and unlabeled sets are from the labeled pool. In contrast, the discriminator network learns how to tell the difference. As it is an adversarial training, the Bi-LSTM VAE is forced to come up with a better representation of the labeled data instances. This aforementioned strategy aids in the selection of the most informative unlabeled samples x_U for the query. It also helps in acknowledging the prevalent problem of selecting outliers for querying using Oracle. Next, the merged resampled and unlabeled latent sets are classified into labeled and unlabeled by using this binary adversarial network. The individual unlabeled data points, on acquiring labels from Oracle, denoted as (x^*, y^*), are then added to the resampled labeled set (X_L, Y_L). Once the updated balanced annotated set, represented as $(x_L, y_L) \cup (x^*, y^*)$, is obtained at the end of each iteration, it is then sent to the set of task-learners i.e classifiers for sentiment classification. As a consequence, the size of the annotated set increases after every loop, and also the overall quality of the set is improved.

To address the issue of uneven distribution, our RS-DAAL architecture ensures that adversarial training is free of class bias and offers the most informative cases for oracle annotation. Our method aims to increase classifier performance by using both labeled and chosen data samples collected by carefully querying an unlabeled data pool, as explained above. In this case, data instances from the unlabeled set are drawn with a limited budget. The loss function $\mathcal{L}_{VAE}$ for unlabeled and labeled data has been expressed as:

$$\begin{aligned}\mathcal{L}_{\mathrm{VAE}} = {} & \mathbb{E}\left[\log p_\theta\left(x_L \mid z_L\right)\right] - \mathrm{D}_{\mathrm{KL}}\left(q_\varphi\left(z_L \mid x_L\right) \| p(z)\right) \\ & + \mathbb{E}\left[\log p_\theta\left(x_U \mid z_U\right)\right] - \mathrm{D}_{\mathrm{KL}}\left(q_\varphi\left(z_U \mid x_U\right) \| p(z)\right)\end{aligned} \tag{1}$$

where the encoder and decoder are denoted by q_φ and p_θ, the Gaussian distribution is $p(z)$, and KL$(\cdot)$ is Kullback-Leibler distance. Here, the loss function seeks to reduce loss by maximizing the lower bound of the chance of creating authentic data points. The reparameterization trick, outlined below, has been applied to compute the gradients properly (Kingma & Welling, 2013).

$$\begin{aligned}\mathbf{z} &= \mu + \sigma \odot \epsilon \\ \epsilon &\sim \mathcal{N}(0, 1)\end{aligned} \tag{2}$$

where μ and σ imply mean and standard deviation, respectively, and $\odot$ signifies element-wise product. Firstly, by mapping the labeled and unlabeled data into the same feature space with identical probability distributions $q(z_L|x_L)$ and $q(z_U|x_U)$, the VAE fools the discriminator. Next, the adversarial network is trained to assign a binary label to the latent representation of $z_L \cup z_U$, which is 1 if the sample belongs to X_L and 0 if it does not. For the discriminator D training, the loss function $\mathcal{L}_D$ is as follows:

$$\begin{aligned}\mathcal{L}_{\mathrm{D}} = {} & -\mathbb{E}\left[\log D\left(q_\varphi\left(z_L \mid x_L\right)\right)\right] \\ & - \mathbb{E}\left[\log\left(1 - D\left(q_\varphi\left(z_U \mid x_U\right)\right)\right)\right]\end{aligned} \tag{3}$$

The following is the concept of the optimization algorithm:

$$\min_{q_\varphi}\max_{D}\mathbb{E}_{z_L\sim p_{x_L}}\left[\log\left(D\left(q_\varphi\left(z_L\mid x_L\right)\right)\right)\right] + \mathbb{E}_{z_U\sim p_{x_U}}\left[\log\left(1-D\left(q_\varphi\left(z_U\mid x_U\right)\right)\right)\right] \tag{4}$$

where q_φ is the VAE encoder and z_L and z_U are feature spaces for labeled and unlabeled data samples, respectively. In this expression, $z_L \sim p_{x_L}$ is denoted by $z_L = q_\varphi(z_L|x_L)$ with $x_L \sim p_{data}$ and $z_U \sim p_{x_U}$ is identical to $z_L \sim p_{x_L}$.

The discriminator is trained to differentiate between unlabeled and re-sampled labeled encoded vectors in feature space, as evident in Fig. 6. Unlike VAAL (Sinha et al., 2019), our method adds the data points, provided by Oracle, to the re-sampled set and not to the original labeled set, due to the fact that the original set is imbalanced and hence, will provide a skewed sentiment prediction.

The proposed resampling-assisted VAAL technique is outlined in algorithm 1. The primary training loop begins at step 3 when the model parameters are set up. It begins by selecting samples from labeled and unlabeled sets. In order to perform unsupervised training of the VAE with Gaussian priors, these two sets of data samples have been used. The loss function 1 gets minimized in an effort to train the Bi-LSTM VAE. The trained encoder is then used to generate latent space-compressed representations of the input COVID-19 tweets (line 7), where z_U and z_L indicate latent vectors corresponding to unlabeled and labeled sets, respectively. Lines 8–10 are overseeing the model's adversarial training. To overcome the inherent imbalanced distribution of the labeled set, the latent vectors belonging to the labeled set are resampled.

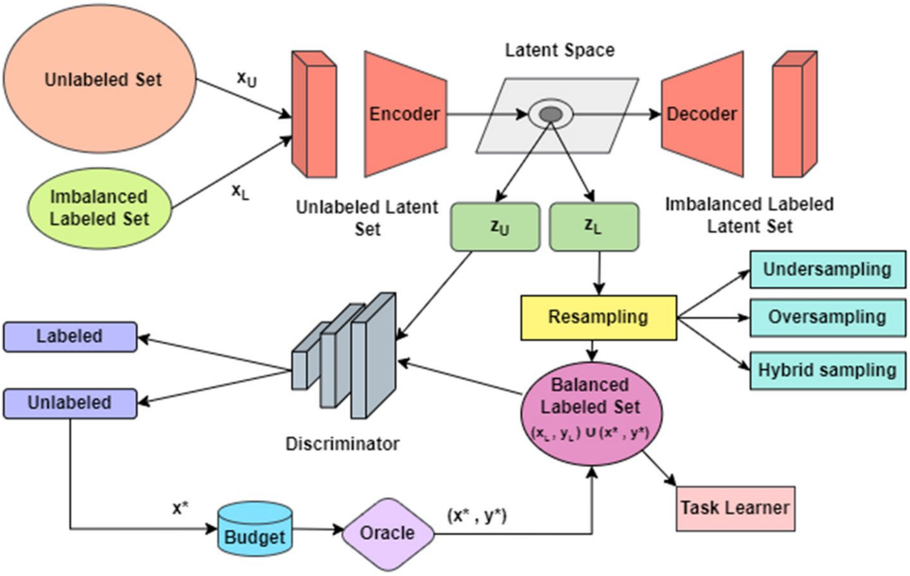


Fig. 6 Addressing imbalanced COVID-19 sentiment classification with synthetic resampling coupled variational adversarial active learning

Algorithm 1 Variational Adversarial Active Learning with Synthetic Resampling

```
Require: Labeled Data $(X_L, Y_L)$, Unlabeled Data $(X_U)$
Initialize $\theta_{VAE}, \theta_D$
repeat
    Draw sample $(x_L, y_L) \sim (X_L, Y_L)$
    Draw sample $x_U \sim X_U$
    Compute $\mathcal{L}_{VAE}$ using Equation 1
    Update $\theta_{VAE}$ by Stochastic Gradient Descent
    $z_U, z_L \leftarrow \theta_{VAE}\{x_U, x_L\}$
    $x^{'}, y^{'} \leftarrow Resample\{x_L, y_L\} \cup Oracle\{x^*\}$
    Compute $\mathcal{L}_D$ using Equation 3
    Update $\theta_D$ by Stochastic Gradient Descent
until $\theta_{VAE}, \theta_D$ Converges
```

The corresponding algorithm 2 describes the Oracle training for RS-DAAL. It is critical to train Oracle using a balanced dataset to prevent it from being biased towards the majority class. Initially, the complete labeled dataset (X_L, Y_L) is used to train a Multilayer Perceptron (MLP) model. In line 4, the selection of the top-k data samples is explained. This is accomplished by taking into account the output layer values of the MLP classifier. Top-k data instances are chosen from each class (X_k, Y_k). This balanced dataset is made out of the most confident data samples from the original labeled dataset. Lines 5–7 sum up the process of how the Oracle gets trained using X_k, Y_k.

Algorithm 2 RS-DAAL Oracle Training

```
Require: Labeled Data $(X_L, Y_L)$, k
1: Initialize $\theta_{MLP}$
2: Compute $\mathcal{L}_{MLP}$ using $(X_L, Y_L)$
3: Update $\theta_{MLP}$ using Adam
4: $X_k, Y_k \leftarrow$ Select top-$k$ most confident training samples from each class
5: Initialize $\theta_{Oracle}$
6: Compute $\mathcal{L}_{Oracle}$ using $(X_k, Y_k)$
7: Update $\theta_{Oracle}$ using Stochastic Gradient Descent
```

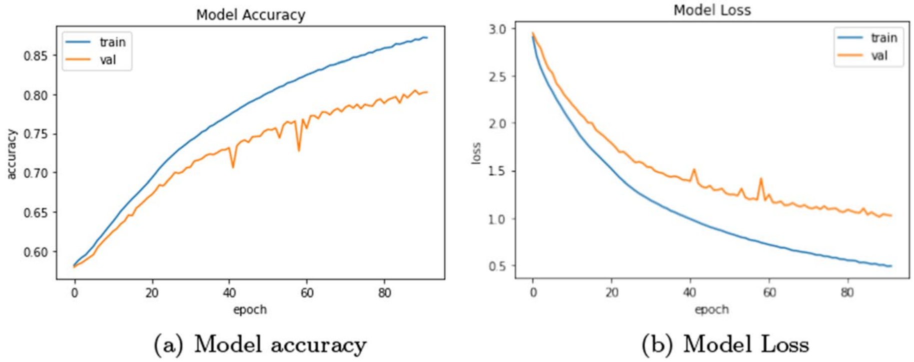
5 Experimental setup

The Bi-LSTM Variational Autoencoder, being an unsupervised learning method, has been trained on the combined labeled and unlabeled set. In order to study the semantic information in latent space, all text sequences have been padded, having the same length of 30, and passed to the model. The embedding dimension has been chosen to be 150, while the batch size is 100. The bidirectional LSTM layer has a memory unit of size 128, thereby the concatenated hidden state dimension is 256. The second Bi-LSTM layer has a memory unit of 64, hence the latent vectors obtained are of size 128. The ReLU activation function has been utilized for the Bi-LSTM layers. A dropout value of 0.4 and a learning rate of 0.001 has been selected. The validation split is 0.5. Additionally, Early stopping which monitors the validation loss, with a patience value of 3, has been implemented. Sparse categorical crossentropy as the loss function and Nesterov-accelerated Adaptive Moment Estimation, or Nadam (Dozat, 2016), as the model optimizer have been employed. Figure 7a and b have displayed the model accuracy and model loss, respectively. It can be observed that after 92 epochs, the training loss, training accuracy, validation loss, and validation accuracy have been 0.4892, 0.8725, 1.0255, and 0.8024, respectively.

The binary adversarial classifier, or discriminator, has been trained with the Encoder weights of the Bi-LSTM VAE, thereby enabling our approach to learn to differentiate between unlabeled and labeled data. The training process has been in a manner similar to that of a GAN. RMSprop with a learning rate of 5×10^{-4}., clip value of 1.0, and decay of 1e-8 have been employed. Besides, for model loss function, binary crossentropy has been used. The threshold for the discriminator prediction has been chosen to be 0.5.

The individual labeled and unlabeled sets are passed through the encoder model to produce the latent vectors once the Bi-LSTM VAE is trained. The acquired labeled latent vector set is, however, imbalanced and hence is re-sampled with 15 different resampling methods. These resampling techniques used in our method include (I) eight under-sampling techniques, viz., OSS (OSS), Neighbourhood Cleaning Rule (NCLR), Near Miss, Instance Hardness Threshold (IHT), Edited Nearest Neighbours (ENN), Repeated Edited Nearest Neighbours (RENN), All KNN, and Cluster Centroids (CC), (II) five over-sampling techniques, viz., Random Oversampler (ROS), SMOTE, ADASYN, Borderline SMOTE (BSMOTE), and SVM-SMOTE, and (III) two hybrid sampling techniques, viz., SMOTE-TOMEK, and SMOTE-ENN. The hyper-parameters chosen for the resampling techniques have been outlined in Table 1. The top-k most confident data instances from the resampled labeled set are given to an MLP classifier with ReLU activation function and Adam optimizer for oracle training. For our experiment, the value of *k* is set to 10,000.

With a budget of 50,000, 5000 data samples have been sent to Oracle for annotation which is then added to the initial balanced labeled set and training is repeated on the improved training set. We have selected a varied range of classifiers, viz. Decision Tree (DT), Random Forest (RF), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Logistic Regression (LR), Bernoulli Naıve Bayes (BNB), Light Gradient Boosted Machine (LGBM), and Multi-Layer Perceptron (MLP), with the help of which the performance of the updated labeled set is evaluated. The classifiers have been implemented using *scikit-learn* (Pedregosa et al., 2011) package. Tenfold cross-validation has been utilized with the



(a) Model accuracy (b) Model Loss

Fig. 7 Bi-LSTM variational autoencoder training for D1 dataset

Table 1 Hyperparameter grid for the synthetic resampling methods

Resampling methods	Hyperparameters used
OSS	Sampling strategy = 'auto' Number of seeds to extract to build set S = 1
NCLR	Sampling strategy = 'auto', Number of neighbors= 3 Strategy used to exclude samples in ENN sampling = 'all', threshold cleaning value= 0.5
NEAR MISS	Sampling strategy = 'auto', version =1 Number of neighbors= 3 Number of neighbors for subset selection= 3
IHT	Sampling strategy = 'auto', Number of folds= 5
ENN	Sampling strategy = 'auto', Number of neighbors= 3 Strategy used to exclude samples= 'all'
RENN	Sampling strategy = 'auto' Number of neighbors= 3, Maximum iterations= 100 Strategy used to exclude samples= 'all'
All KNN	Sampling strategy = 'auto', Number of neighbors= 3 Strategy used to exclude samples= 'all'
CC	Sampling strategy = 'auto', Voting strategy= 'auto'
ROS	Sampling strategy = 'auto'
SMOTE	Number of nearest neighbors = 5
BSMOTE	Number of nearest neighbors = 5 Number of nearest neighbors to determine if a minority sample is in danger= 10
SVM-SMOTE	Number of nearest neighbors = 5 Number of nearest neighbors to determine if a minority sample is in danger= 10, Step size when extrapolating = 0.5
ADASYN	Number of nearest neighbors = 5
SMOTE-TOMEK	Sampling strategy = 'auto'
SMOTE-ENN	Sampling strategy = 'auto'

aim of achieving concrete performance results. Each fold of the testing set contains both positive and negative tweet data. Hyperparameters that have been employed for the task learners, i.e classifiers have been presented in Table 2. In order to compare our proposed framework, we have tested our model with a baseline algorithm. VAAL (Sinha et al., 2019) has been fed with an imbalanced labeled set, and the performance of the task learners has been assessed with respect to our method. Our tests were implemented with an Intel Core i5-1035G1 CPU with Intel UHD Graphics 620, 8 GB RAM, Windows 10 Home 21H1, and TensorFlow 2.5.0.

6 Results and discussion

On the basis of F1-Score, Geometric Mean, Balanced Accuracy, and Area Under the Curve (AUC), we have recorded the performance of our proposed model.

Table 2 Hyperparameter grid for the classification algorithms

Classifiers	Hyperparameters used
DT	Criterion=‘entropy’, Maximum depth of the tree= 11
	Minimum number of samples required to be at a leaf node = 2
RF	Number of trees in the forest =100, Criterion=‘entropy’
KNN	Number of neighbors = 5, Weight function used in prediction = ‘uniform’, Leaf size = 30
SVM	Kernel coefficient for ‘rbf’, ‘poly’, and ‘sigmoid’ =‘auto’
LR	C =0.1, Maximum number of iterations =100000, Penalty=‘l1’
	Algorithm to use in the optimization problem =‘saga’
BNB	Additive smoothing parameter = 1.0
LGBM	Maximum tree leaves for base learners = 31, Type of boosting = ‘gbdt’
	Number of samples for constructing bins = 200000
MLP	Solver for weight optimization = ‘adam’, Learning rate = ‘constant’
	Initial learning rate = 0.001, Maximum number of iterations = 200
	Tolerance for the optimization = 1e-4

$$\text{F1-Score} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \tag{5}$$

$$\text{Geometric mean (GM)} = \sqrt{\frac{TP}{TP + FN} \cdot \frac{TN}{TN + FP}} \tag{6}$$

$$\text{Balanced Accuracy (BACC)} = \frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP}\right) \tag{7}$$

where *TP* is the number of positive tweets successfully categorized by the model, *TN* refers to the negative tweets classified as negative, *FP* is the number of negative tweets inaccurately recognized as positive, and *FN* is the number of positive tweets identified as negative tweets.

The results and analysis section is divided as follows:

1. We have conducted a thorough analysis of classifier performance, as outlined in Sect. 6.1, providing a detailed examination of their effectiveness.
2. In Sect. 6.2, we have delved into resampling approaches, offering a comprehensive overview of their impact on our framework.
3. Sect. 6.3 focuses on assessing the efficacy of our framework through varied resampling levels, showcasing its adaptability and robustness.
4. A comparative evaluation in Sect. 6.4 has positioned our model against an imbalanced baseline, highlighting its superiority in handling imbalance.
5. Statistical analysis, detailed in Sect. 6.5, demonstrates consistent improvement in AUC scores across 10 iterations, reinforcing the reliability of our framework.
6. Sect. 6.6 places our results in context by comparing them with state-of-the-art methods, underscoring the competitive edge of our approach in addressing the challenges posed by imbalanced datasets.

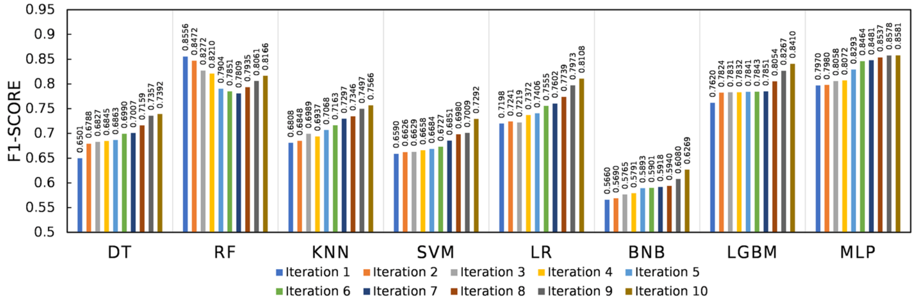


Fig. 8 Performance analysis in terms of F1-score for D1 dataset

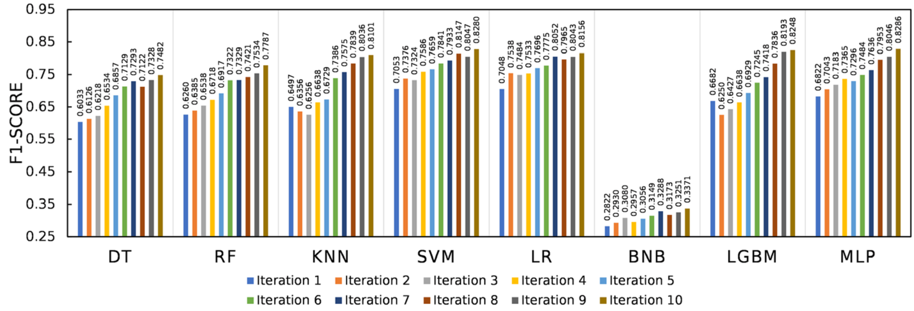


Fig. 9 Performance analysis in terms of F1-score for D2 dataset

6.1 Performance of classifiers

To examine the efficiency of RS-DAAL, over the course of 10 iterations, the values of three evaluation metrics, viz. F1-Score, Geometric mean (GM), and Balanced Accuracy (BACC) have been recorded for D1 and D2 datasets. Since it is an iterative process, after every insertion of new data samples, the constructed annotated dataset has been sent to the task learner, viz. DT, RF, KNN, SVM, LR, BNB, LGBM, and MLP, and the scores mentioned above have been calculated. Here, the datasets have been 100% resampled and the scores have been recorded.

Figures 8 and 9 have reported the F1-score behavior of the 8 classifiers over a span of 10 iterations for D1 and D2 datasets, respectively. It can be observed that the MLP classifier has exhibited the highest F1 scores for both datasets. At the same time, Figs. 10 and 11 illustrate the performance improvement of the classifiers for D1 and D2 datasets over 10 iterations. Existing research (Kuncheva et al., 2019) has proved that geometric mean, due to its multiplicative nature, is a distinctive evaluation metric indicative of the model's performance. Here, a similar trend is observed in these figures as well. Furthermore, Figs. 12 and 13 demonstrate the effectiveness of MLP, LGBM, and LR classifiers in classifying vaccine sentiment.

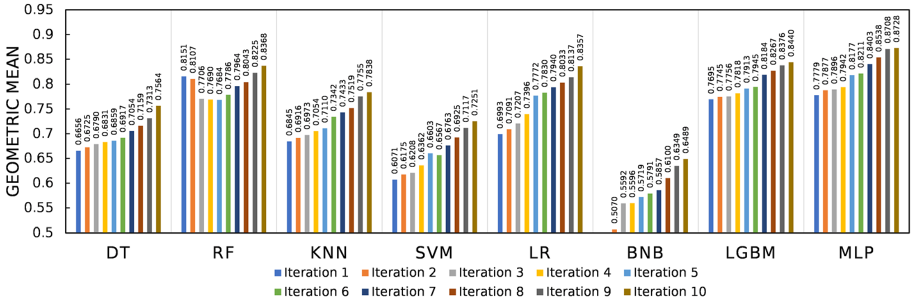


Fig. 10 Performance analysis in terms of Geometric mean for D1 dataset

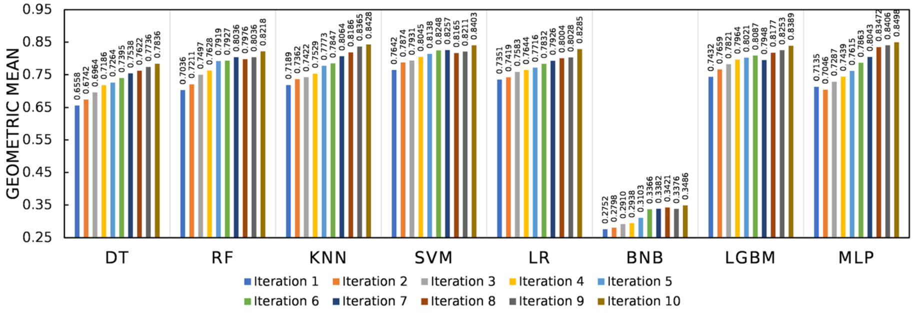


Fig. 11 Performance analysis in terms of Geometric mean for D2 dataset

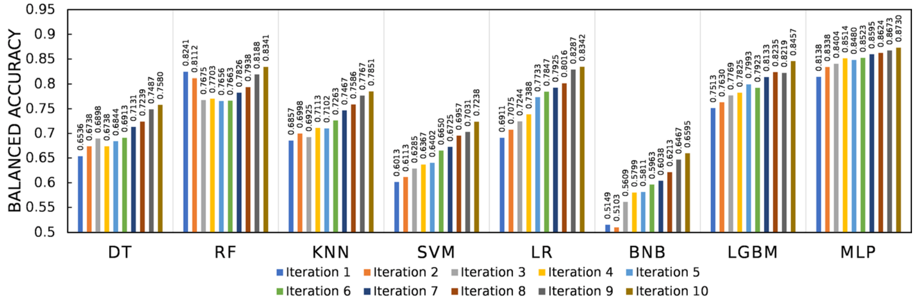


Fig. 12 Performance analysis in terms of balanced accuracy for D1 dataset

6.2 Performance of resampling techniques

On the basis of F1, Geometric mean (GM), and Balanced Accuracy (BACC), the performance of the resampling strategies has been evaluated and analyzed across 10 iterations for D1 and D2 datasets. Here, for each resampling technique, the mean of the classifier scores in terms of F1, GM, and BACC, have been noted. Similar to Sect. 6.1, the datasets have undergone 100% resampling.

Figures 14, 16, and 18 demonstrate the performance analysis of the resampling techniques in terms of F1, GM, and BACC for D1 dataset. At the same time, for the D2 dataset,

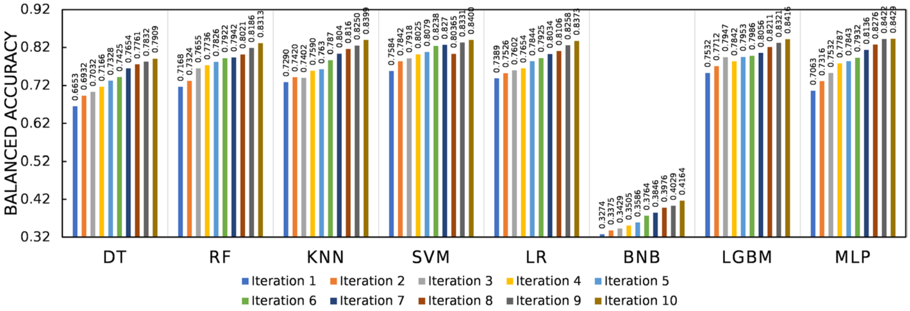


Fig. 13 Performance analysis in terms of balanced accuracy for D2 dataset

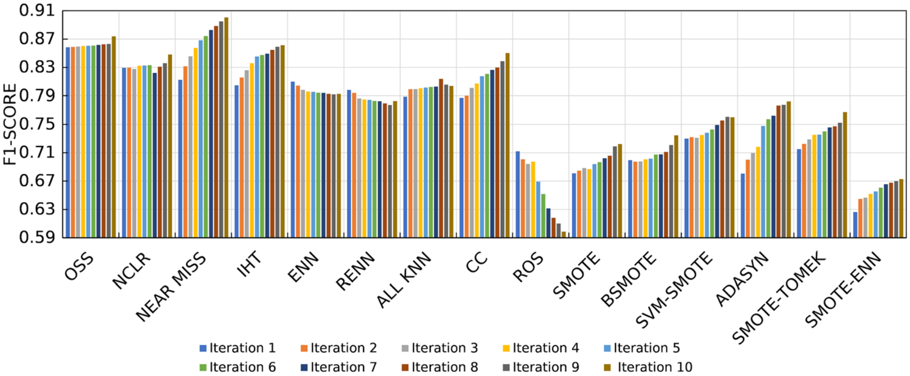


Fig. 14 Performance analysis based on F1-Score for D1 dataset

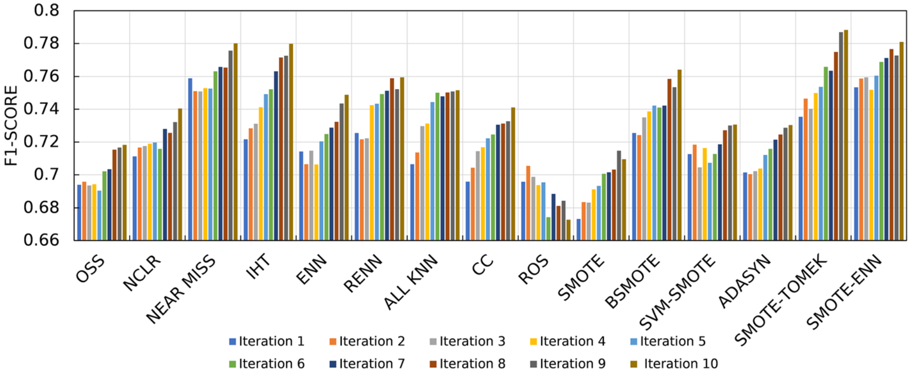


Fig. 15 Performance analysis based on F1-Score for D2 dataset

Figs. 15, 17, and 19 illustrate the performance evaluation with regards to F1, GM and BACC, respectively. IHT undersampling strategy has been the most optimal resampling strategy for the D1 dataset. Meanwhile, the performance of ROS has been poor. Overall, for the D1 dataset, the undersampling methods have excelled in showcasing noteworthy performance scores over a period of 10 iterations.

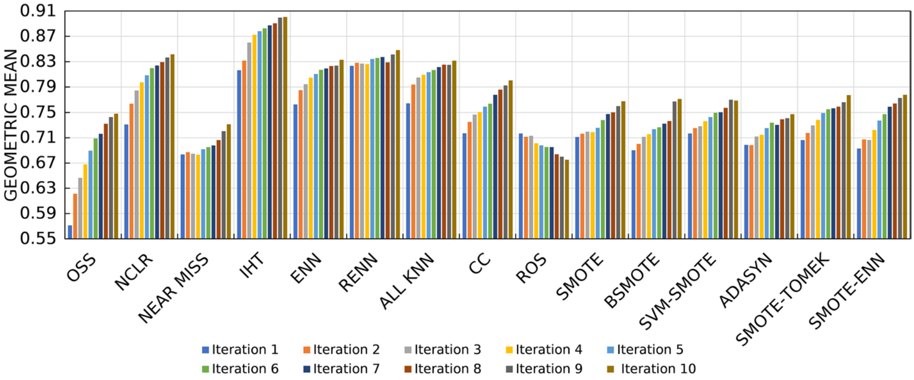


Fig. 16 Performance analysis based on geometric mean for D1 dataset

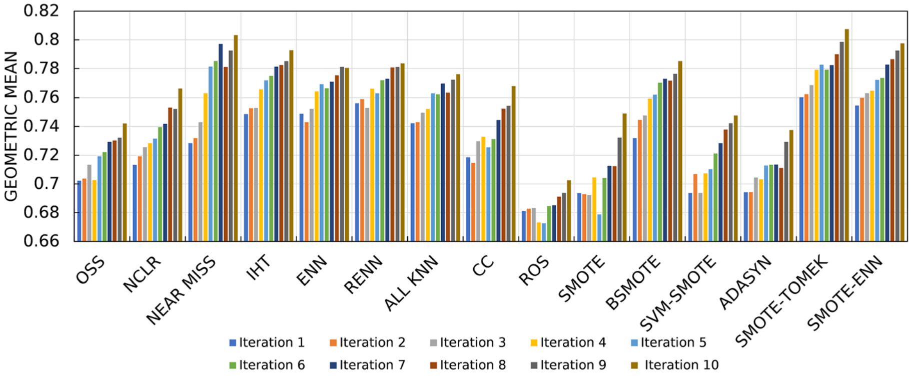


Fig. 17 Performance analysis based on geometric mean for D2 dataset

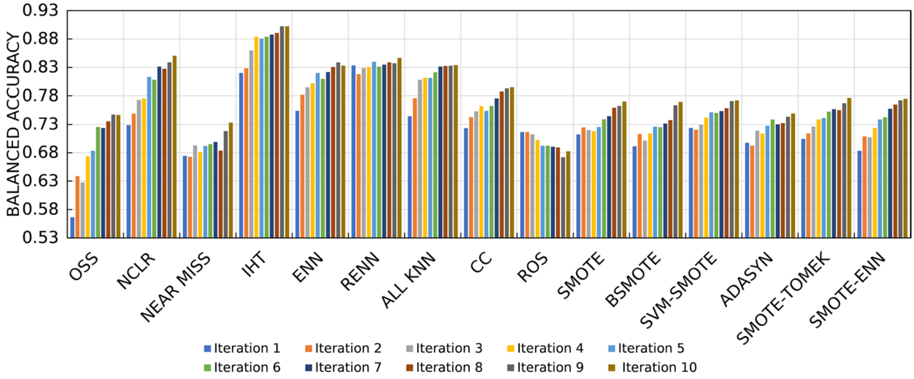


Fig. 18 Performance analysis based on balanced accuracy for D1 dataset

In the case of the D2 dataset, the hybrid-sampling techniques, viz. SMOTE-ENN and SMOTE-Tomek have performed quite remarkably. In terms of GM and BACC, SMOTE-Tomek has exhibited an increase of 6.19%, and 15.11%, respectively. As shown in Fig. 15,

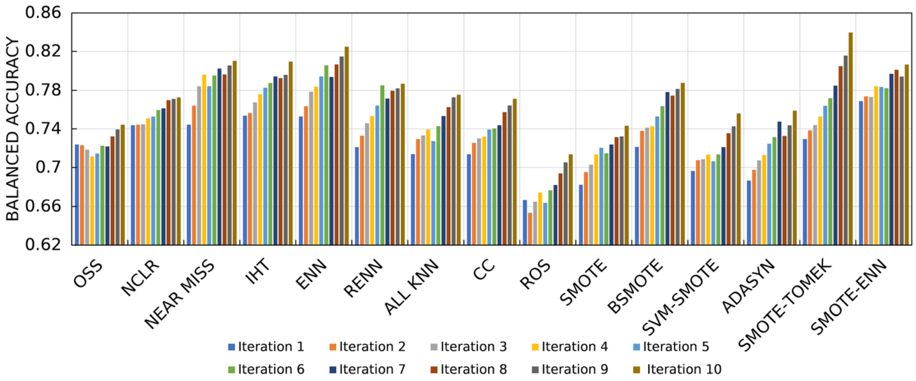


Fig. 19 Performance analysis based on balanced accuracy for D2 dataset

techniques like NEAR MISS, IHT, ENN, RENN, and BSMOTE have also shown notable performance rise in terms of F1-score. All in all, SMOTE-Tomek has been the most optimal resampling strategy for the D2 dataset over 10 iterations.

6.3 Performance comparison using different levels of resampling

This section elaborates on the results obtained by using the best-performing resampling strategies (as apprehended in Sect. 6.2) for D1 and D2 datasets with 4 different levels of resampling, viz. 20, 50, 70, and 100. Here, the MLP classifier has been used, with IHT undersampling for the D1 dataset and SMOTE-Tomek hybrid sampling for the D2 dataset, to classify the positive and negative vaccine sentiments. Table 3 illustrates the performance scores in terms of F1-score, Geometric mean, and Balanced Accuracy, for D1 and D2 datasets after 10th iteration (best scores are highlighted in bold). It can be observed that with an increase in the levels of resampling, the performance improves notably. After 10 iterations, 100% resampling delivers the most optimal performance for D1 and D2 datasets.

6.4 Performance comparison with baseline

Here, comparative research has been conducted with respect to the baseline model in order to validate the efficiency of the proposed framework. The individual performance measures have been examined and analyzed for Iteration 10. In relation to the baseline model, a labeled imbalanced set has been used, here, mentioned as "Imbalance" in the respective

Table 3 Performance comparison with different levels of resampling

Levels of resampling	D1			D2		
	F1-SCORE	GM	BACC	F1-SCORE	GM	BACC
20%	0.8847	0.8645	0.8718	0.827	0.8476	0.8502
50%	0.8943	0.8853	0.8995	0.8467	0.8698	0.8686
70%	0.9106	0.9148	0.9287	0.8632	0.8816	0.8907
100%	**0.9156**	**0.9312**	**0.9429**	**0.8703**	**0.891**	**0.8956**

tables. We have consciously preferred not to depend on the accuracy score when dealing with imbalanced data since it might be deceiving and lead to incorrect conclusions (Sourbier et al., 2022; Ren et al., 2022; Borowska & Stepaniuk, 2022; Cao et al., 2013).

Tables 4, 5, 6 reflect the performance with 100% resampling on comparison with the baseline model in terms of F1-Score, Geometric mean (GM), and Balanced Accuracy (BACC) for D1 and D2 datasets (best scores are highlighted in bold). The highest scores for each classifier in the D1 and D2 datasets have been highlighted in bold. In the case of IHT with MLP classifier in the D1 dataset, a 23.99% increment in F1-Score is observed when compared to the baseline. While for GM and BACC in the D1 dataset, MLP with IHT has demonstrated a 26.88% and 27.29% increase on comparing with baseline. In terms

Table 4 A comparative study in terms of F1-score

		DT	RF	KNN	SVM	LR	BNB	LGBM	MLP
IMBALANCE	D1	0.7113	**0.9198**	0.7187	0.6034	0.7137	0.5647	0.8296	0.7384
	D2	0.5638	0.5227	0.6632	0.7239	0.6922	0.2235	0.7896	0.7959
OSS	D1	0.8549	0.8801	0.8696	0.8914	0.9106	0.7491	0.9113	0.9233
	D2	0.6496	0.812	0.7867	0.7727	0.7838	0.3135	0.8084	0.8193
NCLR	D1	0.7857	0.8666	0.8885	0.8792	0.8707	0.7339	0.8655	0.8957
	D2	0.7478	0.7492	0.7671	0.7944	0.8274	0.3474	0.8426	0.8474
NEAR MISS	D1	**0.9233**	0.9077	**0.9238**	**0.9151**	**0.931**	0.7327	**0.9322**	**0.9376**
	D2	**0.7976**	**0.8345**	0.8434	0.8547	0.8527	0.3576	0.8466	0.8533
IHT	D1	0.8259	0.8595	0.8865	0.8647	0.8742	0.7835	0.8796	0.9156
	D2	0.7882	0.7937	0.8529	**0.8763**	0.8448	0.3813	0.8385	0.8624
ENN	D1	0.7496	0.8124	0.8094	0.719	0.8264	0.6683	0.8507	0.9044
	D2	0.7549	0.7539	0.8136	0.8457	0.8034	0.3649	0.8148	0.8396
RENN	D1	0.749	0.8117	0.8305	0.7067	0.8175	0.5846	0.8585	0.9051
	D2	0.7686	0.7725	0.8337	0.8654	0.8129	0.3671	0.8211	0.8348
ALL KNN	D1	0.7614	0.832	0.8359	0.7329	0.8631	0.6223	0.8697	0.9125
	D2	0.7423	0.8056	0.8379	0.8438	0.7932	0.3576	0.8389	0.7937
CC	D1	0.8402	0.8625	0.8534	0.8569	0.8974	0.6587	0.9085	0.9235
	D2	0.6934	0.7844	0.783	0.8563	0.8026	0.3462	0.8448	0.8178
ROS	D1	0.5401	0.6842	0.6237	0.4682	0.6024	0.5738	0.6538	0.6415
	D2	0.6257	0.7256	0.7129	0.7682	0.7239	0.2764	0.7954	0.7539
SMOTE	D1	0.6926	0.7907	0.6436	0.7689	0.7812	0.4699	0.7999	0.8307
	D2	0.7468	0.6968	0.7802	0.8155	0.8011	0.2452	0.7731	0.8173
BSMOTE	D1	0.7068	0.8156	0.5967	0.6809	0.7689	0.6534	0.8143	0.8398
	D2	0.7877	0.8053	0.8278	0.8413	0.8425	0.3305	0.8286	0.8497
SVM-SMOTE	D1	0.6819	0.8273	0.6881	0.7646	0.7889	0.6464	0.8475	0.8362
	D2	0.7824	0.7308	0.8135	0.7755	0.8375	0.3034	0.8059	0.7952
ADASYN	D1	0.7557	0.7887	0.6766	0.7988	0.7865	**0.7922**	0.8373	0.8239
	D2	0.7729	0.7856	0.7865	0.7869	0.7938	0.2964	0.7976	0.8224
SMOTE-TOMEK	D1	0.7028	0.8272	0.6566	0.7924	0.8279	0.5932	0.8835	0.8549
	D2	0.7681	0.8178	**0.8666**	0.8664	**0.8638**	**0.397**	0.8559	**0.8703**
SMOTE-ENN	D1	0.6698	0.7897	0.7961	0.3904	0.7656	0.3562	0.792	0.8233
	D2	**0.7963**	0.8128	0.846	0.8572	0.8513	0.3717	**0.8596**	0.8524

Table 5 A comparative study in terms of geometric mean

		DT	RF	KNN	SVM	LR	BNB	LGBM	MLP
IMBALANCE	D1	0.5141	0.58	0.546	0.2369	0.5051	0.5742	0.6189	0.7339
	D2	0.5934	0.5811	0.6725	0.5483	0.6177	0.2758	0.7932	0.7853
OSS	D1	0.7162	0.7678	0.7629	0.6572	0.8164	0.6041	0.8062	0.8546
	D2	0.6842	0.8417	0.8132	0.8524	0.7538	0.3368	0.8202	0.8341
NCLR	D1	0.806	0.8512	0.8765	0.7905	0.866	0.7467	0.8888	0.9063
	D2	0.7961	0.8166	0.8157	0.8263	0.8219	0.3563	0.8328	0.8633
NEAR MISS	D1	0.7367	0.817	0.7166	0.5659	0.8324	0.641	0.7498	0.791
	D2	**0.8439**	0.8548	0.8765	0.8578	0.8728	0.3742	0.8699	0.876
IHT	D1	**0.8801**	**0.916**	0.8868	**0.9019**	**0.9131**	**0.8576**	**0.9202**	**0.9312**
	D2	0.7705	0.815	0.8436	0.8689	0.8632	**0.4415**	0.8736	0.866
ENN	D1	0.7927	0.8553	0.842	0.7957	0.8535	0.7208	0.8774	0.9245
	D2	0.8017	0.8404	0.8666	0.8544	0.8076	0.3732	0.852	0.8478
RENN	D1	0.8146	0.8528	0.8827	0.7948	0.8547	0.7473	0.9129	0.9259
	D2	0.8031	0.8367	0.8728	0.8748	0.8027	0.3744	0.8538	0.8507
ALL KNN	D1	0.7929	0.8538	0.8774	0.7687	0.8801	0.6691	0.8868	0.9238
	D2	0.8124	0.8356	0.8432	0.8645	0.8439	0.3769	0.8038	0.8286
CC	D1	0.7715	0.8431	0.7701	0.7801	0.8754	0.557	0.8873	0.9176
	D2	0.7452	0.8122	0.8239	0.8729	0.8326	0.3587	0.8532	0.8435
ROS	D1	0.6366	0.7617	0.7073	0.5622	0.6899	0.601	0.7275	0.7151
	D2	0.6888	0.7627	0.7916	0.7433	0.7633	0.2538	0.8122	0.8049
SMOTE	D1	0.7097	0.8685	0.6915	0.7853	0.8405	0.5421	0.8595	0.8445
	D2	0.7936	0.7653	0.8027	0.8329	0.8216	0.3053	0.8319	0.8386
BSMOTE	D1	0.7612	0.8493	0.6424	0.7567	0.8178	0.6304	0.8494	0.8638
	D2	0.8035	0.8496	0.8649	0.8626	0.8628	0.3628	0.8123	0.8632
SVM-SMOTE	D1	0.7393	0.8272	0.7125	0.7951	0.8116	0.5843	0.8276	0.853
	D2	0.8126	0.7928	0.8322	0.8027	0.8264	0.2976	0.8018	0.8146
ADASYN	D1	0.6867	0.7861	**0.8901**	0.5407	0.854	0.6132	0.7656	0.8417
	D2	0.7524	0.7933	0.8165	0.7635	0.8172	0.3027	0.8124	0.842
SMOTE-TOMEK	D1	0.7326	0.8614	0.7017	0.7863	0.8355	0.5492	0.8584	0.8904
	D2	0.8339	**0.8653**	0.8882	**0.8853**	**0.8833**	0.3314	**0.881**	**0.891**
SMOTE-ENN	D1	0.7694	0.8405	0.7968	0.5958	0.7949	0.6703	0.8428	0.9091
	D2	0.8128	0.8444	**0.8904**	0.8421	0.8537	0.3828	0.8724	0.8823

of GM in the D1 dataset, ENN, RENN, and ALL KNN have also documented notable performance gains. For the D2 dataset, SMOTE-Tomek with MLP has recorded a 13.46% 15.8% in terms of GM and BACC in comparison to the baseline with the imbalanced set. With respect to BACC, NEAR MISS, SMOTE-ENN, NCLR have also reported remarkable performance increments at the end of 10 iterations.

6.5 Statistical analysis

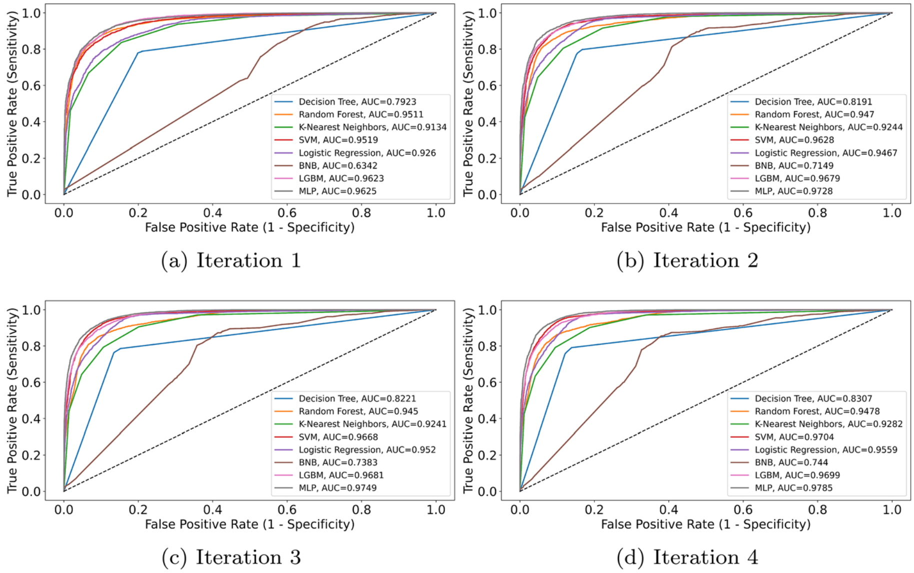
For an extensive and meticulous analysis, the trade-off between True Positive Rate (or Sensitivity) and False Positive Rate (1-Specificity) has been measured with the help of ROC

Table 6 A comparative study in terms of balanced accuracy

		DT	RF	KNN	SVM	LR	BNB	LGBM	MLP
IMBALANCE	D1	0.5108	0.5767	0.5348	0.2429	0.5126	0.5698	0.6272	0.7407
	D2	0.5825	0.5632	0.6853	0.5601	0.6522	0.3322	0.7846	0.7734
OSS	D1	0.7086	0.7543	0.7676	0.6653	0.8195	0.5935	0.8134	0.8506
	D2	0.6996	0.8643	0.8328	0.8352	0.7326	0.3428	0.8235	0.8236
NCLR	D1	0.8045	0.8538	0.8729	0.7986	0.8726	0.7744	0.8984	0.9285
	D2	0.7947	0.8248	0.8459	0.8331	0.8205	0.3592	0.8301	0.8732
NEAR MISS	D1	0.7439	0.8059	0.721	0.5736	0.8074	0.6621	0.7696	0.7825
	D2	**0.8523**	0.8644	0.8636	0.8611	0.8864	0.4122	0.8588	0.8829
IHT	D1	**0.8738**	**0.9037**	**0.8949**	**0.9065**	**0.9086**	**0.8642**	**0.9244**	**0.9429**
	D2	0.7706	0.8151	0.8448	0.8689	0.8632	0.574	0.8738	0.8661
ENN	D1	0.7947	0.8429	0.8398	0.7865	0.8425	0.7389	0.8974	0.9214
	D2	0.8031	0.8423	0.8606	0.8474	0.8233	**0.7539**	0.8438	0.8239
RENN	D1	0.8163	0.8589	0.8739	0.7985	0.8487	0.7528	0.9086	0.9143
	D2	0.8054	0.8387	0.8699	0.8544	0.8154	0.4237	0.8571	0.8302
ALL KNN	D1	0.7996	0.8586	0.8829	0.7742	0.8722	0.6729	0.8875	0.9247
	D2	0.8226	0.8362	0.8221	0.8543	0.8422	0.4056	0.7966	0.8244
CC	D1	0.7637	0.8395	0.7496	0.784	0.8794	0.5753	0.8635	0.9048
	D2	0.7535	0.8221	0.8273	0.8703	0.8378	0.3627	0.8512	0.8457
ROS	D1	0.6496	0.7486	0.7175	0.5646	0.6974	0.6138	0.7295	0.7385
	D2	0.7129	0.7855	0.7935	0.7457	0.8064	0.2849	0.7864	0.7935
SMOTE	D1	0.7154	0.8751	0.6853	0.7742	0.8564	0.5539	0.8595	0.8445
	D2	0.7632	0.7773	0.7923	0.8215	0.8528	0.2948	0.8422	0.8014
BSMOTE	D1	0.7782	0.8503	0.6375	0.7428	0.8293	0.6296	0.8375	0.8529
	D2	0.8148	0.8429	0.8456	0.8633	0.8764	0.3777	0.8328	0.8489
SVM-SMOTE	D1	0.7428	0.8364	0.7195	0.7985	0.8073	0.5928	0.8395	0.8429
	D2	0.8244	0.8032	0.8207	0.8165	0.8312	0.3155	0.8245	0.8117
ADASYN	D1	0.6687	0.7879	0.8857	0.5503	0.8497	0.6352	0.7556	0.8597
	D2	0.7952	0.8374	0.8122	0.7928	0.8122	0.3327	0.8465	0.8426
SMOTE-TOMEK	D1	0.7486	0.8534	0.7297	0.7486	0.8357	0.5486	0.8574	0.8914
	D2	0.8393	**0.8695**	**0.8935**	**0.8905**	**0.8887**	0.5538	**0.886**	**0.8956**
SMOTE-ENN	D1	0.7617	0.8423	0.7987	0.5911	0.7858	0.6849	0.843	0.8947
	D2	0.8123	0.8458	0.8743	0.8453	0.8697	0.4522	0.8714	0.8803

curves. Here, for D1 and D2 datasets, we have chosen IHT undersampling and SMOTE-Tomek hybrid sampling techniques and examined the improvement of the AUC scores over 10 iterations. From our analysis, it has been observed that IHT and SMOTE-Tomek have been most successful in alleviating the class imbalance problem in the text corpus in each dataset. In implementing our proposed Deep AL model, we have studied how the addition of Oracle-based selected data to the balanced labeled set has enhanced the performance of the classification algorithms.

Figures 20, 21, and 22 have demonstrated how the respective classifiers have been capable of distinguishing between positive and negative vaccine tweets over a span of 10 iterations for D1 dataset. AUC, being the degree of separability, has assisted us in interpreting the model performance in terms of the binary classification task. Similarly,



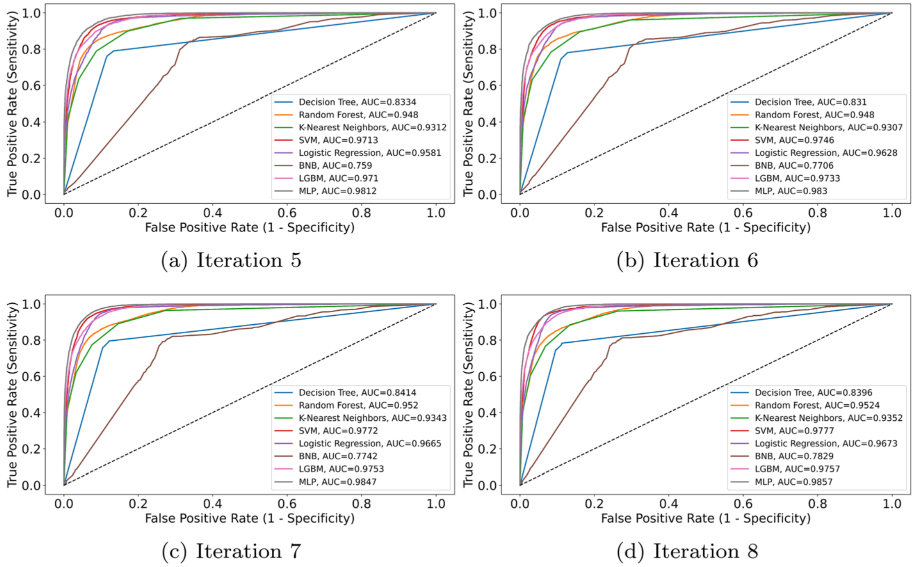
(a) Iteration 1

(b) Iteration 2

(c) Iteration 3

(d) Iteration 4

Fig. 20 Analysis of ROC curves over 10 iterations in D1 dataset



(a) Iteration 5

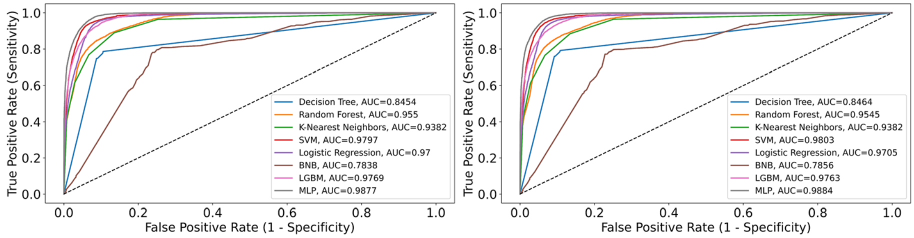
(b) Iteration 6

(c) Iteration 7

(d) Iteration 8

Fig. 21 Analysis of ROC curves over 10 iterations in D1 dataset

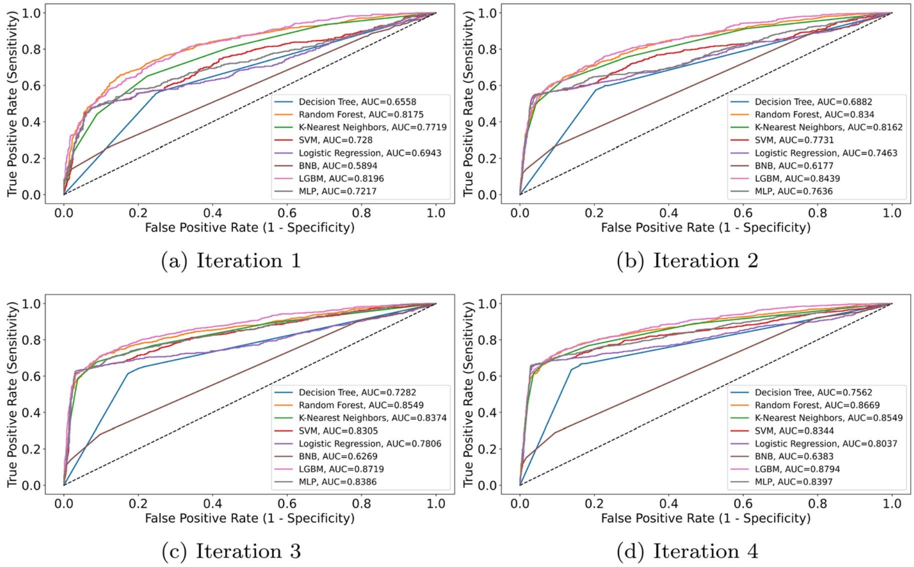
Figs. 23, 24, and 25 illustrate the ROC curves for D2 dataset. A noteworthy increment is observed for both datasets over 10 iterations. In particular, a consistent increase in the AUC score from Iteration 1 to 10 for the MLP classifier has been observed for D1 and D2 datasets. At the end of Iteration 10, MLP with IHT and MLP with SMOTE-Tomek for datasets D1 and D2 respectively have recorded the highest AUC scores. While the



(a) Iteration 9

(b) Iteration 10

Fig. 22 Analysis of ROC curves over 10 iterations in D1 dataset



(a) Iteration 1

(b) Iteration 2

(c) Iteration 3

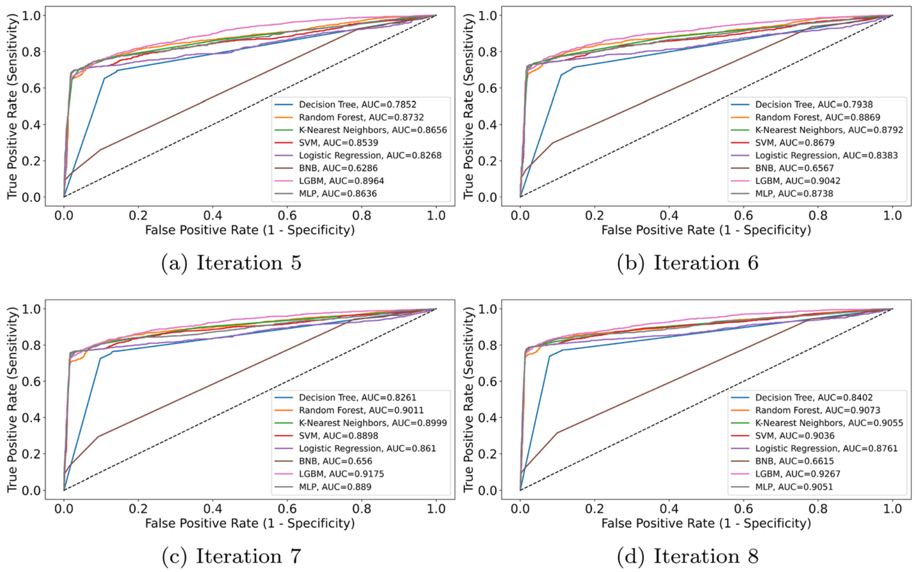
(d) Iteration 4

Fig. 23 Analysis of ROC curves over 10 iterations in D2 dataset

BNB classifier has showcased the lowest AUC scores for D1 and D2 datasets after 10 iterations.

6.6 Comparison with state-of-the-art

In this section, we compare our proposed architecture with existing state-of-the-art methods on the basis of F1-score, and the same is reported in Table 7. All comparisons are reported for dataset ‘D1’. One primary reason for this choice is the substantial disparity in the sample sizes between the datasets ‘D1’ and ‘D2’. Dataset ‘D1’ boasts a significantly larger number of samples compared to Dataset ‘D2’. This difference in sample sizes can significantly impact the statistical robustness and generalizability of the results. Hence, in order to maintain a fair and consistent evaluation and to ensure statistical significance, we concentrated our state-of-the-art comparisons on the larger Dataset ‘D1’. Here, the best-obtained results using RS-DAAL with IHT and MLP for the D1 dataset have been



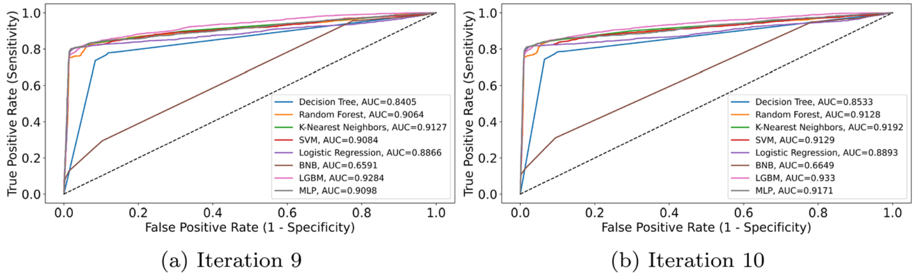
(a) Iteration 5

(b) Iteration 6

(c) Iteration 7

(d) Iteration 8

Fig. 24 Analysis of ROC curves over 10 iterations in D2 dataset



(a) Iteration 9

(b) Iteration 10

Fig. 25 Analysis of ROC curves over 10 iterations in D2 dataset

Table 7 Comparison of proposed work with existing state-of-the-art results

	F1-score
Akpatsa et al. (2022)	0.835
To et al. (2021)	0.455
Yue et al. (2022)	0.8173
Kunneman et al. (2020)	0.36
Jingcheng et al. (2017)	0.7442
Proposed method	0.9156

considered for comparison. The authors in Akpatsa et al. (2022) and Kunneman et al. (2020) have employed support vector machines to analyze COVID-19 vaccine tweets. In To et al. (2021), a Bi-LSTM model is trained to identify anti-vaccination attitudes generated during the COVID-19 pandemic. The support vector machine-based hierarchical classification model with optimized feature sets and model parameters is implemented in Jingcheng et al. (2017). Lastly, a contrastive adaptation network for early misinformation detection algorithm has been used in Yue et al. (2022). It can be observed that our proposed approach has successfully outperformed existing state-of-the-art methods.

6.7 Summary of results analysis

Overall, the results can be summarized as follows; In Sect. 6.1, MLP is the best-performing classifier in terms of the performance measures. Section 6.2 demonstrates that IHT undersampling is the most optimal sampling strategy for the D1 dataset and the hybrid sampling strategies, i.e., SMOTE-Tomek and SMOTE-ENN, have exhibited the highest results in the case of D2 dataset. As shown in Sect. 6.3, 100% resampling is the most optimal level of resampling in terms of F1-score, Geometric mean, and Balanced Accuracy. Section 6.4 indicates that IHT undersampling for the D1 dataset and SMOTE-Tomek for the D2 dataset show remarkable performance improvement over the baseline model, in the case of each classifier algorithm. Post Iteration 10 in Sect. 6.5, MLP with IHT secured the highest AUC score for D1, while MLP with SMOTE-Tomek led in the D2 dataset. Lastly, Sect. 6.6 highlights that our proposed methodology in comparison to various state-of-the-art methods have outperformed significantly.

Further, we calculate the inter-rater agreement scores in terms of Cohen's Kappa statistic of human annotators for both datasets and report it in the 'Human Annotator' columns of Table 8. Here, a correlation analysis is conducted between the inter-rater agreement scores and the performance of the best-performing configuration (MLP) of the proposed RS-DAAL model in terms of F1-Score, Geometric mean, and Balanced Accuracy for both datasets. The correlation coefficients obtained for the performance metrics are 0.8678, 0.8252, and 0.9929 respectively. It indicates a positive relationship between inter-rater agreement scores (Cohen's kappa) and classifier performance. The higher the agreement among human annotators, the better the classifier tends to perform.

7 Conclusion

The current study has proposed a new deep active learning framework by incorporating the latent space resampling method to mitigate the imbalanced class problem for efficient detection of COVID-19 vaccine sentiment prediction. The proposed RS-DAAL method has utilized a large amount of unlabeled data along with a small labeled dataset. Eventually, most informative unlabeled samples have been picked up by training the VAE and discriminator in an adversarial fashion. To mitigate the effect of the imbalanced labeled dataset, a

Table 8 Comparison between human annotator and our proposed method

	F1-score		Geometric mean		Balanced accuracy	
	Human annotator	Proposed method	Human annotator	Proposed method	Human annotator	Proposed method
D1	0.9824	0.9376	0.9739	0.9312	0.9907	0.9429
D2	0.9043	0.8703	0.8853	0.891	0.8934	0.8956

new resampling phase has been introduced inside the adversarial training. Experimental studies have revealed that the RS-DAAL model has been able to improve task learner performance after every iteration as the training progressed indicating an enhancement in the quality of the labeled data pool. Results have indicated that undersampling methods can be better suited to mitigating imbalanced class problems in the context of the current study. Resampling methods such as IHT, ENN, RENN, and ALL KNN have performed significantly well for all task learners. In addition, a comparative study with the baseline VAAL method has established that the proposed RS-DAAL method is better equipped to identify COVID-19 vaccine sentiment from social media platforms. To establish the improvement of classifiers in predicting the sentiment labels, a wide variety of task learners have been thoroughly investigated. It has been found that the performance of the MLP classifier has been improved to a greater extent compared to other classifiers in the current study.

However, the current study is limited to investigating the proposed model in the context of COVID-19 sentiment classification only. Future studies could investigate the possibility of applying the model in other text classification tasks where an abundant amount of unlabelled data is available. The computational complexity involved in training the proposed model could also be improved in future studies.

Nevertheless, future studies can be directed toward developing more trustworthy COVID-19 sentiment prediction models to help healthcare workers in the endeavor to vaccinate the population as swiftly and comprehensively as possible. The proposed framework can be extended to handle multilingual social media posts, enabling a comprehensive understanding of global vaccine sentiment across different languages and regions. In addition, the possibility of using other types of surrogate models as the oracle can also be investigated.

Author Contributions Conceptualization: SC, Methodology: SC, Formal analysis and investigation: SB, SC, Writing: SB, SC, Writing—review and editing: SC, SB, AKD, Supervision: AKD, SB

Funding The current study is not supported financially by any organization or individual.

Declarations

Conflict of interest The authors declare that there is no Conflict of interest.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Availability of data and material Original Data is available at 'https://www.kaggle.com/datasets/gpreda/all-covid19-vaccines-tweets'. Annotated data will be made available upon request to the Corresponding Author.

Code availability Will be made available upon request to the Corresponding Author.

References

Abdelwahab, M., & Busso, C. (2019). Active learning for speech emotion recognition using deep neural network. In *2019 8th International conference on affective computing and intelligent interaction (ACII)* (pp. 1–7). IEEE.

Aggarwal, U., Popescu, A., & Hudelot, C. (2020). Active learning for imbalanced datasets. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision* (pp. 1428–1437).

Akpatsa, S. K., Li, Xiaoyu, L., Hang, & Obeng, V.-H. K. S. (2022). Evaluating public sentiment of covid-19 vaccine tweets using machine learning techniques. *Informatica 46*(1).

Al-Hajri, S., Al-Kuwari, M. G., & Al-Thani, M. H. (2021). The covid-19 vaccine social media challenge: Strategies for addressing vaccine hesitancy in the age of misinformation. *Vaccine, 39*(29), 3859–3861.

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

Alamoodi, A. H., Zaidan, B. B., Al-Masawa, M., Taresh, S. M., Noman, S., Ahmaro, I. Y. Y., Garfan, S., Chen, J., Ahmed, M. A., Zaidan, A. A., et al. (2021a). Multi-perspectives systematic review on the applications of sentiment analysis for vaccine hesitancy. *Computers in Biology and Medicine, 139*, 104957.

Alamoodi, A. H., Zaidan, B. B., Zaidan, A. A., Albahri, O. S., Mohammed, K. I., Malik, R. Q., Almahdi, E. M., Chyad, M. A., Tareq, Z., Albahri, A. S., et al. (2021b). Sentiment analysis and its applications in fighting covid-19 and infectious diseases: A systematic review. *Expert Systems with Applications, 167*, 114155.

Alanazi, N. (2021). Opinion mining challenges and case study: Using twitter for sentiment analysis towards Pfizer/BioNTech, Moderna, AstraZeneca/Oxford, and Sputnik COVID-19 Vaccines. Ph.D. thesis, Lamar University-Beaumont.

Amjad, A., Qaiser, S., Anwar, A., Ali, R., et al. (2021). Analysing public sentiments regarding covid-19 vaccines: A sentiment analysis approach. In *2021 IEEE international smart cities conference (ISC2)* (pp. 1–7). IEEE.

Ash, J. T., Zhang, C., Krishnamurthy, A., Langford, J., & Agarwal, A. (2019). Deep batch active learning by diverse, uncertain gradient lower bounds. arXiv:1906.03671

Bashar, M. A., & Nayak, R. (2021). Active learning for effectively fine-tuning transfer learning to downstream task. *ACM Transactions on Intelligent Systems and Technology (TIST), 12*(2), 1–24.

Basiri, M. E., Nemati, S., Abdar, M., Asadi, S., & Rajendra Acharrya, U. (2021). A novel fusion-based deep learning model for sentiment analysis of covid-19 tweets. *Knowledge-Based Systems, 228*, 107242.

Beck, N., Sivasubramanian, D., Dani, A., Ramakrishnan, G., & Iyer, R. (2021). Effective evaluation of deep active learning on image classification tasks. arXiv:2106.15324

Bhoj, N., Khari, M., & Pandey, B. (2021). Improved identification of negative tweets related to covid-19 vaccination by mitigating class imbalance. In *2021 13th International conference on computational intelligence and communication networks (CICN)* (pp. 23–28). IEEE.

Borowska, K., & Stepaniuk, J. (2022). Rough-granular approach in imbalanced bankruptcy data analysis. *Procedia Computer Science, 207*, 1832–1841.

Cao, P., Zhao, D., & Zaiane, O. R. (2013). An optimized cost-sensitive svm for imbalanced data learning. In *Advances in knowledge discovery and data mining: 17th Pacific-Asia conference, PAKDD 2013, Gold Coast, Australia, April 14–17, 2013, proceedings, Part II 17* (pp. 280–292). Springer.

Chakraborty, K., Bhatia, S., Bhattacharyya, S., Platos, J., Bag, R., & Hassanien, A. E. (2020). Sentiment analysis of covid-19 tweets by deep learning classifiers—a study to show how popularity is affecting accuracy in social media. *Applied Soft Computing, 97*, 106754.

Dash, A., Gamboa, J. C. B., Ahmed, S., Liwicki, M., & Afzal, M. Z. (2017). Tac-gan-text conditioned auxiliary classifier generative adversarial network. arXiv:1703.06412.

Dhiman, G., Vignesh Kumar, A., Nirmalan, R., Sujitha, S., Srihari, K., Yuvaraj, N., Arulprakash, P., & Arshath Raja, R. (2023). Multi-modal active learning with deep reinforcement learning for target feature extraction in multi-media image processing applications. *Multimedia Tools and Applications, 82*(4), 5343–5367.

Dong, S. (2021). Multi class svm algorithm with active learning for network traffic classification. *Expert Systems with Applications, 176*, 114885.

Dor, L. E., Halfon, A., Gera, A., Shnarch, E., Dankin, L., Choshen, L., Danilevsky, M., Aharonov, R., Katz, Y., & Slonim, N. (2020). Active learning for bert: An empirical study. In *Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)* (pp. 7949–7962).

Dozat, T. (2016). Incorporating nesterov momentum into adam.

Du, J., Jun, X., Song, H., Liu, X., & Tao, C. (2017). Optimization on machine learning based approaches for sentiment analysis on hpv vaccines related tweets. *Journal of Biomedical Semantics, 8*(1), 1–7.

Figueroa, R. L., Zeng-Treitler, Q., Ngo, L. H., Goryachev, S., & Wiechmann, E. P. (2012). Active learning for clinical text classification: Is it better than random sampling? *Journal of the American Medical Informatics Association, 19*(5), 809–816.

Geifman, Y., & El-Yaniv, R. (2017). Deep active learning over the long tail. arXiv:1711.00941

Gissin, D., & Shalev-Shwartz, S. (2019). Discriminative active learning. arXiv:1907.06347

Goudjil, M., Koudil, M., Bedda, M., & Ghoggali, N. (2018). A novel active learning method using svm for text classification. *International Journal of Automation and Computing, 15*(3), 290–298.

Hacohen, G., Ben-David, S., & Shalev-Shwartz, S. (2022). Active learning on a budget: Opposite strategies suit high and low budgets. In *Proceedings of the 38th international conference on machine learning*.

Han, W., Fan, R., Wang, L., Feng, R., Li, F., Deng, Z., & Chen, X. (2020). Improving training instance quality in aerial image object detection with a sampling-balance-based multistage network. *IEEE Transactions on Geoscience and Remote Sensing.*

Huang, Y., Liu, Z., Jiang, M., Xian, Yu., & Ding, X. (2019). Cost-effective vehicle type recognition in surveillance images with deep active learning and web data. *IEEE Transactions on Intelligent Transportation Systems, 21*(1), 79–86.

Imran, A. S., Daudpota, S. M., Kastrati, Z., & Batra, R. (2020). Cross-cultural polarity and emotion detection using sentiment analysis and deep learning on covid-19 related tweets. *IEEE Access, 8*, 181074–181090.

Joloudari, J. H., Hussain, S., Nematollahi, M. A., Bagheri, R., Fazl, F., Alizadehsani, R., Lashgari, R., & Talukder, A. (2023). Bert-deep cnn: State of the art for sentiment analysis of covid-19 tweets. *Social Network Analysis and Mining, 13*(1), 99.

Kim, K., Park, D., Kim, K. I., & Chun, S. Y. (2021). Task-aware variational adversarial active learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 8166–8175).

Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. arXiv:1312.6114

Kuncheva, L. I., Arnaiz-González, Á., Díez-Pastor, J.-F., & Gunn, I. A. D. (2019). Instance selection improves geometric mean accuracy: A study on imbalanced data classification. *Progress in Artificial Intelligence, 8*(2), 215–228.

Kunneman, F., Lambooij, M., Wong, A., van den Bosch, A., & Mollema, L. (2020). Monitoring stance towards vaccination in twitter messages. *BMC Medical Informatics and Decision Making, 20*(1), 1–14.

Kwolek, B., Koziarski, M., Bukała, A., Antosz, Z., Olborski, B., Wąsowicz, P., Swadźba, J., & Cyganek, B. (2019). Breast cancer classification on histopathological images affected by data imbalance using active learning and deep convolutional neural network. In *International conference on artificial neural networks* (pp. 299–312). Springer.

Li, Y., Fan, B., Zhang, W., Ding, W., & Yin, J. (2021). Deep active learning for object detection. *Information Sciences, 579*, 418–433.

Liu, J., Cao, L., & Tian, Y. (2020a). Deep active learning for effective pulmonary nodule detection. In *International conference on medical image computing and computer-assisted intervention* (pp. 609–618). Springer.

Liu, M., Tu, Z., Wang, Z., & Xu, X. (2020b). Ltp: A new active learning strategy for bert-crf based named entity recognition. arXiv:2001.02524

Liu, Y., Li, Z., Zhou, C., Jiang, Y., Sun, J., Wang, M., & He, X. (2019). Generative adversarial active learning for unsupervised outlier detection. *IEEE Transactions on Knowledge and Data Engineering, 32*(8), 1517–1528.

Longpre, S., Reisler, J., Huang, E. Greg, L., Yi, F., Andrew, R., Nikhil, & DuBois, C. (2022). Active learning over multiple domains in natural language tasks. arXiv:2202.00254

Luo, J., Wang, J., Cheng, N., & Xiao, J. (2021). Loss prediction: End-to-end active learning approach for speech recognition. In *2021 International joint conference on neural networks (IJCNN)* (pp. 1–7). IEEE.

Lwin, M. O., Jiahui, L., Sheldenkar, A., Schulz, P. J., Shin, W., Gupta, R., & Yang, Y. (2020). Global sentiments surrounding the covid-19 pandemic on twitter: Analysis of twitter trends. *JMIR Public Health and Surveillance, 6*(2), e19447.

Mayer, C., & Timofte, R. (2020). Adversarial sampling for active learning. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision* (pp. 3071–3079).

Miller, B., Linder, F., & Mebane, W. R. (2020). Active learning approaches for labeling text: Review and assessment of the performance of active learning approaches. *Political Analysis, 28*(4), 532–551.

Mitchell, A., Jurkowitz, M., Baxter Oliphant, J., & Shearer, E. (2021). *The connection between social media use and vaccine hesitancy*. Salon.

Mittal, S., Tatarchenko, M., Çiçek, Ö., & Brox, T. (2019). Parting with illusions about deep active learning. arXiv:1912.05361

Mottaghi, A, & Yeung, S. (2019) Adversarial representation active learning. arXiv:1912.09720

Müller, M., Salathé, M., & Kummervold, P. E. (2020). Covid-twitter-bert: A natural language processing model to analyse covid-19 content on twitter. arXiv:2005.07503

Muqtadiroh, F. A., Purwitasari, D., Yuniarno, E. M., Nugroho, S. M. S., & Purnomo, M. H. (2021). Analysis the opinion of school-from-home during the covid-19 pandemic using lstm approach. In *2021 International seminar on intelligent technology and its applications (ISITIA)* (pp. 408–413). IEEE.

Nam, J. G., Park, S., Hwang, E. J., Lee, J. H., Jin, K.-N., Lim, K. Y., Vu, T. H., Sohn, J. H., Hwang, S., Goo, J. M., et al. (2019). Development and validation of deep learning-based automatic detection algorithm for malignant pulmonary nodules on chest radiographs. *Radiology, 290*(1), 218–228.

Naseem, U., Khushi, M., Khan, S. K., Shaukat, K., & Moni, M. A. (2021). A comparative analysis of active learning for biomedical text mining. *Applied System Innovation, 4*(1), 23.

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.*

Noor, S., Guo, Y., Shah, S. H. H., Fournier-Viger, P., & Saqib Nawaz, M. (2020). Analysis of public reactions to the novel coronavirus (covid-19) outbreak on twitter. *Kybernetes.*

Nwafor, E., Vaughan, R., & Kolimago, C.. (2021). Covid vaccine sentiment analysis by geographic region. In *2021 IEEE international conference on big data (big data)* (pp. 4401–4404). IEEE.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research, 12*, 2825–2830.

Peris, A., & Casacuberta, F. (2018). Active learning for interactive neural machine translation of data streams. arXiv:1807.11243

Prabhu, S., Mohamed, M., & Misra, H. (2021). Multi-class text classification using bert-based active learning. arXiv:2104.14289

Prabucki, T. P. (2021). Sentiment analysis of sars-cov-2 vaccination tweets using deep neural networks.

Preda, G. (2021). All covid-19 vaccines tweets.

Rahman, Md., Islam, M. N., et al. (2022). Exploring the performance of ensemble machine learning classifiers for sentiment analysis of covid-19 tweets. In *Sentimental analysis and deep learning* (pp. 383–396). Springer.

Ren, J., Wang, Y., Mao, M., & Cheung, Y. (2022). Equalization ensemble for large scale highly imbalanced data classification. *Knowledge-Based Systems, 242*, 108295.

Ren, P., Xiao, Y., Chang, X., Huang, P.-Y., Li, Z., Gupta, B. B., Chen, X., & Wang, X. (2021). A survey of deep active learning. *ACM Computing Surveys (CSUR), 54*(9), 1–40.

Sahan, M., Smidl, V., & Marik, R.. (2021). Active learning for text classification and fake news detection. In *2021 International symposium on computer science and intelligent controls (ISCSIC)* (pp. 87–94). IEEE.

Sattar, N. S., & Arifuzzaman, S. (2021). Covid-19 vaccination awareness and aftermath: Public sentiment analysis on twitter data and vaccinated population prediction in the USA. *Applied Sciences, 11*(13), 6128.

Shui, C., Zhou, F., Gagné, C., & Wang, B.. (2020). Deep active learning: Unified and principled method for query and training. In *International conference on artificial intelligence and statistics* (pp. 1308–1318). PMLR.

Siddhant, A., & Lipton, Z. C. (2018). Deep Bayesian active learning for natural language processing: Results of a large-scale empirical study. arXiv:1808.05697

Sinha, S., Ebrahimi, S., & Darrell, T. (2019). Variational adversarial active learning. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 5972–5981).

Sourbier, N., Bonnot, J., Majorczyk, F., Gesny, O., Guyet, T., & Pelcat, M. (2022). Imbalanced classification with tpg genetic programming: Impact of problem imbalance and selection mechanisms. In *Proceedings of the genetic and evolutionary computation conference companion* (pp. 608–611).

Stafanovičs, A., Bergmanis, T., & Pinnis, M. (2020). Mitigating gender bias in machine translation with target gender annotations. arXiv:2010.06203

Stark, F., Hazırbas, C., Triebel, R., & Cremers, D. (2015). Captcha recognition with active deep learning. In *Workshop new challenges in neural computation* (Vol. 2015, p. 94). Citeseer.

To, Q. G., To, K. G., Huynh, V.-A.N., Nguyen, N. T. Q., Ngo, D. T. N., Alley, S. J., Tran, A. N. Q., Tran, A. N. P., Pham, N. T. T., Bui, T. X., et al. (2021). Applying machine learning to identify anti-vaccination tweets during the covid-19 pandemic. *International Journal of Environmental Research and Public Health, 18*(8), 4069.

Tran, T., Do, T.-T., Reid, I., & Carneiro, G. (2019). Bayesian generative active deep learning. In *International conference on machine learning* (pp. 6295–6304). PMLR.

Villavicencio, C., Macrohon, J. J., Alphonse Inbaraj, X., Jeng, J.-H., & Hsieh, J.-G. (2021). Twitter sentiment analysis towards covid-19 vaccines in the philippines using naïve Bayes. *Information, 12*(5), 204.

Wang, G., & Ren, P. (2020). Hyperspectral image classification with feature-oriented adversarial active learning. *Remote Sensing, 12*(23), 3879.

Wang, W., Lu, Y., Wu, B., Chen, T., Chen, D. Z., & Wu, J. (2018). Deep active self-paced learning for accurate pulmonary nodule segmentation. In *International conference on medical image computing and computer-assisted intervention* (pp. 723–731). Springer.

Wilson, S. L., & Wiysonge, C. (2020). Social media and vaccine hesitancy. *BMJ global health, 5*(10), e004206.

Xing, W., Chen, C., Zhong, M., Wang, J., & Shi, J. (2021). Covid-al: The diagnosis of covid-19 with deep active learning. *Medical Image Analysis, 68*, 101913.

Xue, J., Chen, J., Chen, C., Zheng, C., Li, S., & Zhu, T. (2020). Public discourse and sentiment during the covid 19 pandemic: Using latent Dirichlet allocation for topic modeling on twitter. *PloS one, 15*(9), e0239441.

Yan, Y.-F., Huang, S.-J., Chen, S., Liao, M., & Xu, J. (2020). Active learning with query generation for cost-effective text classification. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 34, pp. 6583–6590).

Yang, L., Zhang, Y., Chen, J., Zhang, S., & Chen, D. Z. (2017). Suggestive annotation: A deep active learning framework for biomedical image segmentation. In *International conference on medical image computing and computer-assisted intervention* (pp. 399–407). Springer.

Yoo, D., & Kweon, I. S. (2019). Learning loss for active learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 93–102).

Yuan, D., Chang, X., Liu, Q., Yang, Y., Wang, D., Shu, M., He, Z., & Shi, G. (2023). Active learning for deep visual tracking. *IEEE Transactions on Neural Networks and Learning Systems*.

Yue, Z., Zeng, H., Kou, Z., Shang, L., & Wang, D. (2022). Contrastive domain adaptation for early misinformation detection: A case study on covid-19. In *Proceedings of the 31st ACM international conference on information and knowledge management* (pp. 2423–2433).

Zhang, B., Li, L., Yang, S., Wang, S., Zha, Z.-J., & Huang, Q. (2020). State-relabeling adversarial active learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 8756–8765).

Zhang, Y., Lease, M., & Wallace, B. (2017). Active discriminative text representation learning. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 31).

Zhang, Y., Zhang, X., Zhang, R., Wang, R., Zhang, Q., Wang, Y., Liang, Y., Liang, H., & Liu, J. (2021). Vaccine hesitancy and behavior change theory-based social media intervention: A randomized controlled trial. *Vaccine, 40*(4), 647–654.

Zhou, S., Chen, Q., & Wang, X. (2013). Active deep learning method for semi-supervised sentiment classification. *Neurocomputing, 120*, 536–546.

Zhu, J.-J., & Bento, J. (2017). Generative adversarial active learning. arXiv:1702.07956

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Sankhadeep Chatterjee[1] · **Saranya Bhattacharjee**[2] · **Asit Kumar Das**[1] · **Soumen Banerjee**[3]

✉ Sankhadeep Chatterjee
chatterjeesankhadeep.cu@gmail.com

Saranya Bhattacharjee
saranyab1912@gmail.com

Asit Kumar Das
akdas@cs.iiests.ac.in

Soumen Banerjee
prof.sbanerjee@gmail.com

[1] Department of Computer Science and Technology, Indian Institute of Engineering Science and Technology, Shibpur, Howrah, India

[2] Department of Computer Science and Engineering, University of Engineering and Management, Kolkata, India

[3] Narula Institute of Technology, Agarpara, Kolkata, West Bengal 700109, India